Equilibration and Typicality in Quantum Processes

Pedro Figueroa Romero



School of Physics and Astronomy
Monash University

A thesis submitted for the degree of
Doctor of Philosophy

September 2020

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

© Pedro Figueroa Romero (2020)

Abstract

Forgetfulness is a common occurrence in natural processes. After all, if each carbon atom remembered its detailed past, then each of these would have a unique behaviour and there would be no sense in classifying atoms and molecules. Moreover, without forgetfulness, repeatability would be impossible. Despite this, small systems constantly leak information about their state to their surroundings, and quantum mechanics tells us that this information can never be deleted, so that it invariably returns to influence their future behaviour.

How can physical nature be forgetful if it is not allowed to forget? Precisely, in the theory of open quantum systems, memory is the rule and forgetfulness the exception. This conundrum is not dissimilar from that of the emergence of the laws of thermodynamics purely from underlying quantum mechanical laws, which dates from the inception of quantum mechanics itself, and is still fertile ground for important foundational and practical questions.

In this thesis, we address the emergence of forgetfulness—more precisely called *Markovianity*—through a generalization of the concepts of equilibration on average and typicality, previously employed to study the emergence of statistical mechanics from quantum mechanics: the first shows how time-dependent quantities of quantum systems evolve towards fixed values and stay close to them for most times, while the second one argues that small subsystems of a composite are in equilibrium for almost all states of the whole. Using the process tensor framework for quantum stochastic processes, we introduce an extended notion of equilibration, characterizing the conditions under which a multitime quantum process can be approximately described by an equilibrium one. Furthermore, without resorting to the Born-Markov assumption of weak coupling, we show that Markovian processes are typical, and prove that there are physical non-Markovian processes that with high probability look almost Markovian for all orders of correlations, in a phenomenon that we call *Markovianization*.

The results within this thesis bridge the aforementioned gap between what we see in the real world and what idealised physical theories say. The main implications of these findings are twofold: foundationally, they give birth to almost Markovian processes from closed quantum dynamics, while for applications and experiments, they pave the way to predict and quantify the rate at which memory effects become relevant.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Melbourne, September 2020

Acknowledgements

I have grown incommensurately in the last three and a half years, and despite reaching this point in such dire times (amidst the COVID-19 pandemic), I can only be grateful for all the people who supported me in one way or another.

I am deeply thankful to both Kavan Modi and Felix Pollock for giving me the chance to do this research, for their patience, enthusiasm, kindness and guidance; they are two of the brightest minds that I have been fortunate to learn from and I hold both in the highest regard as scientists and as human beings. I was also very fortunate to have been part of the unique Monash Quantum Information Science (MonQIS) group, I am grateful for the support and insightful conversations with all its members, as well as for the memes and the shenanigans. My special thanks to Francesco Campaioli, Mathias Jørgensen, Roberto Muñoz and Magdalini Zonnios for their support, friendship and good times. Thanks as well to Cody Duncan, honorary classical MonQIS member, for helping me proofreading my early drafts.

I specially thank Cristian Sierra, who showed up when I was in the middle of my PhD and is now a dearest friend and brother for life, he was there to help me get through the darkest days as well as to have fun on the sunny ones, I thank him for sharing both the mundane and the outstanding with me.

I am sincerely grateful to the whole School of Physics & Astronomy for the accommodating and amicable atmosphere, together with all the academic and administrative facilities and support. Financially, this work would not have been possible without the Monash Graduate Scholarship (MGS) and the Monash International Postgraduate Research Scholarship (MIPRS).

At this stage of my life, I am now fairly convinced that luck plays a higher role in life than most talents one could put forth, and in this sense, I am incredibly lucky for having my family, *mamá, papá y mi hermana Ale*, and their support from so far away for all this time. To be writing these lines was literally unbelievable ten years ago as a *high-school* dropout, but I never lacked their unconditional support and this is only possible due to them.

Finally, in this same spirit, I am grateful to Felicia Grant for filling my life with light and colour, bringing warmth and safety in these uncertain times, and being my constant reminder of what really matters in life.

Publications

The main results in this thesis have been presented in the following publications:

[1] *Equilibration on average in quantum processes with finite temporal resolution*
Pedro Figueroa-Romero, Felix A. Pollock and Kavan Modi
Accepted (15/09/2020) in Physical Review E, in press.

[2] *Almost Markovian processes from closed dynamics*
Pedro Figueroa-Romero, Kavan Modi, and Felix A. Pollock
Quantum 3, 136 (2019).

[3] *Markovianization by design*
Pedro Figueroa-Romero, Felix A. Pollock and Kavan Modi
Preprint: arXiv:2004.07620 [quant-ph].

By date of writing and publication, these are ordered as [2], [1], [3], but for sake of presentation we follow the order cited above in the corresponding Chapters 3, 4 and 5.

No hay nada más vertiginoso que mirar atrás.

Contents

Abstract iii

Declaration iv

Acknowledgements v

Publications vi

Introduction xiii

Outline . . . xiv

I Background 1

0 Quantum Mechanics 101 2

0.1 Quantum systems and quantum states . . . 2

0.2 Measurements and observables . . . 4

0.3 Closed system dynamics . . . 5

0.4 Distinguishability of quantum states: trace distance . . . 6

0.5 Entanglement . . . 7

1 The foundations of Statistical Mechanics 9

1.1 Equilibrium and the second law of thermodynamics . . . 9

1.2 Equilibration on average . . . 11

1.2.1 The equilibrium state . . . 12

1.2.2 Temporal fluctuations . . . 13

1.2.3 Fluctuations within a finite time . . . 18

1.2.4 Trace distance equilibration . . . 20

1.3 Typicality . . . 22

1.3.1 Random states and the Haar measure . . . 23

1.3.2 Twirling and the Schur-Weyl duality . . . 25

1.3.3 Average states and average purity . . . 26

1.3.4 Concentration of measure . . . 28
1.3.5 Typicality of entangled states . . . 30
1.3.6 Entanglement as a canonical principle . . . 32

2 Quantum Processes **35**
2.1 Quantum channels . . . 36
2.2 Three representations of quantum channels . . . 37
2.2.1 Operator sum representation . . . 38
2.2.2 Dilations . . . 39
2.2.3 Choi-Jamiołkowski isomorphism . . . 40
2.3 Open quantum dynamics . . . 42
2.3.1 The Born-Markov approximation . . . 43
2.3.2 The initial correlation problem . . . 45
2.3.3 The superchannel . . . 46
2.4 Multiple time-steps . . . 50
2.4.1 Classical stochastic processes . . . 50
2.4.2 The classical Markov condition . . . 53
2.4.3 The process tensor . . . 55
2.4.4 The quantum Markov condition . . . 61
2.4.5 An unambiguous measure for non-Markovianity . . . 63

II Main Results **66**

3 Equilibration in Quantum Processes across Multiple Points in Time **67**
3.1 Equilibration due to finite temporal resolution . . . 68
3.2 Multitime equilibration due to finite temporal resolution . . . 71
3.3 Genuine multitime equilibration . . . 78
3.4 Conclusions . . . 79

4 Markovian Typicality **80**
4.1 Random quantum processes . . . 81
4.2 The moments of the unitary group & average processes . . . 82
4.3 A bound on non-Markovianity & average noisiness . . . 87
4.3.1 Limiting cases . . . 89
4.4 Concentration around Markovian processes . . . 91
4.4.1 Numerical sampling . . . 95
4.5 Observing non-Markovianity . . . 96
4.6 Conclusions . . . 97

5 Markovianization by Design **99**
5.1 Approximate unitary designs . . . 100

5.2 Large deviations on unitary designs around Markovian processes 103
5.3 Markovianization by circuit design 108
5.4 Conclusions 111

III Envoi **113**

6 Conclusions **114**
6.1 Outlook 115

Appendices **117**

A Notation **118**

B Haar distributed process tensors **122**
B.1 Average processes - Constant interaction 122
B.1.1 Superchannel case 123
B.2 Average purity - Random interaction 124
B.3 Average purity - Constant interaction 126
B.3.1 Small subsystem limit 128
B.3.2 Long time limit 129

C Sampling from the unitary group **130**

Bibliography **132**

List of Figures

1.1 The equilibrium state as a dephasing of the initial state . . . 13
1.2 Equilibration on average on expectation values . . . 16
1.3 Levy's lemma: concentration of measure on the sphere . . . 30

2.1 Stinespring dilation of a quantum channel . . . 39
2.2 Choi-Jamiołkowski isomorphism . . . 41
2.3 Circuit diagram of the superchannel . . . 46
2.4 Complete positivity of the superchannel . . . 47
2.5 The Choi state of the superchannel . . . 48
2.6 The action of the superchannel in terms of Choi states . . . 49
2.7 Trajectory in classical stochastic processes . . . 52
2.8 A k-step process tensor . . . 56
2.9 Choi state representation of the process tensor . . . 57
2.10 Containment property of the process tensor . . . 59
2.11 Correlated operations on a process tensor . . . 60
2.12 Causal break . . . 62
2.13 Markovian process with open quantum evolution . . . 64

3.1 Multitime fuzzy clocks . . . 71
3.2 Equilibrium process and fuzzy multitime equilibration . . . 73

4.1 Sampling a quantum process at random and quantum Markovian processes . . . 82
4.2 Concentration around Markovian processes in large dimensional environments . . . 94
4.3 Average non-Markovianity on numerical sampling . . . 95

5.1 Unitary designs, from simplicity to randomness . . . 101
5.2 Large deviations bound on non-Markovianity with unitary designs . . . 108
5.3 Markovianization by circuit design . . . 109
5.4 Markovianization by circuit design: non-commuting gate depth . . . 110

Acronyms

CJI Choi-Jamiołkowski isomorphism.

CPTNI Completely Positive Trace Non-Increasing.

CPTP Completely Positive Trace Preserving.

CP Completely Positive.

E Environment.

NCP Not Completely Positive.

PDF Probability Density Function.

POVM Positive-Operator Valued Measure.

SΓ System-Ancilla.

SE System-Environment.

S System.

TNI Trace Non-Increasing.

TP Trace Preserving.

Introduction

Thermodynamics is without a doubt the most resilient, universal and, in a sense, *strange* theory of physics. It was originally devised as a phenomenological theory, surviving all major revolutions in science and serving as the highest authority for most technological developments. Similarly, the concept of energy pervades the social imaginarium (albeit often in pseudo-scientific ways) and there's arguably no more popular physical law than the first law of thermodynamics, which, together with the remaining three laws (or two, depending who one asks), no one has yet been able to contradict in a real laboratory.

On the other hand, quantum mechanics is one of the fundamental pillars on which any physical theory must stand. By circa 1930, the mathematical formalism of quantum mechanics was established thoroughly by John von Neumann in Ref. [4], and given that statistical mechanics—the mathematical spinal cord of thermodynamics—was already a mature field of research, it is no surprise that he would quickly approach a quantum theory of statistical mechanics and raise foundational questions that remained largely unsolved until very recently.

A characteristic feature of quantum mechanics is that of generating a distinct type of correlations which is non-local in nature, and this was just recently seen to be key in explaining the quantum emergence of statistical mechanics. Correlations, however, can also be temporal, and it turns out that many foundational questions regarding these correlations in quantum mechanics can be posed in an analogous way to those regarding the foundations of statistical mechanics. Namely, how do quantum systems become *forgetful*, i.e. behaving independently of their detailed past? Or how resilient is a system in thermal equilibrium when temporal correlations are present and when the act of observing the system itself disturbs it?

The overarching goal of this thesis is to provide an in-depth and novel account of the relationship between the phenomenon of equilibration, which is a pillar in the foundations of statistical mechanics, and general quantum processes that incorporate a full account of correlations and memory effects; namely, under what conditions do quantum processes with seemingly negligible temporal correlations arise, given that processes with non-vanishing correlations are the norm? And furthermore, how do we account for the pervasiveness of these in nature? In investigating these questions, we find a non-trivial interplay between complexity and randomness, further characterizing how equilibration comes about in quantum processes.

Outline

The rest of text is organised as follows: Part I consists of Chapters 0 to 2, containing an overview and explanation of existing literature and no original results, followed by Part II, which consists of Chapters 3 to 5 and discusses the original results obtained during this PhD.

Chapter 0 We briefly introduce the concepts and the mathematical framework of quantum mechanics such as quantum states, observables, measurements, closed system evolution and distinguishability of quantum states.

Chapter 1 We first motivate this chapter with a brief discussion of the second law of thermodynamics and why a derivation purely from quantum mechanics is needed. This is followed by a discussion of the main results in the literature showing how quantum systems dynamically fulfil the second law given solely the quantum mechanical framework, in a concept known as equilibration on average. Finally, we discuss how the backbone of the second law, namely the fundamental postulate of statistical mechanics, emerges naturally through entanglement in quantum mechanics without any a-priori assumptions by appealing to a notion of typicality.

Chapter 2 We begin by describing more general quantum operations with the concept of quantum channels and three of their main representations, namely dilations, the operator sum representation and the Choi Jamiołkowski isomorphism. We then describe open quantum dynamics and the problem of initial correlations, together with a resolution known as the superchannel. Following this, we motivate the generalization encoding the initial correlations problem via multiple interventions, whereby temporal correlations between more than two points become relevant, in turn leading to the generalization of the superchannel known as the process tensor. We describe how the process tensor framework generalizes the concept of classical stochastic processes and the concept of Markovianity, contrasting with several different approaches in the literature which have proved to be problematic. We finally discuss how the process tensor framework naturally provides an unambiguous measure of non-Markovianity.

Chapter 3 In this chapter we blend the concepts in the first chapters to study the conditions under which a process with a finite temporal resolution can be approximately described by an equilibrium one, which is equivalent to having operations being implemented with a fuzzy clock or to having a system with uniformly fluctuating energies. We first define what we mean by an equilibrium process and by a fuzzy clock, and we then derive a generalization of the concept of equilibration on average to one which can be operationally assessed at multiple times, placing an upper-bound on a new observable distinguishability measure comparing a multitime process with a fuzzy clock against a fixed equilibrium one. We will see that the conditions for equilibration to occur can be extended, with genuine multitime contributions depending on the fuzzy process and the amount of disturbance

of the observer's operations on it. We finally motivate a parallel between the concepts of subsystem equilibration, thermalization and the emergence of statistical mechanics, and process equilibration, *Markovianization* and the emergence of memorylessness in nature. This chapter is based on Ref. [1]

Chapter 4 By bridging the ideas from the discussion of the emergence of statistical mechanics and the postulate of equal a-priori probabilities to the realm of quantum processes, we argue that this naturally leads to the question of the emergence of memoryless processes, known as Markovian, purely from the rules of quantum mechanics. Motivated by the results on typicality for quantum states, we are able to formally prove that a quantum process drawn uniformly at random will be almost memoryless with high probability whenever it is undergone within a large-dimensional environment. We argue that our results have a parallel interpretation to the case of the emergence of statistical mechanics in the sense of replacing ad-hoc assumptions and approximations to render quantum processes memoryless. We finally discuss the limitations of the typicality approach, the most contentious of which is further discussed in the following chapter. This chapter is based on Ref. [2]

Chapter 5 Almost all quantum processes drawn at random within a large environment will be almost Markovian. However, nature seldom behaves randomly. In this chapter we identify a class of physically motivated quantum processes –known as unitary designs– that satisfy a statement known as a large deviations bound, quantifying the probability that these differ greatly from their Markovian counterparts. We show that, similar to the way that quantum systems thermalize, quantum processes *Markovianize* in the sense that they can converge towards Markovian processes in the correct limits, and in particular as the overall complexity of the interactions in the whole system increases. We further exemplify our result making use of an efficient construction of an approximate unitary design with an n-qubit quantum circuit mediated by two-qubit interactions only, showing how seemingly simple systems can speedily become forgetful. We finally discuss potential applications as well as further open questions related to the forgetfulness of nature. This chapter is based on Ref. [3].

PART I

Background

CHAPTER 0

Quantum Mechanics 101

Quantum theory is a set of rules allowing the computation of probabilities for the outcomes of tests which follow specified preparations.

Asher Peres ([5])

In this chapter we briefly introduce the basic notation and mathematical concepts from quantum mechanics required for all the remaining topics covered this PhD. These can be consulted in standard textbooks, such as Ref. [6–8].

0.1 Quantum systems and quantum states

The essential ingredients of quantum mechanics are quantum states, the transformations between them and the measurement outcomes that we observe.

The space of definite states of every quantum system is isomorphic to the space of rays in a Hilbert space $\mathscr{H}$, that is, the equivalence class of proportional vectors in $\mathscr{H}$. Henceforth we will restrict ourselves to finite, d-dimensional Hilbert spaces over the set of complex numbers $\mathbb{C}$ with inner product $\langle\phi|\psi\rangle$ for $|\phi\rangle$, $|\psi\rangle$ vectors in $\mathscr{H}$. A bipartite quantum system AB comprising systems A and B is associated with a tensor product $\mathscr{H}_{\mathsf{AB}} = \mathscr{H}_{\mathsf{A}} \otimes \mathscr{H}_{\mathsf{B}}$, which is such that, if $\{|\alpha\rangle\}_{\alpha=1}^{d_{\mathsf{A}}}$ and $\{|\beta\rangle\}_{\beta=1}^{d_{\mathsf{B}}}$ are bases for A and B, respectively, then any vector in $\mathscr{H}_{\mathsf{AB}}$ can be represented by $|\Psi\rangle = \sum_{\alpha,\beta} \psi_{\alpha\beta}|\alpha\beta\rangle$ for coefficients $\psi_{\alpha\beta} \in \mathbb{C}$ and where $|\alpha\beta\rangle := |\alpha\rangle \otimes |\beta\rangle$. Any n-partite space $\mathscr{H} \cong \mathscr{H}_1 \otimes \cdots \otimes \mathscr{H}_n$ then is built similarly by extension.

The most general state of a quantum system is then specified by a density operator ρ, which is an element of the space of bounded operators $\mathscr{B}(\mathscr{H})$[1] and which additionally is set to satisfy

$$\rho = \rho^\dagger \,(\text{Hermiticity}), \quad \rho \geq 0 \,(\text{positivity}) \quad \text{and} \quad \operatorname{tr}(\rho) = 1 \,(\text{unit trace}), \tag{1}$$

where Hermiticity means ρ equals its conjugate transpose, here denoted by $\dagger$, which implies in turn that all its eigenvalues are real. We denote the subset of density operators on a given space by $\$(\mathscr{H}) \subset \mathscr{B}(\mathscr{H})$. Positivity explicitly means positive semi-definite, $\langle\varphi|\rho|\varphi\rangle \geq 0$ for any vector

[1] In general these also must have a finite trace; this is ensured for finite-dimensional systems.

$|\varphi\rangle \in \mathscr{H}$, which implies that all of the eigenvalues of ρ are non-negative; we will commonly refer to this property *for quantum states* simply as positivity. Together with unit trace, these will ensure all probabilities corresponding to the outcomes of a measurement of a quantum system are real, positive and add up to unity. We will commonly refer to density operators simply as quantum states.

A quantum state $\rho \in \$(\mathscr{H})$ is called pure if there exists a vector $|\psi\rangle$ such that $\rho = |\psi\rangle\langle\psi|$, or equivalently if its rank, i.e. the dimension of its image, is equal to one. From the spectral theorem then it follows that every quantum state is a convex mixture of pure states, $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, with $p_i \geq 0$ and $\sum p_i = 1$. The purity of a quantum state is given by

$$\frac{1}{d} \leq \mathrm{tr}(\rho^2) \leq 1, \tag{2}$$

which is known as the purity of the state ρ. The upper-bound is saturated when the state is pure, whilst the lower bound is reached for the so-called maximally mixed state, given by $\rho = \mathbb{1}/d$, where $\mathbb{1}$ is the identity operator on $\mathscr{H}$, here the $d \times d$ identity matrix. This is interpreted as the state of maximal ignorance, as the system has equal probability to be in any possible pure state.

Furthermore, any mixed state can be expressed as a reduced state of a pure state in a larger Hilbert space: this is known as purification. Here *reduced* means a state of a subset of degrees of freedom of the full system, i.e. for a quantum state in a bipartite system, $\rho \in \$(\mathscr{H}_\mathsf{A} \otimes \mathscr{H}_\mathsf{B})$, we define

$$\rho_\mathsf{A} := \mathrm{tr}_\mathsf{B}(\rho) := \sum_{\beta=1}^{d_\mathsf{B}} (\mathbb{1}_\mathsf{A} \otimes \langle\beta|)\, \rho\, (\mathbb{1}_\mathsf{A} \otimes |\beta\rangle), \tag{3}$$

as the reduced state on space A, where $\mathrm{tr}_\mathsf{B}(\cdot)$ is called a partial trace, defined as a trace over subspace B and with $\mathbb{1}_\mathsf{A}$ the identity operator solely on A. Similarly, $\rho_\mathsf{B} = \mathrm{tr}_\mathsf{A}(\rho)$ is the reduced state of ρ on subspace B. It is clear by inspection that the resulting ρ_A, ρ_B are legitimate quantum states. Thus purification means that every mixed quantum state $\rho \in \$(\mathscr{H})$ can be expressed as $\rho = \mathrm{tr}_\Gamma[|\Psi\rangle\langle\Psi|]$ for some pure state $|\Psi\rangle \in \mathscr{H} \otimes \mathscr{H}_\Gamma$. The system Γ is usually referred to as an ancillary space or just an ancilla.[2] This is easily seen by the so-called Schmidt decomposition, which ensures that any bipartite Hilbert space vector can be written in the form $|\Psi\rangle = \sum_{i=1}^{D} \sqrt{\varphi_i} |u_i v_i\rangle$ where here $|u_i\rangle$ and $|v_i\rangle$ are orthonormal states in the respective subsystems, $D = \min(d, d_\Gamma)$, and φ_i are strictly positive coefficients such that $\sum \varphi_i = 1$ [5]. Then for such $|\Psi\rangle$ we obtain the reduced state $\rho = \sum \varphi_i |u_i\rangle\langle u_i|$, so we can always go in the opposite direction by decomposing any quantum state via the spectral theorem and using it to construct a pure state in an extended space incorporating an ancilla. Notice that in the case of a purification, $d \leq d_\Gamma$ and that such purification will not be unique. This concept has far reaching consequences as will be shown below and can be clearly motivated physically by thinking of the ancillary space as an environment.

[2] *Ancilla* is the Latin term for maidservant; despite other (mainly negative) connotations it potentially carries, it is now standard in quantum information science as a synonym of *auxiliary*.

0.2 Measurements and observables

Whereas in classical mechanics we can describe the state of a system in a somehow passive way, in quantum mechanics we need access to the density operator through other operators that play an active role in a sense we will now describe. The concept of an observable is tightly related with that of a measurement: the most general measurements are represented by a finite ordered set $\{\mathrm{M}_i\}$ called a Positive-Operator Valued Measure (`POVM`), where the elements M_i are such that

$$\sum_i \mathrm{M}_i = \mathbb{1}, \quad \text{with} \quad \mathrm{M}_i = \mathrm{M}_i^\dagger \quad \text{and} \quad \mathrm{M}_i \geq 0, \tag{4}$$

i.e. Hermitian positive semidefinite operators forming a partition of the identity operator in $\mathscr{H}$. A `POVM` measurement applied to a state ρ produces the i^{th} outcome with probability $\mathrm{tr}(\mathrm{M}_i\rho)$, with the definition of a `POVM` ensuring that these sum up to unity. A `POVM` is called *informationally complete* if its statistics fully determine the density matrix, which requires at least d^2 elements [7].

In particular, whenever the `POVM` consists of d elements with all being orthogonal *projectors*, i.e. $\{\mathrm{M}_i = \Pi_i\}$ satisfying

$$\Pi_i\Pi_j = \Pi_i\delta_{ij}, \qquad i,j = 1,2,\ldots,d, \tag{5}$$

this is called a *projective measurement*. We then refer to a Hermitian operator as an observable whenever its real eigenvalues correspond to measurable outcomes. In particular then, an observable with spectral decomposition $A = \sum_{i=1}^{d} \alpha_i\Pi_i$ describes a projective measurement with probability $\mathrm{tr}(\Pi_i\rho)$ of obtaining the i^{th} outcome α_i.

Upon measuring, a quantum system will generally change its state. In general a `POVM` is not enough to determine the post-measurement state. In Chapter 2 we will introduce the notion of so-called Kraus operators which will let us deal with this; in particular, for projective measurements, the projectors themselves are Kraus operators and after measurement the state is generally[3] transformed to $\rho \to \rho' = \sum_i \Pi_i\,\rho\,\Pi_i$, or in particular, if the i^{th} outcome is observed, then

$$\rho \to \rho' = \frac{\Pi_i\,\rho\,\Pi_i}{\mathrm{tr}[\Pi_i\rho]}, \tag{6}$$

which is called a *selective* measurement in Ref. [7]. Selective measurements are repeatable in the sense that if they are performed again the post measurement state remains the same; this is not true for general `POVM`s.

Finally, we will denote the *expectation value* of A on the state ρ by

$$\langle A\rangle_\rho := \mathrm{tr}(A\rho) = \sum_i \alpha_i\,\mathrm{tr}(\Pi_i\rho), \tag{7}$$

so that $\mathrm{tr}(\Pi_i\rho)$ is the corresponding probability for the i^{th} outcome α_i, which in essence constitutes Born's rule [9, 10]. In particular, the *measurement statistics* of a `POVM` will refer to the vector of probabilities $\mathrm{tr}(\Pi_i\rho)$.

[3] More generally, even in the case of orthogonal projectors as `POVM` elements, projectors as Kraus operators need not be implied in the post-measurement state.

0.3 Closed system dynamics

We now discuss the dynamical picture for quantum systems. A *closed* system described by ρ at a given time $t \in \mathbb{R}^+$, where $\mathbb{R}^+$ denotes the set of positive real numbers, will evolve unitarily according to the Schrödinger equation,[4]

$$i\frac{\partial}{\partial t}\rho(t) = [H, \rho(t)], \tag{8}$$

where $H \in \mathscr{B}(\mathscr{H})$ is the observable known as the Hamiltonian of the system, and where we set units $\hbar = 1$. This operator can be said to be generating the dynamics of ρ, and whenever it does not depend on time, as we will consider throughout this thesis, it gives rise to the unitary time-evolution operator

$$U(t) = \exp(-iHt). \tag{9}$$

In general, any operator $V \in \mathscr{B}(\mathscr{H})$ such that $VV^\dagger = V^\dagger V = \mathbb{1}$ is called unitary. We can see that given a state ρ, any $\sigma = V\,\rho\,V^\dagger$ remains a density operator. We thus have the solution to Eq. (8) as

$$\rho(t) = U(t)\,\rho(0)\,U^\dagger(t), \tag{10}$$

where we will usually denote the initial state at time $t = 0$ simply as $\rho(0) = \rho$.

We will commonly write the Hamiltonian in a spectral decomposition

$$H = \sum_{n=1}^{\mathfrak{D}} E_n P_n, \tag{11}$$

where P_n is the spectral projector onto the n^{th} eigenspace of H with energy (eigenvalue) E_n. Here $\mathfrak{D} = |\text{spec}(H)| \leq d$ is the number of distinct energies E_n; if H is degenerate with the n^{th} level having degeneracy ℓ, then $P_n = \sum_{j=1}^{\ell} |n_j\rangle\langle n_j|$ with $\{|n_\ell\rangle\}$ a basis for the n^{th} energy eigenspace.

The expectation value of an observable A with respect to a time-evolved state $\rho(t)$ can then be written as

$$\langle A\rangle_{\rho(t)} = \text{tr}[A\rho(t)] = \text{tr}[A\,U(t)\,\rho\,U^\dagger(t)] = \text{tr}[U^\dagger(t)\,A\,U(t)\,\rho], \tag{12}$$

by the cyclic property of the trace. This can be equivalently thought of as the expectation value $\langle A(t)\rangle_\rho$ of a time-evolved operator $A(t) := U^\dagger(t)\,A\,U(t)$ at time t on the state ρ. This is commonly known as the Heisenberg picture, with the observable A evolving according to

$$-i\frac{\partial}{\partial t}A(t) = [H, A(t)], \tag{13}$$

so that whenever $[H, A(t)] = 0$, the expectation value in Eq. (12) is constant and, similarly to the classical case, A is called a conserved quantity.

[4] In general such evolution for any operator is referred to as the von Neumann or the quantum Liouville equation.

0.4 Distinguishability of quantum states: trace distance

Another important aspect we will require is to be able to quantify how different, or how distinguishable, two quantum states are. In particular, a distinguishability measure called the trace distance will be central to most of our discussions. We can put the task of distinguishing a pair of quantum states in terms of measuring the *distance* between them, i.e. of quantifying how close or far are two states from each other.

Let us first define exactly what we mean by a distance measure.

Definition 0.4.1 (Distance measure)**.** Let $\mathscr{V}$ be a vector space and $x, y \in \mathscr{V}$ any two points. A distance measure $\Delta : \mathscr{V} \times \mathscr{V} \to \mathbb{R}_0^+$ between x and y satisfies:

i. Positivity: $\Delta(x, y) \geq 0$ with equality if and only if $x = y$.
ii. Symmetry: $\Delta(x, y) = \Delta(y, x)$.
iii. Triangle inequality: $\Delta(x, y) \leq \Delta(x, v) + \Delta(v, y)$ for any $v \in \mathscr{V}$.

While there can be a plethora of valid distance measures, not necessarily all of these will constitute a distinguishability measure. By distinguishability measure we specifically refer to a distance measure with an operational meaning, i.e. one that can be ultimately phrased and quantified via measurements. In both the classical and quantum cases, there is no unique way of quantifying distinguishability; however, while the classical case depends just on the statistical state, in the quantum one this depends both on the way the state is measured and the quantum states. That is, considering a `POVM`, say $\{\mathrm{M}_i\}$, the relevant quantity is given by the probabilities $\mathrm{tr}[\mathrm{M}_i\rho]$, as discussed in Section 0.1.

One way to do this, for any two quantum states $\rho, \sigma \in \$(\mathscr{H})$, is to quantify the probability of error in guessing which one of the two is the given state for the system in a single measurement.

Consider then a distance $D_{\{M_i\}}$ defined by

$$D_{\{\mathrm{M}_i\}}(\rho, \sigma) := \frac{1}{2} \sum_i |\,\mathrm{tr}\,\mathrm{M}_i(\rho - \sigma)|, \tag{14}$$

with the $1/2$ being a normalization factor. This corresponds to the distinguishability between ρ and σ given the `POVM` $\{\mathrm{M}_i\}$, as it compares the probabilities for each outcome on either state given a measurement. Specifically, considering a system which was prepared in either state ρ or state σ, we care about guessing which of these the system is actually in, not minding both destroying the actual state and accidentally guessing the wrong outcome. To this end, the best strategy we can take, given the i^{th} outcome, is to guess that the state is ρ if $\mathrm{tr}[\mathrm{M}_i\rho] \geq \mathrm{tr}[\mathrm{M}_i\sigma]$ and to guess it is in σ otherwise. Then we automatically can be right half of the time, with the probability of success in correctly guessing the correct state being

$$\mathbb{P}^{\text{success}}_{\{\mathrm{M}_i\}} = \frac{1}{2} - \frac{1}{2} \sum_i |\,\mathrm{tr}[\mathrm{M}_i\rho] - \mathrm{tr}[\mathrm{M}_i\sigma]| = \frac{1}{2}[1 - D_{\{\mathrm{M}_i\}}(\rho, \sigma)]. \tag{15}$$

If we take a subset of all possible POVMs, say $\mathbb{M}$, we could also define

$$D_{\mathbb{M}} := \max_{\{\mathrm{M}_i\}\in\mathbb{M}} D_{\{\mathrm{M}_i\}} \tag{16}$$

as a distinguishability restricted to such subset. It can be readily verified that this is a legitimate distance measure, and furthermore we have the hierarchy

$$0 \leq D_{\mathbb{M}}(\rho,\sigma) \leq D(\rho,\sigma) \leq 1, \tag{17}$$

where we have defined

$$D(\rho,\sigma) := \frac{1}{2}\|\rho-\sigma\|_1, \tag{18}$$

as the so-called trace distance, where $\|\cdot\|_1$ is the Schatten 1-norm or also sometimes called trace-norm, which can be defined as the case $p=1$ of the family of norms

$$\|X\|_p := \mathrm{tr}\left[|X|^p\right]^{1/p}, \tag{19}$$

called the Schatten p-norms. This is because the trace distance would give the distinguishability measure with the optimal of all possible measurements [7].

Importantly, the Schatten norms satisfy the hierarchy

$$\|\cdot\|_1 \geq \|\cdot\|_2 \geq \ldots \geq \|\cdot\|, \tag{20}$$

where here $\|X\|_\infty := \|X\|$ will be referred to as the *operator norm*, and corresponds to the largest singular value[5] of X.

There are several reasons why the trace distance is important and usually preferred among other state distinguishability measures, which are nevertheless legitimate in their own right. While some of these will become apparent when we present so-called quantum maps in Section 2.4, its operational relevance is overall what makes it a suitable distinguishability measure. Specifically, we can see from Eq. (15) that the trace distance is precisely the one that maximizes the probability of success with the optimal amongst all possible measurements. Similarly, other scenarios where the trace distance is relevant can be seen e.g. in Ref. [11].

0.5 Entanglement

The history of entanglement is well-known, with the discussion beginning when the famous EPR paper [12] came to light, and after which Schrödinger, in correspondence with Einstein [13], coined the term *entanglement* to describe the new kind of correlation. Schrödinger would later add that entanglement is not *one* but rather *the* characteristic trait of quantum mechanics. A full in-depth, geometrical discussion of entanglement can be seen e.g. in Ref. [7].

While the consequences of bipartite entanglement are far reaching (often capturing the popular imagination as well), it is a concept that has a very simple definition. Consider a bipartite

[5] The singular values of a matrix X are the square roots of the eigenvalues of $X^\dagger X$. Thus, for a Hermitian matrix, the singular values are the absolute values of its eigenvalues.

space $\mathscr{H} \cong \mathscr{H}_\mathsf{A} \otimes \mathscr{H}_\mathsf{B}$, then a state $\rho \in \$(\mathscr{H})$ is called *separable* if either it is a product state, $\rho = \rho_\mathsf{A} \otimes \rho_\mathsf{B}$, or if it can be written as a convex combination of product states, $\rho = \sum p_i \rho_\mathsf{A}^{(i)} \otimes \rho_\mathsf{B}^{(i)}$ with $\sum p_i = 1$. Otherwise, the state ρ is called *entangled*.

The simplest example is that of a pair of qubits, with both $\mathscr{H}_\mathsf{A}$ and $\mathscr{H}_\mathsf{B}$ being two-dimensional. Written in the so-called computational basis $\{|0\rangle, |1\rangle\}$, we can find an orthogonal basis with this property (as we will see, in its most extreme form) called the Bell basis, given by the four vectors

$$|\varphi^\pm\rangle := \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), \qquad |\vartheta^\pm\rangle := \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle), \tag{21}$$

which we can see are not separable. We can also see that $\mathrm{tr}_\mathsf{B}(|\varphi^+\rangle\langle\varphi^+|) = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \mathbb{1}_\mathsf{A}/2$, and similarly for all other reduced states, so that despite the whole being pure, this reduced state is maximally mixed. In other words, despite having full certainty of the global state, we get maximal ignorance in either subpart. This also naturally leads to a relation between the purity and how entangled a bipartite state is: the more entangled, the lower the purity of the reduced states.

We can generalize to bipartite systems of dimension $\mathsf{d} = d_\mathsf{A} = d_\mathsf{B}$ through the state that projects into the vector

$$|\Psi\rangle = \frac{1}{\sqrt{\mathsf{d}}} \sum_{i=1}^{\mathsf{d}} |ii\rangle, \tag{22}$$

which we will generically call *the* maximally entangled state, for which effectively we can readily see that $\mathrm{tr}_\mathsf{A}(|\Psi\rangle\langle\Psi|) = \mathrm{tr}_\mathsf{B}(|\Psi\rangle\langle\Psi|) = \mathbb{1}/\mathsf{d}$. Notice that if the dimensions are different, at most the smaller subsystem can be maximally mixed.

With this we can now introduce an analogue measure of entanglement given by the von Neumann entropy of the reduced states

$$S(\rho_\mathsf{A}) = -\,\mathrm{tr}[\rho_\mathsf{A} \log \rho_\mathsf{A}], \tag{23}$$

and similarly for ρ_B, where log can be taken to be base 2, either known as entanglement entropy. Consider a Schmidt decomposition of a state $|\phi\rangle \in \mathscr{H}_\mathsf{A} \otimes \mathscr{H}_\mathsf{B}$,

$$|\phi\rangle = \sum_{i=1}^{D} \sqrt{\lambda_i} |u_i v_i\rangle, \tag{24}$$

with $|u_i\rangle$, $|v_i\rangle$ orthonormal states in each system, $D = \min(d_\mathsf{A}, d_\mathsf{B})$ and positive coefficients $\sum \lambda_i = 1$. Then, writing $\Phi := |\phi\rangle\langle\phi|$, the reduced state on either system is diagonal, $\Phi_\mathsf{A} = \sum \lambda_i |u_i\rangle\langle u_i|$ and $\Phi_\mathsf{B} = \sum \lambda_i |v_i\rangle\langle v_i|$, thus, as the logarithm of a diagonal matrix is the matrix of logarithms of its entries, it follows that

$$S(\Phi_\mathsf{A}) = -\sum_{i=1}^{D} \lambda_i \log(\lambda_i) = S(\Phi_\mathsf{B}), \tag{25}$$

which is known as the Shannon entropy of the distribution given by the eigenvalues λ_i, and indeed it turns out this is maximized for $\lambda_i = 1/D$, i.e. when the distribution in the smaller subspace is uniform.

CHAPTER 1

The foundations of Statistical Mechanics

If physical theories were people,
thermodynamics would be the village witch.

Goold et al. ([14])

1.1 Equilibrium and the second law of thermodynamics

Thermodynamics—the branch of physics that deals with the different manifestations of energy and the relation between them—is often regarded as a cornerstone in physics, and anyone that aims to gain a deep knowledge about the nature of reality will certainly have to master its concepts. It is, however, a different kind of theory from, say quantum mechanics, in the sense that it pervades all of physics without necessarily being a *fundamental* theory of physical reality in the same way that we regard quantum mechanics as fundamental.

Quantum mechanics indeed can be seen as a fundamental theory of physics at small spatial scales which in the suitable limit will contain the physics at ordinary macroscopic scales. That is, classical physics, and all of its predictions, are in principle attainable from quantum mechanics taken in the appropriate limit. In particular, this means that the laws of thermodynamics should *emerge* from the microscopic physics given by quantum mechanics.

Historically, this was acknowledged since the inception of quantum theory itself by its founding fathers, the first one being perhaps Erwin Schrödinger, who invoked what he calls a *statistical hypothesis* [15] (with English translation in Ref. [16]) about initial energy level populations in two weakly interacting systems. Schrödinger's aim was to describe the long time behaviour of such systems, and he found that the states satisfying his hypothesis are well described by thermal states (satisfy a canonical distribution). Similarly, John von Neumann in Ref. [4] (with translation in Ref. [17]) sets out to explain how the irreversible behaviour of entropy emerges from quantum mechanics and how ensemble properties can be assumed in macroscopic (real and imperfect) physical systems.

Particularly puzzling is the behaviour found in the second law of thermodynamics, where the underlying quantum dynamics should give rise to an ever-increasing entropy and evolve towards

an equilibrium configuration. The dynamics described by quantum mechanics by the Schrödinger equation is unitary, which implies that the information of the system in question is conserved throughout its evolution. This has important consequences for the second law of thermodynamics, as it implies that the dynamics is reversible and shouldn't *forget* its initial configuration when converging to a fixed state.

While the works of Schrödinger and von Neumann made progress in addressing such issues, how the second law emerges from unitary quantum dynamics is far from trivial and remained largely unresolved for years. It has only been recently that such question has come back with renewed wave of interest, given not only major advances in experimental and computational techniques but also other unifying theoretical ideas that put fundamental problems such as this one in a new light and render them tractable [14, 18, 19].

The breadth of topics related to the foundations of statistical mechanics is quite large, with many other areas of physics being affected by them [20–27]; for the purpose of this thesis, however, in this section and then further, we will focus on two main ideas that directly addressed the issues arising from the quantum mechanical foundations of statistical mechanics; these can be described as kinematic and dynamical, and are known as *typicality* and *equilibration*, respectively. The first one refers to an idea that anticipates that almost all quantum systems will be almost in equilibrium, so that most evolutions will carry quantum systems to equilibrium and stay close to it for most times. The second is precisely concerned with the characterization of such evolutions and thus how is equilibrium is achieved. Modern approaches further distinguish as *thermalization* a more restrictive case which can be regarded as a proper thermodynamic equilibration [18], but this is outside the scope of this thesis.

While each of the laws of thermodynamics enjoy their fair share of popularity, the second law is particularly celebrated for its consequences, e.g. that of providing an arrow of time, banning perpetual motion machines, or giving us the conclusion that even our universe will meet death one day. It is one of the most fundamental principles in science and its significance stretches to practically everything we can think of in physical reality. The second law is concerned primarily with the direction in which physical processes can occur: in a nutshell, it is a statement about evolution of physical systems always proceeding towards a fixed state of equilibrium.

The statement of the second law of thermodynamics usually invokes the concept of *entropy*.[1] While in classical thermodynamics it was introduced as a quantity related to the efficiency of thermodynamic processes, its interpretation in statistical mechanics (classical or quantum) relates to how probable a macrostate, i.e. a global property of a system, is. In a nutshell, the second law states that the entropy of a thermally isolated system can only increase, so that systems can only evolve towards the most probable macrostate. While this agrees with our experience of reality, reconciling this phenomenological statement with microscopic reversible laws has been, and continues to be, a challenge.

[1] The word was introduced by Rudolf Clausius in 1865 as a composition of terms standing for *energy* and *transformation*. The modern interpretation of entropy is that of an information measure and physics is rather one particular application; the scope of what can be said about entropy far exceeds what is covered in this thesis.

While the quantitative description of the second law is given by the so-called H-theorem, first derived by Ludwig Boltzmann in 1872 [28] for an ideal gas, it was soon realized that this description was not in itself a proof but a statement that followed as a consequence of an implicit assumption. Such an assumption is known as the *equal a-priori probabilities* assumption, otherwise known as the fundamental postulate of statistical mechanics:

Definition 1.1.1 (Postulate of equal a-priori probabilities [29])**.** A closed system is equally likely to be in any of the microstates accessible to it. We refer to this either as the fundamental postulate of statistical mechanics or the postulate of equal a-priori probabilities.

This is seemingly a perfectly reasonable assumption to make, however, it is still an arbitrary one that needs to be put in by hand. Any successful explanation of the second law purely from quantum mechanics must then account for this fundamental postulate of statistical mechanics.

The second key element that a quantum emergence of the second law must fulfil is that of explaining how is equilibrium reached dynamically, i.e. how is it possible that closed quantum dynamics, as given by the Schrödinger equation, becomes forgetful in the sense of converging to a fixed equilibrium state. This puzzle can be posed simply by saying that the dynamics of closed quantum systems is *unitary*; this implies that dynamics will be recurrent and time-reversal invariant, as explained below. Thus, if quantum systems equilibrate, it must be said precisely in what sense and in which way they do.

In the following sections we present some of the results that have resolved to a great extent both explaining the equal a-priori probabilities postulate as well as the dynamic emergence of equilibrium, and which will be useful further for the main results of this thesis. We highlight throughout that the foundations of statistical mechanics is a vibrantly active topic of research forming part of so-called *quantum thermodynamics*, with interest from all computational, applied and foundational fronts, thus some cutting-edge topics that are also of a high relevance, such as the *eigenstate thermalization hypothesis* [19, 24–27] or equilibration timescales, are not discussed in this thesis.

1.2 Equilibration on average

One of the very first problems that we are faced with when trying to approach the question of convergence towards equilibrium in quantum mechanics is that of recurrences, i.e. that quantum states evolving unitarily eventually return to their initial states. This was proved in Ref. [30], and despite this being the case as well in classical mechanics with so-called Poincaré recurrences, the quantum discussion can be seen to have some differences [31]. For finite dimensional systems, however, such as the ones we will consider here, this picture is intuitively clear since there are finitely many mutually distinguishable states towards which evolution can occur; this can be shown quantitatively as in Ref. [32] and recurrence timescales can be discussed as well [33, 34].

Similarly, time-reversal-invariance is clear from Eq. (8) and Eq. (13). Genuinely equilibrating

closed quantum systems thus cannot exist. A way to make sense of a quantum evolution towards equilibrium is by showing that indeed quantum systems approach an equilibrium state, albeit just staying *close to it for most of the time*, with recurrences being rare. This is in essence what is meant by *equilibration on average*, which we should note, however, is much more generic to what one usually associates with evolution towards *thermal* equilibrium, known as thermalization. Specifically, we use the following general definition.

Definition 1.2.1 (Equilibration on average [18])**.** A time dependent property *equilibrates on average* if its value remains close to a given equilibrium value for most times.

In this chapter we will focus mainly on equilibration on average for the expectation value of observables in closed systems.

1.2.1 The equilibrium state

To make a formal statement about equilibration on average, we first need to define what we mean by an equilibrium value. As equilibration on average assesses how close a time-dependent property is from equilibrium *for most times*, a natural candidate for an equilibrium state is the time-averaged state. Specifically, given an initial state $\rho \in \$(\mathscr{H})$, an observable $A = \mathscr{B}(\mathscr{H})$, and denoting time-averaging by an overline, if the time-average of the expectation value of an observable $\overline{\langle A\rangle_{\rho(t)}} = \overline{\mathrm{tr}[A\rho(t)]}$ equilibrates, then it should do so to $\langle A\rangle_{\overline{\rho(t)}} = \mathrm{tr}[A\overline{\rho(t)}]$. This seems like an obvious observation, but this nevertheless motivates the definition

$$\omega := \lim_{T\to\infty} \overline{\rho}^T, \tag{1.1}$$

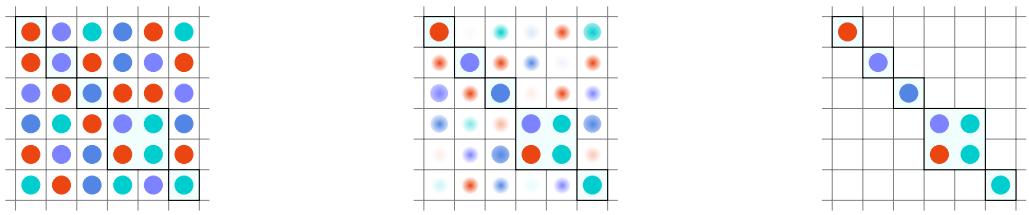
as the equilibrium state of the system, where

$$\overline{\rho}^T := \frac{1}{T}\int_0^T \rho(t)\, dt, \tag{1.2}$$

is the uniform time-average of $\rho(t)$ over a finite-interval $[0, T]$. For quantum systems with a time-independent Hamiltonian, the limit in Eq. (1.1) is well-defined [35] in essence because of the unitary evolution [36], and we get

$$\begin{aligned}
\omega &= \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathrm{e}^{-iHt}\rho\,\mathrm{e}^{iHt}dt \\
&= \lim_{T\to\infty} \frac{1}{T}\int_0^T \sum_{n,m=1}^{\mathfrak{D}} \mathrm{e}^{-it(E_n-E_m)} P_n\,\rho\,P_m\,dt \\
&= \sum_{n,m=1}^{\mathfrak{D}} \left(\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathrm{e}^{-it(E_n-E_m)}dt\right) P_n\,\rho\,P_m \\
&= \sum_{n=1}^{\mathfrak{D}} P_n\,\rho\,P_n,
\end{aligned} \tag{1.3}$$

as $\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathrm{e}^{-it(E_n-E_m)}dt = \delta_{nm}$, where δ_{nm} refers to a Kronecker delta, equal to 1 if $n = m$ and equal to 0 otherwise. That is, the state ω corresponds to the completely dephased state of ρ with respect to the Hamiltonian H. We illustrate this in Fig. 1.1.



$$\rho(0) = \sum_{n,m=1}^{\mathfrak{D}} P_n \, \rho \, P_m \qquad \overline{\rho}^T = \sum_{n,m=1}^{\mathfrak{D}} G^{(T)}_{nm} P_n \, \rho \, P_m \qquad \omega = \sum_{n=1}^{\mathfrak{D}} P_n \, \rho \, P_n$$

Figure 1.1: **(Motivated by Ref. [18].) Dephasing as the mechanism for equilibration:** An initial state $\rho(0) = \rho$ in the energy eigenbasis gets dephased with respect to the Hamiltonian H when averaged over time, $\overline{\rho}^T$, where here we denote $G^{(T)}_{nm} = \overline{\mathrm{e}^{-it(E_n - E_m)}}^T$, ultimately being totally dephased in the infinite time-average limit, corresponding to $\omega = \lim_{T\to\infty} \overline{\rho}^T$.

Remark 1.2.2. In Ref. [35] (and with more detail in Ref. [18]) it is shown that ω is the unique state that maximises the von Neumann entropy among all others with the same expectation on all conserved quantities. That is, for all $\sigma \in \$(\mathscr{H})$ such that $\mathrm{tr}(O_i\sigma) = \mathrm{tr}(O_i\omega)$ on all observables $\{O_i : [O_i, H] = 0\}$, the state ω is the one that maximizes the von Neumann entropy.

Specifically, the von Neumann entropy of a state $\rho \in \$(\mathscr{H})$, defined previously by means of Eq. (23), is a concave function [5], i.e. for any states $\{\sigma_i \in \$(\mathscr{H})\}$ and positive real numbers $\{\lambda_i \in \mathbb{R}^+ : \sum \lambda_i = 1\}$, we have $S\left(\sum \lambda_i \sigma_i\right) \geq \sum \lambda_i S(\sigma_i)$. It then follows from Theorem V.2.1[2] in Ref. [37] that $S(\omega) \geq S(\rho)$, since the dephasing with respect to H is a so-called *pinching*.[3]

Moreover, given any $\sigma_1, \sigma_2 \in \$(\mathscr{H})$, these yield the same eigenvalues on all conserved quantities O_i if and only if $\sum P_n \sigma_1 P_n = \sum P_n \sigma_2 P_n$, and uniqueness can be shown by optimization [18].

This result is remarkable in that one can get a maximum entropy principle[4] purely from unitary quantum dynamics, as opposed to obtaining it from a probabilistic interpretation imposed onto a classical statistical theory. The same principle is satisfied in the respective conditions for a thermodynamic equilibrium with $\rho_{\mathcal{G}} \approx \exp(-\beta H)/\,\mathrm{tr}[\exp(-\beta H)]$ for an inverse temperature β such that $\mathrm{tr}(\rho_{\mathcal{G}} H) = E$ [5]. This maximum entropy principle, however, still does not explain the equal a-priori probabilities postulate or how quantum states evolve towards equilibrium.

1.2.2 Temporal fluctuations

We are interested then in the behaviour for most times of the difference between the time-evolved expected value of an arbitrary observable, $\langle A \rangle_{\rho(t)} = \mathrm{tr}[A\rho(t)]$, and that in equilibrium,

[2] There one has to note that a function f is said to be operator concave if $-f$ is operator convex.

[3] This is the argument given in Ref. [18, 35]; however, it also can be seen to follow from the so-called *data processing inequality* [38], which implies $S(\Phi(\rho)) \geq S(\rho)$ for any unital quantum channel Φ, of which the dephasing map with respect to H is a particular case. These concepts, however, will be introduced until the following chapter.

[4] While L. Boltzmann in Ref. [28] and W. Gibbs in Ref. [39] obtained an ever-increasing entropy result by physical considerations, E.T. Jaynes derived a maximum entropy principle in Ref. [40] for the classical case and in Ref. [41] for the quantum one (which is already conceptually similar to modern ones such as Ref. [42]).

$\langle A\rangle_\omega = \mathrm{tr}[A\omega]$. By definition, the infinite temporal average of the difference will be identically equal to zero. The results in Ref. [43–45] arguably constitute the seminal modern approach to dynamical equilibration; their main focus is on the temporal fluctuations of the expectation value of an observable around equilibrium, i.e. the variance

$$\overline{|\langle A\rangle_{\rho(t)} - \langle A\rangle_\omega|^2} = \overline{|\,\mathrm{tr}[A(\rho(t)-\omega)]|^2}. \tag{1.4}$$

Consider first a system with a non-degenerate Hamiltonian $H = \sum E_n|n\rangle\langle n|$ (so that $\mathfrak{D} = d$) and a pure initial state $\Psi = |\psi\rangle\langle\psi| \in \$(\mathscr{H})$, then we have

$$\begin{aligned}\Psi(t) - \omega &= \sum_{n,m=1}^{d} \mathrm{e}^{-it(E_m-E_n)}\Psi_{mn}|m\rangle\langle n| - \sum_{n=1}^{d}\Psi_{nn}|n\rangle\langle n| \\ &= \sum_{n\neq m}^{d} \mathrm{e}^{-it(E_m-E_n)}\Psi_{mn}|m\rangle\langle n|,\end{aligned} \tag{1.5}$$

where $\Psi_{mn} = \langle m|\psi\rangle\langle\psi|n\rangle$, and so, for any general operator $A \in \mathscr{B}(\mathscr{H})$,

$$\overline{|\langle A\rangle_{\Psi(t)} - \langle A\rangle_\omega|^2} = \overline{\left|\sum_{n\neq m}^{d} \mathrm{e}^{-it(E_m-E_n)}\Psi_{mn}A_{nm}\right|^2} = \sum_{\substack{n\neq m\\ \nu\neq\mu}}^{d} \overline{\mathrm{e}^{-it(E_m-E_\mu-E_n+E_\nu)}}\Psi_{mn}\Psi_{\nu\mu}A_{nm}A^*_{\mu\nu}. \tag{1.6}$$

An additional assumption is now made in Ref. [44, 45], labelled a non-degenerate gap condition, so that no gap between energy levels occur more than once in the energy spectrum. For the average in Eq. (1.6) it implies that

$$\begin{matrix} E_m - E_\mu = E_n - E_\nu \\ m \neq n,\ \mu \neq \nu \end{matrix} \quad \text{iff} \quad \begin{matrix} m = \mu \\ n = \nu \end{matrix}\ , \tag{1.7}$$

so that

$$\overline{\mathrm{e}^{-it(E_m-E_\mu-E_n+E_\nu)}} = \delta_{m\mu}\delta_{n\nu}. \tag{1.8}$$

This condition was already considered by von Neumann [17], who called it a non-resonance condition, and it can be motivated physically when considering a bipartition of the full system to ensure interaction between both parts [18, 44]. Now, $|\Psi_{nm}|^2 = \Psi_{nn}\Psi_{mm}$, hence

$$\begin{aligned}\overline{|\langle A\rangle_{\Psi(t)} - \langle A\rangle_\omega|^2} &= \sum_{m\neq n}^{d} \Psi_{mm}\Psi_{nn}|A_{nm}|^2 \\ &\leq \sum_{m,n}^{d} \Psi_{mm}A_{nm}\Psi_{nn}A^*_{mn} \\ &= \mathrm{tr}\left[\left(\sum_{n=1}^{d}\Psi_{nn}|n\rangle\langle n|A\right)\left(\sum_{m=1}^{d}\Psi_{mm}|m\rangle\langle m|A^\dagger\right)\right] \\ &= \mathrm{tr}[\omega\, A\, \omega\, A^\dagger],\end{aligned} \tag{1.9}$$

where in the second line we included all indices $m = n$. Now we can apply the Cauchy-Schwarz inequality, which reads $\mathrm{tr}[P^\dagger Q]^2 \leq \mathrm{tr}[P^\dagger P]\,\mathrm{tr}[Q^\dagger Q]$ for any $P, Q \in \mathscr{M}_\ell$, with $\mathscr{M}_\ell$ the space of complex $\ell \times \ell$ matrices; this gives

$$\overline{|\langle A\rangle_{\Psi(t)} - \langle A\rangle_\omega|^2} \leq \left(\mathrm{tr}\left[AA^\dagger\omega^2\right]\mathrm{tr}\left[A^\dagger A\omega^2\right]\right)^{1/2}. \tag{1.10}$$

Now we can employ Hölder's inequality, which says that

$$|\,\mathrm{tr}[P^\dagger Q]| \leq \|P\|_a^{1/a}\|Q\|_b^{1/b}, \qquad \text{with} \quad \frac{1}{a}+\frac{1}{b}=1, \tag{1.11}$$

for any $a, b \in \mathbb{R}_0^+$, where $\mathbb{R}_0^+$ denotes the set of non-negative real numbers, and where $\|X\|_p$ is the Schatten p-norm, defined in Eq. (19). We can thus take $a = 1$, $b \to \infty$, so that

$$\overline{|\langle A\rangle_{\Psi(t)} - \langle A\rangle_\omega|^2} \leq \|A\|^2\,\mathrm{tr}(\omega^2), \tag{1.12}$$

as all quantities are positive.

This is already in essence the derivation made in Ref. [45]. The purity of the equilibrium state is given by

$$\mathrm{tr}(\omega^2) = \sum_{n=1}^{d}\langle n|\Psi|n\rangle\langle n|\Psi|n\rangle = \sum_{n=1}^{d}\left(\mathrm{tr}\left[|n\rangle\langle n|\Psi\right]\right)^2, \tag{1.13}$$

which, written as in the second equality, can be directly read as the sum of squares of probabilities for each of the energy eigenstates to be occupied by the initial state. Indeed, if we consider a general degenerate Hamiltonian $H = \sum E_n P_n$ and any given initial state ρ, we can define

$$d_{\mathrm{eff}}^{-1}(\rho) := \sum_{n=1}^{\mathfrak{D}}(\mathrm{tr}[P_n\rho])^2, \tag{1.14}$$

which is the so-called inverse *effective dimension* of the state ρ, also labelled inverse participation ratio [46, 47]. It satisfies $1 \leq d_{\mathrm{eff}} \leq \mathfrak{D} \leq d$, with the lower bound saturated when the initial state is an energy eigenstate and either upper bound saturated when the occupation probability is the same across all eigenspaces or eigenstates. Notice that $d_{\mathrm{eff}}(\rho(t)) = d_{\mathrm{eff}}(\rho)$ is independent of time, and it only corresponds to the purity of ω when either the Hamiltonian is non-degenerate or when the initial state is pure.

The extension of Eq. (1.12) to mixed states when the Hamiltonian is non-degenerate can be made by exploiting the fact that $|\rho_{nm}|^2 \leq \rho_{nn}\rho_{mm}$, which follows from the positivity of ρ.[5] If the Hamiltonian is degenerate, $H = \sum_{n=1}^{\mathfrak{D}} E_n P_n$ with the n^{th} level having degeneracy ℓ, but the initial state is pure, $\Psi = |\psi\rangle\langle\psi|$, we have

$$|\psi(t)\rangle = \sum_{n=1}^{\mathfrak{D}}\sum_{j=1}^{\ell}\mathrm{e}^{-itE_n}|n_j\rangle\langle n_j|\psi\rangle, \tag{1.15}$$

[5] Every principal 2×2 submatrix ϱ_{ij} of ρ is positive semi-definite [48], then for any such matrix, $\det(\varrho_{ij}) = \rho_{ii}\rho_{jj} - |\rho_{ij}|^2 \geq 0$ from positivity of eigenvalues.

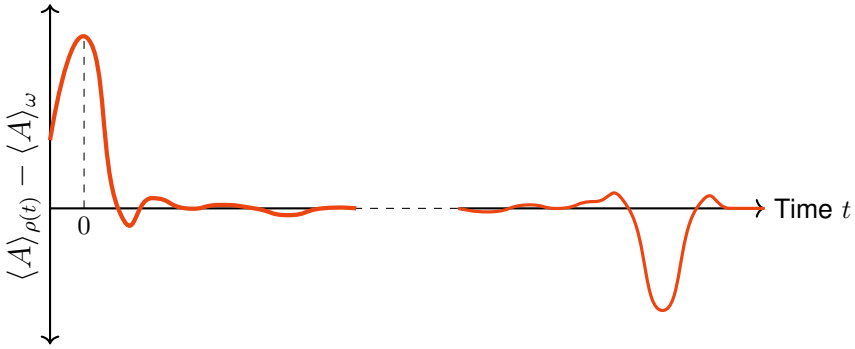


Figure 1.2: **Equilibration on average on expectation values:** An observable A on a space with state $\rho(t)$ equilibrates on average if its expectation value remains close to the one on equilibrium ω for most times; this is guaranteed whenever the fluctuations around equilibrium are suppressed, which occurs whenever the overlap of ρ with the energy eigenstates is large.

so that the eigenbasis $\{|n_j\rangle\}$ for H can be chosen such that the initial state $|\psi\rangle$ has an overlap with only a single eigenstate $|n_{j'}\rangle$ for each energy level E_n. This means $|\psi\rangle$ will evolve as if acted on with a non-degenerate Hamiltonian $H' = \sum_{n=1}^{\mathfrak{D}} E_n |n\rangle\langle n|$. In this case, it is clear that we still have $\mathrm{tr}(\omega^2) = d_{\mathrm{eff}}^{-1}(\Psi)$.

More generally, if we consider both mixed states and degenerate energies, the general form of Eq. (1.12) can be seen to follow by purification [45]. That is, given $\rho \in \$(\mathscr{H})$, take $\Psi \in \$(\mathscr{H}\otimes\mathscr{H})$ evolving under $\tilde{H} = H \otimes \mathbb{1}$, i.e. with the ancilla evolving trivially, thus not affecting the spectrum or degeneracies of the original system. Then we have $\mathrm{tr}[A\rho] = \mathrm{tr}[(A\otimes\mathbb{1}_{\mathscr{H}})\Psi]$, i.e. the expectation values of A on the original system coincide with those of $A\otimes\mathbb{1}$ and also $\|A\otimes\mathbb{1}\| = \|A\|$, i.e. the maximum singular value of A on the original system also coincides with that of $A\otimes\mathbb{1}$. Finally $d_{\mathrm{eff}}(\Psi) = d_{\mathrm{eff}}(\rho)$, as the full trace can be split as a composition of partial traces, i.e. $\mathrm{tr}_{\tilde{\mathscr{H}}} = \mathrm{tr}_{\mathscr{H}_1}\circ\mathrm{tr}_{\mathscr{H}_2} = \mathrm{tr}_{\mathscr{H}_2}\circ\mathrm{tr}_{\mathscr{H}_1}$. Notice, however, that even though $\mathrm{tr}(\omega'^2) = d_{\mathrm{eff}}^{-1}(\rho)$, the purities do not coincide, and in fact $\mathrm{tr}(\omega'^2) \geq \mathrm{tr}(\omega^2)$.

The result in Eq. (1.12) can thus more generally be written as

$$\overline{|\langle A\rangle_{\rho(t)} - \langle A\rangle_\omega|^2} \leq \frac{\|A\|^2}{d_{\mathrm{eff}}(\rho)}, \tag{1.16}$$

which highlights both the role of $d_{\mathrm{eff}}(\rho)$ in determining whether a system will equilibrate or not, as well as that of $\|A\|$ as a scale term for how well the observable can tell between the time-evolving state and equilibrium. That is, $d_{\mathrm{eff}}(\rho)$ tells us if the temporal fluctuations around equilibrium will be suppressed or not, whilst the norm $\|A\|$ can be further restricted to an experimentally reasonable resolution [43]. We can illustrate the whole idea of observable equilibration on average as in Fig. 1.2.

The punchline of Eq. (1.16) is that expectation values of observables in closed systems will equilibrate whenever the total occupation of the energy eigenstates by the initial state is small, or equivalently, when the overlap of the initial state with every single energy eigenstate is large so that Eq. (1.14) is small. That the effective dimension is typically large can be argued for small

subsystems of large whole systems and, in particular, it is normally expected for macroscopic systems just by the strikingly large size these have. Taking into account that a macroscopic system has $\sim \mathcal{O}(10^{23})$ degrees of freedom, it's conceivable that an experimentalist will only be able to prepare a state that overlaps significantly only a very few of them so that the composite will still look quite mixed across all energy eigenstates.

Even when the experimentalist can prepare the initial state with low uncertainty and the levels can be occupied extremely unequally, it can be argued that at realistic scales the effective dimension will remain very large [49]. Notice that Eq. (1.16) applies the same if the observable A acts only on a subpart, System (S), of a larger composite System-Environment (SE) because $\|A_\mathsf{S} \otimes \mathbb{1}_\mathsf{E}\| = \|A_\mathsf{S}\|$ and $\mathrm{tr}[A_\mathsf{S}\,\mathrm{tr}_E(X)] = \mathrm{tr}[(A_\mathsf{S} \otimes \mathbb{1}_\mathsf{E})X]$. Moreover, with the mathematical concepts that will be introduced in the next section, it was shown in Ref. [44] that the effective dimension is typically large, i.e. for any pure state chosen at random from a large subspace of a Hilbert space the probability for the effective dimension to be small is exponentially small in the subspace dimension.

While these arguments make the case for the effective dimension being typically large, it seems clear that in general it would not be efficient to compute such a quantity. In Ref. [50] a feasible way to decide if this is the case is presented for k-local Hamiltonians, i.e. Hamiltonians acting on a lattice of quantum systems with $H = \sum h_i$, with h_i acting on the nearest k sites to the i^{th} one. There it is shown that indeeed the effective dimension is typically large for states with exponentially decaying correlations, i.e. with ρ such that $|\langle AB\rangle_\rho - \langle A\rangle_\rho\langle B\rangle_\rho|$ maximized over operators A, B, decays exponentially in the distance between the support of the sites on which A and B act. Similarly, an equivalence of equilibrium ensembles, i.e. a statement that different macroscopic descriptions of the thermal equilibrium state lead to the same predictions, was shown in Ref. [51–53].

Finally, we can make a statistically relevant statement that fully evokes the definition of equilibration on average for expectation values of observables, such as that in Ref. [43], by making use of Chebyshev's inequality, which states that $\mathbb{P}[|X - \mu| \geq \kappa\,\sigma] \leq \kappa^{-2}$, holding for any $\kappa > 0$ with X a random variable with mean μ and variance σ^2. By picking $\kappa = \|A\| \left(d_{\text{eff}}^{1/3}\sigma\right)^{-1}$ with $\sigma := \overline{|\,\mathrm{tr}[A(\rho(t) - \omega)]|^2}$, together with the bound in Eq. (1.16), this leads to

$$\mathbb{P}_t\left[\left|\langle A\rangle_{\rho(t)} - \langle A\rangle_\omega\right| \geq \frac{\|A\|}{d_{\text{eff}}^{1/3}}\right] \leq \frac{1}{d_{\text{eff}}^{1/3}}, \tag{1.17}$$

which explicitly states that for any time $t \in \mathbb{R}^+$ drawn uniformly at random, the expectation value of A with respect to $\rho(t)$ will be close to that with respect to ω for most times, whenever the effective dimension is large. This also quantitatively captures the notion mentioned previously and sketched in Fig. 1.2 that the departures from equilibrium are rare.

1.2.3 Fluctuations within a finite time

While the previous results show that dynamical equilibration is attained under remarkably mild conditions, two clear questions remain: one relates to the restriction on non-degeneracy of energy gaps and the second regards the timescale on which equilibration occurs. While the non-degenerate gaps condition ensures fully interacting systems regardless of how these are partitioned, it is still a restrictive one that could leave out physically relevant Hamiltonians. The second question is implicit in the previous result, as, while it ensures that equilibration for expectation values *will* take place, it says nothing about *when* it will take place.

One of the most significant steps taken in this direction was that in Ref. [54]: the non-degenerate energy gap condition was relaxed to one quantifying the number of degenerate gaps, and the averaging time window was restricted to be finite. As per Eq. (1.16), this is expected to lead to an upper-bound that gives a correction with respect to the number of degenerate gaps and the width of the time-averaging window. Let us begin by considering again a pure state $\Psi \in \$(\mathscr{H})$, and as per the argument above, a degenerate Hamiltonian $H = \sum E_n P_n$ which, however, has its eigenbasis chosen so that Ψ evolves as if it was doing so under a non-degenerate H. Then we now care about an analogous quantity to the variance above, but now with a uniform time-average within an interval of width T, i.e.

$$\overline{|\langle A\rangle_{\rho(t)} - \langle A\rangle_{\omega}|^2}^{T} = \sum_{\substack{n\neq m\\ \nu\neq\mu}}^{\mathfrak{D}} \overline{\mathrm{e}^{-it(E_m - E_\mu - E_n + E_\nu)}}^{T} \Psi_{mn}\Psi_{\nu\mu}A_{nm}A^*_{\mu\nu}, \tag{1.18}$$

where the finite time-average is defined in Eq. (1.2). Now a way to simplify notation into a more familiar one is to label the energy gaps with $\ell := (m,n)$, $\lambda := (\mu,\nu)$, so that

$$M^{(T)}_{\ell\lambda} := \overline{\mathrm{e}^{it(\mathcal{E}_\ell - \mathcal{E}_\lambda)}}^{T} \qquad \text{where} \qquad \begin{aligned}\mathcal{E}_\ell &= E_m - E_n\\ \mathcal{E}_\lambda &= E_\mu - E_\nu\end{aligned}, \tag{1.19}$$

and $v_\ell := v_{(m,n)} = \Psi_{mn}A_{nm}$, $v_\lambda := v_{(\mu,\nu)} = \Psi_{\mu\nu}A_{\nu\mu}$. Then it becomes clear that

$$\begin{aligned}\overline{|\langle A\rangle_{\rho(t)} - \langle A\rangle_{\omega}|^2}^{T} &= \sum_{\ell,\lambda} v^*_\lambda\, M^{(T)}_{\lambda\ell}\, v_\ell\\ &\leq \|M\|\|v\|_2^2 = \|M\| \sum_{n\neq m}^{\mathfrak{D}} \Psi_{mn}A_{nm}\Psi_{nm}A^*_{mn}\\ &\leq \frac{\|M\|\|A\|^2}{d_{\mathrm{eff}}(\rho)},\end{aligned} \tag{1.20}$$

where the second line is equivalent to $v^\dagger M v \leq \lambda\|v\|_2^2$ for the Hermitian matrix M with components $M^{(T)}_{\lambda\ell}$ and maximum singular value λ, and where in the third line $\|\cdot\|_2$ is the Schatten 2-norm defined in Eq. (19), with all remaining steps following as in Eq. (1.16). This is effectively a correction to Eq. (1.16), which reduces to it when H has no degenerate energy gaps and when the infinite time limit is taken.

The components of M are explicitly given by

$$M_{\ell\lambda}^{(T)} = \begin{cases} 1 & \text{if} \quad \mathcal{E}_\ell = \mathcal{E}_\lambda \\ \dfrac{\exp[iT(\mathcal{E}_\ell - \mathcal{E}_\lambda)] - 1}{iT(\mathcal{E}_\ell - \mathcal{E}_\lambda)} & \text{otherwise} \end{cases}, \tag{1.21}$$

and we can use the hierarchy of Schatten norms in Eq. (20) to get

$$\|M\| \leq \max_\lambda \sum_{\ell=1}^{\mathfrak{D}(\mathfrak{D}-1)} |M_{\ell\lambda}|, \tag{1.22}$$

where from Eq. (1.21) it follows that $|M_{\ell\lambda}| \leq 1$, as all components have modulus less than one. Now the second key definition is to let $N(\varepsilon)$ be the maximum number of energy gaps in any interval of size $\varepsilon > 0$,

$$N(\varepsilon) := \max_E \{\ell : \ell \in \mathscr{E}, \mathcal{E}_\ell \in [E, E+\varepsilon]\}, \tag{1.23}$$

where we defined the set of labels $\mathscr{E} := \{(n,m) : n, m \in \{1, 2, \ldots, \mathfrak{D}\}, n \neq m\}$. The maximum degeneracy of any energy gap is given by $\mathfrak{G}_\mathcal{E} := \lim_{\varepsilon \to 0^+} N(\varepsilon)$, with the non-degenerate case corresponding to $\mathfrak{G}_\mathcal{E} = 1$. This implies that there are at most $N(\varepsilon)$ energy gaps such that

$$(k - 1/2)\varepsilon \leq \mathcal{E}_\ell - \mathcal{E}_\lambda < (k + 1/2)\varepsilon, \tag{1.24}$$

for any non-zero integer k, otherwise just taking $|M_{\ell\lambda}| \leq 1$. Then Eq. (1.22) can be bounded by,

$$\|M\| \leq N(\varepsilon) \left\{ 1 + 2 \sum_{k=1}^{\mathfrak{D}(\mathfrak{D}-1)/2} \frac{2}{T(k - 1/2)\varepsilon} \right\}, \tag{1.25}$$

where the first term comes from the $k = 0$ contribution and the second term uses Eq. (1.24); the sum is maximised by having as many values with small $|k|$ as possible, which gives the factor of 2 at the front of the sum. Now the sum can be further bounded [54] so that

$$\|M\| \leq N(\varepsilon) \left(1 + \frac{8 \log_2 \mathfrak{D}}{\varepsilon T} \right). \tag{1.26}$$

This renders the final result for equilibration on average for expectation values of observables within a finite time interval:

Theorem 1.2.3 (Observable equilibration in finite time [45]). Given a quantum system in state $\rho(t) \in \$(\mathscr{H})$ at time t, evolving via a time-independent Hamiltonian with $\mathfrak{D}$ distinct energies, then for any operator $A \in \mathscr{B}(\mathscr{H})$ and any energy $\varepsilon > 0$ and time $T > 0$,

$$\overline{|\langle A \rangle_{\rho(t)} - \langle A \rangle_\omega|^2}^T \leq \frac{\|A\|^2}{d_{\text{eff}}(\rho)} N(\varepsilon) \left(1 + \frac{8 \log_2 \mathfrak{D}}{\varepsilon T} \right), \tag{1.27}$$

where $\|\cdot\|$ denotes largest singular value, and with inverse effective dimension $d_{\text{eff}}^{-1}(\varrho)$ and $N(\varepsilon)$, the maximum number of energy gaps in an interval or width ε, defined in Eq. (1.14) and in Eq. (1.23), respectively.

It is entirely clear now that Eq. (1.27) reduces to Eq. (1.16) in the non-degenerate energy gaps case, $\mathfrak{G}_{\mathcal{E}} := \lim_{\varepsilon\to 0^+} N(\varepsilon) = 1$ together with the infinite time-window limit $T \to \infty$. Thus the previous conditions on the effective dimension d_{eff} as well as the resolution of A for equilibration to occur remain, while now the correction factor is also required to be small. The first term only requires a low energy gap degeneracy, as opposed to restrict to no gap degeneracy at all, while the second already touches upon one of the most relevant problems for equilibration, namely determining a relevant timescale within which it will occur.

The biggest issue with having equilibration on average in an infinite time-average window is self evident in that equilibration could take the age of the universe to manifest, and while the prediction for its occurrence would still be correct, it would nevertheless be meaningless. By inspecting Eq. (1.27), a time-averaging window of width

$$T \gtrsim \frac{\log_2 \mathfrak{D}}{\varepsilon}, \tag{1.28}$$

will give equilibration provided the remaining quantities in the bound are small, becoming of the order of the infinite-time-average bound in Eq. (1.16). Notice the minimum ε is not an ideal choice: while this parameter can be picked arbitrarily, how large both $N(\varepsilon)$ and the bound in Eq. (1.28) have to be taken into account. This is problematic as it leads to a dependence in the system size and could be at odds with recurrence times [32–34, 55].

While this result offers an insight into the equilibration time scale problem, the question is far from settled as opposed to that about dynamical equilibration. While a general upper bound is much sought after, with promising candidates as in Ref. [56, 57], where it is claimed that for physical observables one such bound is independent of the system size,[6] the problem has so far proved too complex in general and only progress in some classes of systems has been made.[7] Moreover, even performing classical simulations is out of reach to directly approach this, since it has to be done on large systems and long timescales. This is thus one of the most relevant open problems in equilibration to this day, and as a consequence is out of the scope of this thesis.

1.2.4 Trace distance equilibration

The main issue that arises from approaching equilibration on average in terms of expectation values of observables is that, even when the expectation values with respect to two different states can be equal, this does not necessarily mean that observations cannot distinguish the states. This can be made explicit with an example given in Ref. [45], where an observable yields an equal mixture of -1 and $+1$ outcomes with respect to one state and always 0 with respect to another state, so that a measurement of the observable will *always* distinguish them, despite their expectation values being identical. This can be phrased alternatively as saying that, even though an observable might not distinguish a pair of states for most times, this does not imply that for most times it can not distinguish them.

[6] See too, however, the much recent Ref. [58]

[7] A review of equilibration timescales and the classes of systems where results have been obtained can be found in Ref. [59].

These issues can be addressed by considering the distinguishability between two quantum states, which can be phrased in terms of quantifying an operationally relevant distance measure between such pair of states, in our case between a time-evolved state and the equilibrium state. We thus follow Ref. [45, 54] and employ the trace distance D, defined in Eq. (18), between $\rho(t)$ and ω, and so we say that a system equilibrates on average if $\overline{D(\rho(t),\omega)}^T \ll 1$ over a finite time-window of width T. Some of the main results on this front are the ones in Ref. [44, 45, 54] under the very same conditions that we specify in Theorem 1.2.3 previously for observable equilibration; furthermore, the proof for observable equilibration can be used in a straightforward way to obtain

$$\overline{D_{\mathbb{M}}(\rho(t),\,\omega)}^T \leq \frac{\varkappa(\mathbb{M})}{4\sqrt{d_{\text{eff}}}}\sqrt{N(\varepsilon)f(\varepsilon T)}, \tag{1.29}$$

where we define

$$f(\varepsilon T) := 1 + \frac{8\log_2 \mathfrak{D}}{\varepsilon T}, \tag{1.30}$$

and where where $\varkappa(\mathbb{M})$ is the number of possible measurement outcomes in a finite subset of possible measurements $\mathbb{M}$, each with a finite set of outcomes, and the remaining quantities as in Theorem 1.2.3. This result follows specifically because

$$\begin{aligned}\overline{D_{\mathbb{M}}(\rho(t),\,\omega)}^T &\leq \frac{1}{2}\sum_{i,\{\mathrm{M}_i\}\in\mathbb{M}} \overline{|\operatorname{tr}[\mathrm{M}_i(\rho(t)-\omega)]|}^T \\ &\leq \frac{1}{2}\sum_{i,\{\mathrm{M}_i\}\in\mathbb{M}} \sqrt{\overline{|\operatorname{tr}[\mathrm{M}_i(\rho(t)-\omega)]|^2}^T},\end{aligned} \tag{1.31}$$

where in the first line we upper bounded the definition over the maximum with a sum over all POVMs in $\mathbb{M}$ and in the second line we used Jensen's inequality,[8] with the remaining steps as in the derivation for Eq. (1.27), where it is furthermore argued in Ref. [45] that $\sum_{i,\{\mathrm{M}_i\}\in\mathbb{M}} \|\mathrm{M}_i\| \leq \varkappa(\mathbb{M})/2$. The previous conditions for equilibration now remain, with the only difference being the term $\varkappa(\mathbb{M})$, which ought only to be compared with d_{eff}. If we again consider the discussion above for realistic, macroscopic systems with $\mathcal{O}(10^{23})$ degrees of freedom—such that $d \simeq 10^{10^{23}}$—even with e.g. $d_{\text{eff}} \leq d^{1/100}$, the number of possible outcomes $\varkappa(\mathbb{M})$ will remain small compared to the effective dimension, despite it being in itself very possibly a large quantity.

Finally, a bound on the distinguishability with the trace distance was also obtained in Ref. [44, 45, 54] for a subsystem of a larger system-environment composite. Now the interest is on the distinguishability between the reduced states $\rho_{\mathsf{S}}(t) := \operatorname{tr}_{\mathsf{E}}(\rho(t))$ and $\omega_{\mathsf{S}} := \operatorname{tr}_{\mathsf{E}}(\omega)$. The upper bound, with a setup as in Theorem 1.2.3 on a composite Hilbert space $\mathscr{H}_{\mathsf{S}} \otimes \mathscr{H}_{\mathsf{E}}$ reads

$$\overline{D(\rho_{\mathsf{S}}(t),\,\omega_{\mathsf{S}})}^T \leq \frac{d_{\mathsf{S}}}{2\sqrt{d_{\text{eff}}}}\sqrt{N(\varepsilon)f(\varepsilon T)}, \tag{1.32}$$

which, despite being very similar, cannot be obtained from Eq. (1.31) as the number of all possible measurements is infinite. It can, however, be obtained as well from Theorem 1.2.3; the key step in the proof in Ref. [45] is taking an orthonormal basis in S given by d_{S}^2 operators $\{F_i\}$ such that $\rho(t) - \omega = \sum_i \lambda_i(t) F_i$. The explicit form of the operators is reminiscent of a Fourier

[8] $\mathbb{E}[f(X)] \leq f(\mathbb{E}[X])$ with expectation $\mathbb{E}$, a random variable X and a concave function f.

transform and can be seen in Ref. [45]. However, we point out that these are non-Hermitian, so that when Theorem 1.2.3 is applied, it is relevant that it applies to general linear bounded operators $A \in \mathscr{B}(\mathscr{H})$; the other relevant property is that $F_i F_j^\dagger = F_i^\dagger F_j = \delta_{ij} \mathbb{1}_S / d_\mathsf{S}$, so that also $\|F_\mathsf{S}^\dagger \otimes \mathbb{1}_\mathsf{E}\|^2 = 1/d_\mathsf{S}$. With this, we have

$$\begin{aligned}
\overline{\|\rho_\mathsf{S}(t) - \omega_\mathsf{S}\|_1}^T &\leq \sqrt{d_\mathsf{S}}\, \overline{\|\rho_\mathsf{S}(t) - \omega_\mathsf{S}\|_2}^T \\
&\leq \sqrt{d_\mathsf{S}} \sqrt{\sum_{i,j} \overline{\lambda_i(t)\lambda_j^*(t)}^T \operatorname{tr}[F_i F_j^\dagger]} \\
&= \sqrt{d_\mathsf{S}} \sqrt{\sum_i \overline{|\lambda_i(t)|^2}^T} \\
&= \sqrt{d_\mathsf{S}} \sqrt{\sum_i \overline{|\operatorname{tr}\{[\rho(t) - \omega](F_i^\dagger \otimes \mathbb{1}_\mathsf{E})\}|^2}^T},
\end{aligned} \tag{1.33}$$

where in the first line the inequality $\|X\|_1 \leq \sqrt{\dim(X)}\|X\|_2$ was used[9], followed by Jensen's inequality for the square root in the second; the result then follows by applying Theorem 1.2.3 with the operator $A = F_i^\dagger \otimes \mathbb{1}_\mathsf{E}$. The result in Eq. (1.32) implies equilibration in small subsystems S, i.e. whenever we can access only a small subpart of the full system-environment composite with any observable no matter how exotic or unrealistic.

The result in Eq. (1.27) is remarkable in its generality and its reach, as it implies that small subsystems will equilibrate on average within some finite time-window with respect to any measurement we wish to make.

Here we have presented some of the main results for equilibration on average for expectation values and briefly discussed the distinguishability approach; however, a comprehensive discussion of dynamical equilibration with other approaches and further intersections with other topics related to the foundations of statistical mechanics can be found e.g. in Ref. [35, 57, 59, 60].

We now continue to discuss the second big question we posed at the beginning, namely, how the fundamental postulate of statistical mechanics is justified given only the rules of quantum mechanics.

1.3 Typicality

While we came near to touching upon the emergence of the fundamental postulate of statistical mechanics around Eq. (23), with the maximum entropy principle satisfied by the equilibrium state, this has not been fully justified just yet. The approach of equilibrium on average can be classified as a dynamical one, in the sense that it tells us *how* quantum states can evolve towards equilibrium and stay close to it for most times. However, the backbone of the quest to understand equilibration is precisely a notion of equilibrium itself and how it is justified. Furthermore, from concepts like the zeroth law of thermodynamics, the canonical ensemble of statistical mechanics,

[9] Let X a matrix of dimension $n \times n$ and let $\{x_i\}$ be its eigenvalues, then by convexity of the square, $\|X\|_1^2 = n^2(\sum |x_i|/n)^2 \leq n \sum |x_i|^2 = n\|X\|_2^2$.

or the maximum entropy principle, this equilibrium seems to be the most likely state in most cases.

This is precisely the notion that is understood when one speaks of typicality, and it is the kind of notion that Schrödinger and von Neumann had in mind when first trying to reconcile statistical mechanics with quantum mechanics. In Ref. [15] (translated in Ref. [16]), Schödinger studied a pair of systems that are weakly coupled for long times, and found that by assuming a certain proportionality of the population of the initial state in the energy levels with respect to the degeneracy in the non-interacting levels—which he dubbed a *statistical hypothesis*—the reduced states of a small subsystem are well described by a thermal state. Similarly, in Ref. [4] (translated in Ref. [17]), von Neumann presents his *quantum ergodic theorem*, stating in essence that for non-degenerate, non-resonant (i.e. with no degenerate gaps) Hamiltonians, for most decompositions of the Hilbert space[10] and all initial states, for most times the evolving state of the system is macroscopically indistinguishable from a suitable microcanonical state.

This is in the very same spirit as most of the works that approached this question many years later. Some prominent examples are Ref. [62–64]. Notice that even in the work of von Neumann, despite the statement resembling one about equilibration, it is really one about the suitable partitioning of the Hilbert space, holding for most of them under some assumptions about the Hamiltonian but for all initial states. In this thesis we will focus mainly on the result in Ref. [65], which not only addresses the question of typicality but resolves to a large extent the question about the emergence of the fundamental postulate of a-priori probabilities purely from quantum mechanical laws. To do this we first need to specify what does it mean for some property to hold for most quantum states, or even more so, what it means to sample a state at random and seeing that some property for it then holds.

1.3.1 Random states and the Haar measure

To discuss the issue of sampling of quantum states, we should first discuss the geometry of the set of quantum states and how to place a probability measure on it.

As mentioned in Section 0.1, we are concerned here only with studying finite quantum systems which can be associated with the complex Hilbert space $\mathscr{H} \cong \mathbb{C}^d$. A pure state is defined by a vector $|\psi\rangle \in \mathscr{H}$ such that $\langle\psi|\psi\rangle = 1$. Hence it follows that any other vector $|\varphi\rangle$ can be obtained by a unitary transformation $U|\psi\rangle$ where U is a unitary matrix of dimension d, i.e. such that $UU^\dagger = U^\dagger U = \mathbb{1}$. That is, we can define $|\varphi\rangle = U|\psi\rangle$, which is such that $\langle\varphi|\varphi\rangle = 1$, and, similarly, we can see that for the state $\Phi = |\varphi\rangle\langle\varphi|$, Hermiticity and positivity, are satisfied, making Φ a valid pure quantum state. This can be understood intuitively as well if we consider that $\mathbb{C}^d$ is isomorphic to $\mathbb{R}^{2d}$, i.e. any vector $|\psi\rangle \in \mathbb{C}^d$ can be thought of as a real vector in $\mathbb{R}^{2d}$. Furthermore, all of pure states form a $2d-1$ dimensional sphere[11] $\mathbb{S}^{2d-1} = \{v \in \mathbb{R}^{2d} : \|v\|_2 = 1\}$,

[10] Detail and other informal discussion can be seen in Ref. [61].

[11] More precisely, a set of equivalence classes of points of the sphere, as the global phase means that there is a circle of points on the sphere for each distinct pure state.

and so we can think of unitary operators acting on these as rotations. A similar discussion can be made for mixed states e.g. by writing any $\rho \in \$(\mathscr{H})$ as a convex combination of pure states, although the geometrical analogy is not quite the same. Details can be found e.g. in Ref. [7]. For our purposes we only need to have clear that quantum states can always be related by a unitary transformation.

Sampling quantum states *at random* then can be seen to be induced by the sampling of unitaries at random. To do this we need a probability measure on the d dimensional unitary group,[12]

$$\mathbb{U}(d) = \{U \in \mathbb{C}^{d\times d} : UU^\dagger = U^\dagger U = \mathbb{1}\}, \tag{1.34}$$

where $\mathbb{C}^{d\times d}$ denotes complex square matrices of dimension d. Now, a natural way to place a probability measure on the unitary group is to demand that it is an invariant measure in the sense that it should not change under fixed unitary shifts, i.e. it should assign the same measure to all of $\mathbb{U}(d)$.

To begin with, consider the reals: we can measure the size of an interval $\mathbb{I} = [a, b] \subset \mathbb{R}$ of the real line by its length $\ell(\mathbb{I}) = b - a$. Invariance in this case means that $\ell(\mathbb{I} + c) = \ell(\mathbb{I})$ for any $c \in \mathbb{R}$. Now, the length of the interval is simply $\ell(\mathbb{I}) := \int_a^b dx$, so more generally, for any subset $\mathbb{L} \subset \mathbb{R}$ we can measure its size by $\mu_\ell(\mathbb{L}) := \int_{\mathbb{L}} dx$, and we also have

$$\mu_\ell(\mathbb{L} + c) = \int_{\mathbb{L}+c} dx = \int_{\mathbb{L}} d(x + c) = \int_{\mathbb{L}} dx = \mu_\ell(\mathbb{L}), \tag{1.35}$$

and it can be said that μ_ℓ is a translation-invariant measure of the additive group $(\mathbb{R}, +)$.

This property can then be extended to the unitary group $\mathbb{U}(d)$ so that for a given subset $\mathbb{W} \subseteq \mathbb{U}(d)$ and any $V \in \mathbb{U}(d)$ we define

$$\mu_{\mathsf{h}}(\mathbb{W}) = \int_{\mathbb{W}} d\mu_{\mathsf{h}}(U), \tag{1.36}$$

as the Haar measure,[13] μ_{h}, of the set $\mathbb{W}$, which satisfies the left-right invariance property

$$\mu_{\mathsf{h}}(\mathbb{W}) = \mu_{\mathsf{h}}(V\mathbb{W}) = \int_{\mathbb{W}} d\mu_{\mathsf{h}}(VU) = \int_{\mathbb{W}} d\mu_{\mathsf{h}}(UV) = \mu_{\mathsf{h}}(\mathbb{W}V). \tag{1.37}$$

Now what exactly does this mean? Remember the analogy of rotations on real space, then the Haar measure implies that if we measure a "cap" of the sphere and rotate it we get exactly the same measure. Similarly, it means that all unitaries have the same measure, so that it does not matter if we rotate to any other subset of the unitary group, we still get the same answer.

How are we to practically use the Haar measure to actually *measure* the size of a subset of unitaries? It turns out that this can be done relatively easy on a computer [66], as we detail in

[12] A group is a set together with a binary operation satisfying axioms of closure, associativity, identity and invertibility. We take this as a standard concept but present all relevant detail in Appendix B.

[13] The Haar measure is named after Alfréd Haar, who introduced such invariant measure in 1932 more generally over locally compact groups, allowing an analogue of Lebesgue integrals, such as in the cited example over the additive real group.

Appendix C (this was done for the results of Ref. [2], which we discuss in Chapter 4). However, it is an elaborate task in the sense that we need to parametrize the unitary matrix entering $d\mu_{\mathsf{h}}$, and there are d^2 independent real parameters in a unitary matrix,[14] so that

$$d\mu_{\mathsf{h}}(U) = f(\theta_1, \ldots, \theta_{d^2})\, d\theta_1 \cdots d\theta_{d^2}, \tag{1.38}$$

with $\{\theta_1, \ldots, \theta_{d^2}\}$ the parameters of U acting as a set of local real coordinates on the manifold[15] described by $\mathbb{U}(d)$ embedded in $\mathbb{R}^{2d^2}$ and with a given *probability density function*[16] f.

Finally, the Haar measure can be seen to be unique up to a multiplicative constant [68], i.e. for any two Haar measures μ_{h} and μ'_{h}, we can relate these as $\mu_{\mathsf{h}} = \alpha\mu'_{\mathsf{h}}$ for some constant $\alpha \in \mathbb{R}^+$. The Haar measure is finite and positive, so we can make this a legitimate probability measure by normalizing it as $\mu_{\mathsf{h}}(\mathbb{U}) = 1$. In general, this property holds for groups known as *compact* (closed and bounded). However, here we are only concerned with the unitary group, which satisfies such property. This means then that we can say that we *sample a unitary matrix* uniformly at random whenever we take a unitary matrix that is distributed according to the Haar measure, and we denote this by $U \sim \mu_{\mathsf{h}}$.

1.3.2 Twirling and the Schur-Weyl duality

We can now wrap up the previous discussion and connect it with sampling a pure quantum state at random simply by sampling a unitary from the Haar measure. Ultimately what we care about are the statistical properties of random quantum states, or more general quantities depending on unitaries. Specifically, we will rely on the moments of the unitary group to compute statistical properties of quantities relying on compositions of unitaries, such as the purity of a reduced state.

While we will later be concerned with higher order moments of the unitary group with respect to the Haar measure, it is instructive to consider here the calculation of the first and second moments of the unitary group over the Haar measure. For this we can rely on the following.

Theorem 1.3.1 (Schur-Weyl duality [69]). Any operator $\mathcal{O}$ acting on $\mathscr{H}^{\otimes n} := \bigotimes_{i=1}^{n} \mathscr{H}_n$ commutes with all operators $V^{\otimes n}$, where $V \in \mathbb{U}(d)$, if and only if $\mathcal{O}$ is a linear combination of permutation operators:

$$[\mathcal{O}, V^{\otimes n}] = 0, \forall\, V \in \mathbb{U}(d) \qquad \Longleftrightarrow \qquad \mathcal{O} = \sum_{\sigma \in \mathfrak{S}_n} c_\sigma \wp_\sigma, \tag{1.39}$$

with $\mathfrak{S}_n$ denoting the symmetric group over $\{1, 2, \ldots, n\}$ (i.e. the set of permutation operations that can be performed on n symbols), and where the permutation operator is defined as

$$\wp_\sigma |v_1, \ldots, v_n\rangle = |v_{\sigma(1)}, \ldots, v_{\sigma(n)}\rangle, \tag{1.40}$$

for any $|v_1, \ldots, v_n\rangle \in \mathscr{H}^{\otimes n}$.

[14] A complex square matrix of dimension d has $2d^2$ real parameters, and unitarity $UU^\dagger = \sum U_{ij}U^*_{\ell j}|i\rangle\langle\ell| = \mathbb{1}$ imposes $\sum_j U_{ij}U^*_{\ell j} = \delta_{i\ell}$ which gives d^2 real constraints.

[15] A manifold is a topological space for which every point has a neighborhood which is Euclidean.

[16] We define precisely a probability density function in Chapter 3. See Ref. [67] for an explicit construction.

For example, for $n = 3$, there are $3! = 6$ permutations and σ denotes the assignment of each, e.g. the permutation $1 \to 2 \to 1$, $3 \to 3$ corresponds to $\sigma(1) = 2$, $\sigma(2) = 1$, $\sigma(3) = 3$. This can also be denoted in so-called cycle notation simply as $(1, 2)(3)$.

The Schur-Weyl duality is a result from representation theory and it carries deep consequences that go far beyond the results of this thesis; for our purposes, however, it allows us to compute integrals in a so-called *twirled* map, defined as

$$\Xi^{(n)}(X) := \int_{\mathbb{U}(d)} U^{\otimes n} X (U^{\otimes n})^\dagger \, d\mu_{\mathsf{h}}(U), \tag{1.41}$$

where $X \in \mathscr{H}^{\otimes n}$. For now let us simply take this definition for the twirl as a map $\Xi : \mathscr{H}^{\otimes n} \to \mathscr{H}^{\otimes n}$. We will revisit this briefly in Section 2.1.

It is straightforward to see that $\Xi^{(n)}(X)$ commutes with all $V^{\otimes n}$ by the invariant property of the Haar measure,

$$\begin{aligned} V^{\otimes n}\Xi^{(n)}(X) &= \int_{\mathbb{U}(d)} V^{\otimes n} U^{\otimes n} X (V^{\otimes n})^\dagger \, d\mu_{\mathsf{h}}(U) \\ &= \int_{\mathbb{U}(d)} W^{\otimes n} X (W^{\otimes n})^\dagger V^{\otimes n} \, d\mu_{\mathsf{h}}(V^\dagger W) \\ &= \int_{\mathbb{U}(d)} W^{\otimes n} X (W^{\otimes n})^\dagger V^{\otimes n} \, d\mu_{\mathsf{h}}(W) \\ &= \Xi^{(n)}(X) V^{\otimes n}, \end{aligned} \tag{1.42}$$

where we defined $W = VU$ in the second line. Thus, the Schur-Weyl duality ensures that

$$\Xi^{(n)}(X) = \sum_{\sigma \in \mathfrak{S}_n} c_\sigma(X) \, \wp_\sigma, \tag{1.43}$$

where c_σ is now a linear function of X.

1.3.3 Average states and average purity

The decomposition of the twirl map in Eq. (1.43) is quite relevant for our purposes because it lets us swiftly compute the first and second moment with $n = 1, 2$. By the n^{th} statistical moment we are referring to the expectation of products of components of Haar-distributed unitaries, i.e.

$$\mathbb{E}_{\mathsf{h}}[U_{i_1 j_1} \cdots U_{i_n j_n} U^*_{i'_1 j'_1} \cdots U^*_{i'_n j'_n}] = \int_{\mathbb{U}(d)} U_{i_1 j_1} \cdots U_{i_n j_n} U^*_{i'_1 j'_1} \cdots U_{i'_n j'_n} \, d\mu_{\mathsf{h}}(U), \tag{1.44}$$

where U_{ab} are entries of the unitary $U = \sum U_{ab} |a\rangle\langle b|$ in a given basis. Here the notation $\mathbb{E}_{\mathsf{h}}[\cdot] = \int_{\mathbb{U}(d)} [\cdot] \, d\mu_{\mathsf{h}}$ explicitly means the expectation with respect to the Haar measure with $X \sim \mu_{\mathsf{h}}$. Thus computing the n-moments of the unitary group can be rendered equivalent to computing an n-twirl, $\mathbb{E}_{\mathsf{h}}[U^{\otimes n} X (U^{\otimes n})^\dagger] = \Xi^{(n)}(X)$.

For the average we have by definition

$$\mathbb{E}_{\mathsf{h}}[U \, \rho \, U^\dagger] = \int_{\mathbb{U}(d)} U \, \rho \, U^\dagger \, d\mu_{\mathsf{h}}(U), \tag{1.45}$$

so this corresponds to the 1-twirl, $\Xi^{(1)}(\rho)$, which by Eq. (1.43) must imply

$$\mathbb{E}_{\mathsf{h}}[U\,\rho\,U^\dagger] = c(\rho)\,\mathbb{1}, \tag{1.46}$$

because the only possible permutation of one object is to itself and thus the identity is the only possible value for $\wp$. Now, we can exploit both the cyclicity and linearity of the trace, together with the normalization of the Haar measure, to see that

$$\begin{aligned}\operatorname{tr}\{\mathbb{E}_{\mathsf{h}}[U\,\rho\,U^\dagger]\} &= \mu_{\mathsf{h}}(\mathbb{U})\operatorname{tr}(\rho) = 1\\ &= c(\rho)\,\operatorname{tr}(\mathbb{1}) = d\,c(\rho),\end{aligned} \tag{1.47}$$

where in the first line we took the trace of Eq. (1.45), whilst in the second we took that of Eq. (1.46).

Thus, it follows that $c(\rho) = \operatorname{tr}(\rho) = 1$, and hence the average quantum state from the uniform probability measure is the maximally mixed state,

$$\mathbb{E}_{\mathsf{h}}\left[U\,\rho\,U^\dagger\right] = \frac{\mathbb{1}}{d}, \tag{1.48}$$

and it similarly follows, in general, that for any random $d \times d$ matrix X that is Haar-distributed, $\mathbb{E}_{\mathsf{h}}[X] = \operatorname{tr}(X)\mathbb{1}/d$. This result is intuitive since if we take into account that the maximally mixed state is the maximal ignorance state, so for any quantum state drawn uniformly at random, naturally we would expect to get the equiprobable state.

For the second moment we can encounter different relevant quantities, but particularly we will care about the purity $\operatorname{tr}\left[\rho_{\mathsf{A}}^2\right]$ of a reduced state $\rho_{\mathsf{A}} = \operatorname{tr}_{\mathsf{B}}[\rho_{\mathsf{AB}}]$, where $\rho_{\mathsf{AB}} \in \$(\mathscr{H}_{\mathsf{A}} \otimes \mathscr{H}_{\mathsf{B}})$, i.e.

$$\mathbb{E}_{\mathsf{h}}\left\{\operatorname{tr}\left[\left(\rho'_{\mathsf{A}}\right)^2\right]\right\} = \int_{\mathbb{U}(d)} \operatorname{tr}\left\{\operatorname{tr}_{\mathsf{B}}\left[U\,\rho_{\mathsf{AB}}\,U^\dagger\right]\operatorname{tr}_{\mathsf{B}}\left[U\,\rho_{\mathsf{AB}}\,U^\dagger\right]\right\}\,d\mu_{\mathsf{h}}(U), \tag{1.49}$$

where here $\rho'_{\mathsf{A}} = \operatorname{tr}_{\mathsf{B}}[U\,\rho_{\mathsf{AB}}\,U^\dagger]$. This means we can now use the 2-twirl, $\Xi^{(2)}$; towards this, we now have via Schur-Weyl duality

$$\Xi^{(2)}[X] = \alpha(X)\,\mathbb{1} + \beta(X)\,\textsc{swap}, \tag{1.50}$$

where here $\mathbb{1} \in \mathscr{H}^{\otimes 2}$ is a d^2 dimensional identity operator and $\textsc{swap} := \sum |ij\rangle\langle ji|$ is known as a swap operator, which, as the name suggests, swaps the respective states, i.e. $\textsc{swap}\,|uv\rangle = |vu\rangle$, and is such that $\textsc{swap} = \textsc{swap}^\dagger$ and $\textsc{swap}^2 = \mathbb{1}$. This is because the only possible permutations on two elements are precisely the identity and interchanging the elements. Now we need two equations to determine α and β, but this can be done in a similar manner to the 1-twirl: first we have

$$\operatorname{tr}\{\Xi^{(2)}[X]\} = d^2\alpha(X) + d\,\beta(X) = \operatorname{tr}(X), \tag{1.51}$$

where the first equality follows from tracing the Schur-Weyl duality expression, while the second follows from the definition of the twirl. Now, similarly,

$$\operatorname{tr}\{\textsc{swap}\,\Xi^{(2)}[X]\} = d\,\alpha(X) + d^2\,\beta(X) = \operatorname{tr}(\textsc{swap}\,X), \tag{1.52}$$

as SWAP commutes with $U^{\otimes 2}$, so we can solve for α and β,

$$\alpha(X) = \frac{\mathrm{tr}(X)}{d^2-1} - \frac{\mathrm{tr}(\mathrm{SWAP}\, X)}{d(d^2-1)}, \qquad \beta(X) = \frac{\mathrm{tr}(\mathrm{SWAP}\, X)}{d^2-1} - \frac{\mathrm{tr}(X)}{d(d^2-1)}, \tag{1.53}$$

and consequently Eq. (1.50) is now solved.

Going back to the motivation of the purity of a reduced state, first notice that

$$\mathrm{tr}[\mathrm{SWAP}(\rho_\mathsf{A} \otimes \rho_\mathsf{A})] = \sum_{i,j=1}^{d_\mathsf{A}} \langle j|\rho_\mathsf{A}|i\rangle\langle i|\rho_\mathsf{A}|j\rangle = \mathrm{tr}\left[\rho_\mathsf{A}^2\right], \tag{1.54}$$

so that we may readily use the 2-twirl if we take $\mathscr{H} \cong \mathscr{H}_\mathsf{A} \otimes \mathscr{H}_\mathsf{B}$ and $X = \rho_\mathsf{AB}^{\otimes 2}$, then partial trace over the subspace B for each copy,

$$\mathrm{tr}_{\mathsf{BB}}\left\{\Xi^{(2)}\left[\rho_\mathsf{AB}^{\otimes 2}\right]\right\} = \alpha\left(\rho_\mathsf{AB}^{\otimes 2}\right)\, d_\mathsf{B}^2\, \mathbb{1} + \beta\left(\rho_\mathsf{AB}^{\otimes 2}\right)\, d_\mathsf{B}\, \mathrm{SWAP}, \tag{1.55}$$

where implicitly both the identity and the swap act on $\mathscr{H}_\mathsf{A}^{\otimes 2}$. Taking the trace then, we get the expected purity

$$\begin{aligned}
\mathbb{E}_\mathsf{h}\left\{\mathrm{tr}\left[(\rho'_\mathsf{A})^2\right]\right\} &= \mathrm{tr}\left[\mathrm{SWAP}\left(\mathrm{tr}_{\mathsf{BB}}\left\{\Xi^{(2)}\left[\rho_\mathsf{AB}^{\otimes 2}\right]\right\}\right)\right] \\
&= \frac{d_\mathsf{A} d_\mathsf{B}^2}{d_\mathsf{AB}^2-1} - \frac{d_\mathsf{B}\,\mathrm{tr}(\rho_\mathsf{AB}^2)}{d_\mathsf{AB}^2-1} + \frac{d_\mathsf{A}^2 d_\mathsf{B}\,\mathrm{tr}(\rho_\mathsf{AB}^2)}{d_\mathsf{AB}^2-1} - \frac{d_\mathsf{A}}{d_\mathsf{AB}^2-1} \\
&= \frac{d_\mathsf{A}(d_\mathsf{B}^2-1) + d_\mathsf{B}(d_\mathsf{A}^2-1)\,\mathrm{tr}(\rho_\mathsf{AB}^2)}{d_\mathsf{AB}^2-1},
\end{aligned} \tag{1.56}$$

and in particular if the original state is pure, $\mathrm{tr}(\rho_\mathsf{AB}^2) = 1$, this yields

$$\mathbb{E}_\mathsf{h}\left\{\mathrm{tr}\left[(\rho'_\mathsf{A})^2\right]\right\} = \frac{d_\mathsf{A} + d_\mathsf{B}}{d_\mathsf{AB}+1}. \tag{1.57}$$

Moreover notice that

$$\lim_{d_\mathsf{B}\to\infty} \mathbb{E}_\mathsf{h}\left\{\mathrm{tr}\left[(\rho'_\mathsf{A})^2\right]\right\} = \frac{1}{d_\mathsf{A}}, \tag{1.58}$$

i.e. the expected purity of the reduced state ρ_A will approach that of the maximally mixed state whenever $d_\mathsf{B} \gg d_\mathsf{A}$. This implies that, on average, we would have maximum ignorance about the small subsystem despite knowing everything about the state of the full composite. As it turns out, this is the case whenever A and B have the maximum amount of *entanglement* possible. We will introduce precisely what this means below. Bipartite maximally entangled states happen to be typical, in the sense that if we sample a bipartite pure state at random, it will be highly entangled with very high probability. To formalize this notion of typicality let us introduce a concept known as *concentration of measure*, which will let us to go back to the main issue we set out to investigate, which is the emergence of the fundamental postulate of statistical mechanics.

1.3.4 Concentration of measure

While Eq. (1.57) is an interesting result, it does not tell us anything about the distribution of bipartite states and how much their reduced states will differ from the maximally mixed state. A

powerful concept, however, that allows us to dig into the structure of such distribution without having to compute all of the moments of the unitary group is one called concentration of measure.

As the name suggests, concentration of measure refers to a given measure being concentrated in a region of a metric space [70]. More precisely, let $\mathscr{M}$ be a metric space with metric (i.e. distance) $\tilde{\Delta}_{\mathscr{M}}$ and probability measure μ_{M}. Then we say that a function $f : \mathscr{M} \to \mathbb{R}$ satisfies a concentration of measure around its mean if, for any point $x \in \mathscr{M}$ and any $\delta > 0$,

$$\mathbb{P}_{\mathsf{M}}[f(x) \geq \mathbb{E}_{\mathsf{M}}(f) + \delta] \leq \alpha_{\mathsf{M}}(\delta/\mathscr{L}), \tag{1.59}$$

where here, as in the case of the Haar measure, $\mathbb{P}_{\mathsf{M}}$ and $\mathbb{E}_{\mathsf{M}}$ explicitly refer to the probability and expectation with $x \sim \mu_{\mathsf{M}}$, and where $\mathscr{L}$ is the so-called Lipschitz constant of f, and the function α_{M} is known as the concentration function or concentration rate, which must vanish in increasing δ in order to have concentration of measure.

The Lipschitz constant has a special role because, intuitively, it tells us how fast the function f changes, which in turn will have an impact on the concentration rate and thus imply concentration of measure or absence thereof. Here we are concerned with functions mapping to the reals; however, in general, we say that a function $f : \mathbb{X} \to \mathbb{Y}$ between metric spaces $(\mathbb{X}, \tilde{\Delta}_{\mathbb{X}})$ and $(\mathbb{Y}, \tilde{\Delta}_{\mathbb{Y}})$ is $\mathscr{L}$-Lipschitz if there is a real constant $\mathscr{L} \geq 0$ such that

$$\tilde{\Delta}_{\mathbb{Y}}\left(f(x_1), f(x_2)\right) \leq \mathscr{L}\, \tilde{\Delta}_{\mathbb{X}}(x_1, x_2), \tag{1.60}$$

for any $x_1, x_2 \in \mathbb{X}$.

This means that the function in Eq. (1.59) will satisfy a concentration of measure around its mean whenever its rate of change is low, i.e. whenever it has a small Lipschitz constant, given that the concentration rate α_{M} has to be vanishing in increasing δ.

Whilst concentration of measure can be defined as in Eq. (1.59), it is a mathematical concept with a much larger depth [70]. However, as we discussed at the beginning of this section, pure quantum states can be represented as equivalence classes of points in the hypersphere $\mathbb{S}^{2d-1}$, so for our purposes the relevant quantities are the hypersphere as our metric space together with its concentration rate. This result is known as Levy's lemma.[17]

Theorem 1.3.2 (Levy's lemma [65]). Let $f : \mathbb{S}^{d-1} \to \mathbb{R}$ be an $\mathscr{L}$-Lipschitz continuous function on the $d-1$ dimensional hypersphere $\mathbb{S}^{d-1}$, then for any $x \in \mathbb{S}^{d-1}$ and any $\delta > 0$,

$$\mathbb{P}_{\mathsf{h}}[|f(x) - \mathbb{E}_{\mathsf{h}}(f)| \geq \delta] \leq 2\exp\left(-\frac{d\,\delta^2}{9\pi^3\mathscr{L}^2}\right). \tag{1.61}$$

That is, for all functions on a high-dimensional hypersphere that do not change too rapidly, i.e. with a small Lipschitz constant $\mathscr{L}$, the function evaluated on a point picked uniformly at random will be close to its expectation with very high probability. To make this concrete, consider $x = (x_1, \ldots, x_d)$ and the function defined by $f(x) = x_1$, then the expectation is the

[17] Also known as an isoperimetric inequality, it is named after Paul Lévy who derived it in 1919 albeit in a slightly different form.

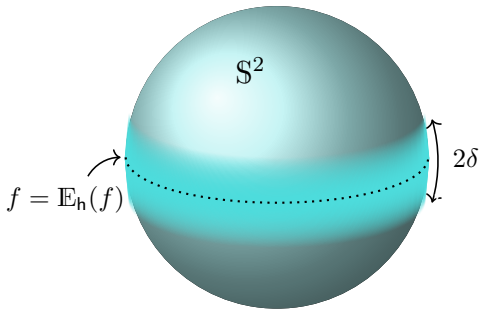


Figure 1.3: **Levy's lemma for the hypersphere** $\mathbb{S}^d$ implies that the probability of picking a point at random outside a band of width 2δ along the equator converges exponentially to zero in $d\delta^2$. For the unit sphere with $d = 2$, concentration is rather weak, however for high-dimensional hyperspheres most random points will lie along the equator with high probability.

equator, $x_1 = 0$, and thus Levy's lemma in Eq. (1.61) tells us that the probability of finding a random point outside a band of width 2δ along the equator converges exponentially to zero in $d\,\delta^2$. Thus *large-dimensional hyperspheres are fat along every equator*, as illustrated in Fig. 1.3.

Let us now go back to the physical scenario. Levy's lemma allows for a neat application of the previous result, letting us show that pure bipartite maximally entangled states are typical, in the sense that most bipartite states are concentrated around them. In other words, if we sample pure bipartite states at random according to the Haar measure, these will be highly entangled with very high probability. Furthermore, as we will see, it is a crucial ingredient for the main result of this section on the emergence of the fundamental postulate of statistical mechanics.

1.3.5 Typicality of entangled states

An average reduced state under the Haar measure will be maximally mixed, as can be seen from Eq. (1.45), as we can always take the partial trace after averaging. However, we can now ask, what is the average distinguishability between a generic reduced state and the maximally mixed state,

$$\mathbb{E}_{\mathsf{h}}\left[D\left(\rho'_{\mathsf{A}}, \frac{\mathbb{1}_{\mathsf{A}}}{d}\right)\right] = \frac{1}{2}\mathbb{E}_{\mathsf{h}}\|\rho'_{\mathsf{A}} - \frac{\mathbb{1}_{\mathsf{A}}}{d_{\mathsf{A}}}\|_1, \tag{1.62}$$

where here $\rho'_{\mathsf{A}} = \mathrm{tr}_{\mathsf{B}}[U\rho_{\mathsf{AB}}U^\dagger]$ with $U \in \mathbb{U}(d_{\mathsf{AB}})$ uniformly distributed, $U \sim \mu_{\mathsf{h}}$. The choice of the trace distance here is motivated as in Sections 0.4 and 1.2.4.

Now we may use both the relation $\|X\|_1 \leq \sqrt{\dim(X)}\|X\|_2$ between Schatten norms and Jensen's inequality for the square root to get

$$\mathbb{E}_{\mathsf{h}}\left[D\left(\rho'_{\mathsf{A}}, \frac{\mathbb{1}_{\mathsf{A}}}{d_{\mathsf{A}}}\right)\right] \leq \frac{1}{2}\sqrt{d_{\mathsf{A}}}\sqrt{\mathbb{E}_{\mathsf{h}}\left[\mathrm{tr}(\rho'^2_{\mathsf{A}})\right] - \frac{1}{d_{\mathsf{A}}}} = \frac{1}{2}\sqrt{\frac{d_{\mathsf{A}}^2 - 1}{d_{\mathsf{AB}} + 1}}, \tag{1.63}$$

where in the second line we used the average purity in Eq. (1.57), assuming that we care about the situation in which the global state ρ_{AB} is pure, $\mathrm{tr}(\rho^2_{\mathsf{AB}}) = 1$. Now, in particular when B is

much greater than A, we get

$$\mathbb{E}_{\mathsf{h}}\left[D\left(\rho'_{\mathsf{A}}, \frac{\mathbb{1}_{\mathsf{A}}}{d_{\mathsf{A}}}\right)\right] \lesssim \frac{1}{2}\sqrt{\frac{d_{\mathsf{A}}}{d_{\mathsf{B}}}}, \tag{1.64}$$

which itself converges to zero in the $d_{\mathsf{B}} \gg d_{\mathsf{A}}$ limit.

Now we can regard this trace distance to the maximally mixed state as a function $D(\mathrm{tr}_{\mathsf{B}}[\cdot], \mathbb{1}_{\mathsf{A}}/d_{\mathsf{A}}) : \mathbb{S}^{2d_{\mathsf{AB}}-1} \to \mathbb{R}^+$ from pure states on the sphere to the reals, and apply Levy's lemma, provided it is Lipschitz continuous. We have, for any two $\rho_{\mathsf{AB}}, \sigma_{\mathsf{AB}} \in \mathbb{S}(\mathscr{H}_{\mathsf{A}} \otimes \mathscr{H}_{\mathsf{B}})$,

$$\left|D\left(\rho_{\mathsf{A}}, \frac{\mathbb{1}_{\mathsf{A}}}{d_{\mathsf{A}}}\right) - D\left(\sigma_{\mathsf{A}}, \frac{\mathbb{1}_{\mathsf{A}}}{d_{\mathsf{A}}}\right)\right| \leq D(\rho_{\mathsf{A}}, \sigma_{\mathsf{A}}) \leq D(\rho_{\mathsf{AB}}, \sigma_{\mathsf{AB}}), \tag{1.65}$$

where in the first inequality we used the triangle inequality and in the second we used the fact that the partial trace does not increase the trace distance.

This last property can be seen by noticing that the difference of density matrices is Hermitian and thus it can be diagonalized with real eigenvalues, i.e. we have $\rho - \sigma = UDU^\dagger$ and we can further split this as $\rho - \sigma = Q - S$ where Q, S are positive semi-definite with orthogonal eigenspaces. Now, because ρ, σ are states, it follows that $\mathrm{tr}(Q) = \mathrm{tr}(S)$, and thus $D(\rho, \sigma) = \mathrm{tr}(Q)$ given that $|\rho - \sigma| = Q + S$. Now, finally $\mathrm{tr}(Q_{\mathsf{A}}) \geq \mathrm{tr}(\Pi\, Q_{\mathsf{A}})$ for any projector Π, so taking the trace distance definition maximizing over projectors, $D(\rho, \sigma) = \max_\Pi \mathrm{tr}[\Pi(\rho - \sigma)]$ it follows that $D(\rho, \sigma) \geq \mathrm{tr}(\Pi\, Q_{\mathsf{B}}) \geq \mathrm{tr}[\Pi(\rho_{\mathsf{A}} - \sigma_{\mathsf{A}})] \geq D(\rho_{\mathsf{A}}, \sigma_{\mathsf{A}})$. This property holds in general for any trace preserving map, but this will be discussed in more detail in Section 2.1.

Thus Eq. (1.65) means that the Lipschitz constant of $D(\mathrm{tr}_{\mathsf{B}}[\cdot], \mathbb{1}_{\mathsf{A}}/d_{\mathsf{A}})$ can simply be taken to $\mathscr{L} = 1$, as for pure states $D(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|)^2 = 1 - |\langle\psi|\phi\rangle|^2$ [6], so that if $\rho_{\mathsf{AB}} = |\phi\rangle\langle\phi|$ and $\sigma_{\mathsf{AB}} = |\psi\rangle\langle\psi|$,

$$\begin{aligned} D(\rho_{\mathsf{AB}}, \sigma_{\mathsf{AB}}) &\leq \sqrt{(1 - |\langle\psi|\phi\rangle|)(1 + |\langle\psi|\phi\rangle|)} \\ &\leq \sqrt{2(1 - \mathrm{Re}[\langle\psi|\phi\rangle])} \\ &\leq ||\phi\rangle - |\psi\rangle|\,, \end{aligned} \tag{1.66}$$

and we now can readily apply Levy's lemma.

Theorem 1.3.3 (Almost all pure quantum states are almost maximally entangled [71]). For any pure state $\rho_{\mathsf{AB}} \in \mathbb{S}(\mathscr{H}_{\mathsf{A}} \otimes \mathscr{H}_{\mathsf{B}})$ of dimension $d_{\mathsf{AB}} = d_{\mathsf{A}} d_{\mathsf{B}}$ with $d_{\mathsf{B}} \gg d_{\mathsf{A}}$, drawn uniformly at random, and for any $\delta > 0$,

$$\mathbb{P}_{\mathsf{h}}\left[D\left(\rho'_{\mathsf{A}}, \frac{\mathbb{1}_{\mathsf{A}}}{d_{\mathsf{A}}}\right) \geq \frac{1}{2}\sqrt{\frac{d_{\mathsf{A}}}{d_{\mathsf{B}}}} + \delta\right] \leq 2\exp\left(-\frac{2d_{\mathsf{AB}}\,\delta^2}{9\pi^3}\right), \tag{1.67}$$

so when B is much bigger than A, the probability of a random reduced state of being distinguishable from the maximally mixed state is vanishing. This implies that almost all pure states will be almost maximally entangled.

A similar result for the average entanglement entropy was derived in Ref. [72], as well as a concentration of measure result in Ref. [73], which can be argued for by using a relation between the entanglement entropy and the trace distance, known as Fannes inequality [6].

Theorem 1.3.3 already hints at implications for the emergence of the fundamental postulate of statistical mechanics; despite the universe being in a pure state, we may find small systems in almost maximal ignorance states, with the reason behind this being precisely entanglement.

1.3.6 Entanglement as a canonical principle

While much progress in topics regarding typicality has been made since von Neumann's results, a turning point in the topic can arguably be attributed to the work of Popescu, Short and Winter in Ref. [65], where the authors explicitly drew a connection between the emergence of the fundamental postulate of statistical mechanics and the typicality of entanglement.

Consider an SE system-environment composite given by the space $\mathscr{H} \cong \mathscr{H}_\mathsf{S} \otimes \mathscr{H}_\mathsf{E}$ as a closed system corresponding to the universe. A global constraint on these, which thermodynamically would correspond to the total energy of the universe, can be modelled in general by restricting the allowed global states to a subspace $\mathscr{H}_\mathsf{R} \subseteq \mathscr{H}$ of dimension d_R. Now the fundamental postulate of statistical mechanics would correspond to a-priori assigning all pure states on R the same probability, i.e. having the equilibrium thermodynamics of the universe under R completely described by the maximally mixed state $\mathbb{1}_\mathsf{R}/d_\mathsf{R}$.

Now the canonical state of the system is defined as the equal a-priori probability state of R with the degrees of the Environment (E) traced out,

$$\Omega_\mathsf{S} = \mathrm{tr}_\mathsf{E}\left(\frac{\mathbb{1}_\mathsf{R}}{d_\mathsf{R}}\right). \tag{1.68}$$

Now the question can be posed in similar terms to that of the typicality of entangled states. The crucial insight in Ref. [65] is that indeed the SE universe might be in a pure state, but despite this, the reduced states in subsystem S will typically be close to the canonical state. Now we consider the average trace distance between an S state ρ_S and the canonical state Ω_S, which we can bound as

$$\mathbb{E}_\mathsf{h}[D\left(\rho'_\mathsf{S}, \Omega_\mathsf{S}\right)] \leq \frac{1}{2}\mathbb{E}_\mathsf{h}[\|\rho'_\mathsf{S} - \Omega_\mathsf{S}\|_2] \leq \frac{1}{2}\sqrt{d_\mathsf{S}}\sqrt{\mathbb{E}[\mathrm{tr}(\rho'^2_\mathsf{S})] - \mathrm{tr}(\Omega^2_\mathsf{S})}, \tag{1.69}$$

where here $\rho'_\mathsf{S} = \mathrm{tr}_\mathsf{E}[U\rho_\mathsf{R}U^\dagger]$ with $U \in \mathbb{U}(d_\mathsf{R})$, and $\rho_\mathsf{R} \in \$(\mathscr{H}_\mathsf{R})$ is a pure state $\mathrm{tr}(\rho_\mathsf{R}) = 1$ of the universe; in the first line we used $\|X\|_1 \leq \sqrt{\dim(X)}\|X\|_2$, while in the second one we used Jensen's inequality together with $\mathbb{E}_\mathsf{h}[\rho'_S] = \Omega_\mathsf{S}$.

We can bound the expectation of the purity of ρ_S by writing, similar to Eq. (1.54),

$$\mathrm{tr}[\rho^2_\mathsf{S}] = \mathrm{tr}[\textsc{swap}(\rho_\mathsf{S} \otimes \rho_\mathsf{S})] = \mathrm{tr}[(\mathbb{1}_\mathsf{EE} \otimes \textsc{swap}_\mathsf{SS})\rho_\mathsf{R} \otimes \rho_\mathsf{R}], \tag{1.70}$$

where SWAP acts on SS and the full trace in the second equality is over RR. Thus we need the 2-twirl

$$\Xi^{(2)}[\rho_\mathsf{R} \otimes \rho_\mathsf{R}] = \frac{1}{d_\mathsf{R}(d_\mathsf{R}+1)}\left(\mathbb{1} + \textsc{swap}\right), \tag{1.71}$$

where we used the Schur-Weyl duality as in Eq. (1.50) with the constants determined by means of Eq. (1.53). The identity and the SWAP are in space $\mathscr{H}_\mathsf{R} \otimes \mathscr{H}_\mathsf{R}$. This 2-twirl is equivalent to

$\mathbb{E}[\rho'_{\mathsf{R}} \otimes \rho'_{\mathsf{R}}]$, where $\rho'_{\mathsf{R}} = U\rho_{\mathsf{R}}U^\dagger$. Thus we have

$$\begin{aligned}\mathbb{E}[\mathrm{tr}(\rho'^2_{\mathsf{S}})] &= \mathrm{tr}[(\mathbb{1}_{\mathsf{EE}} \otimes \mathrm{SWAP}_{\mathsf{SS}})\mathbb{E}_{\mathsf{h}}[\rho_{\mathsf{R}} \otimes \rho_{\mathsf{R}}]] \\ &= \frac{1}{d_{\mathsf{R}}(d_{\mathsf{R}}+1)}\,\mathrm{tr}[(\mathbb{1}_{\mathsf{EE}} \otimes \mathrm{SWAP}_{\mathsf{SS}})(\mathbb{1}_{\mathsf{RR}} + \mathrm{SWAP}_{\mathsf{RR}})],\end{aligned} \tag{1.72}$$

now we notice that $\mathrm{SWAP}_{\mathsf{RR}} = \mathbb{1}_{\mathsf{RR}}(\mathrm{SWAP}_{\mathsf{EE}} \otimes \mathrm{SWAP}_{\mathsf{SS}})$, where the identity $\mathbb{1}_{\mathsf{RR}}$ here simply means that the SE swaps are restricted to R, thus

$$\begin{aligned}\mathbb{E}[\mathrm{tr}(\rho'^2_{\mathsf{S}})] &= \frac{\mathrm{tr}[\mathbb{1}_{\mathsf{RR}}(\mathbb{1}_{\mathsf{EE}} \otimes \mathrm{SWAP}_{\mathsf{SS}})]}{d_{\mathsf{R}}(d_{\mathsf{R}}+1)} + \frac{\mathrm{tr}[\mathbb{1}_{\mathsf{RR}}(\mathrm{SWAP}_{\mathsf{EE}} \otimes \mathbb{1}_{\mathsf{SS}})]}{d_{\mathsf{R}}(d_{\mathsf{R}}+1)}, \\ &\leq \mathrm{tr}\left[\left(\frac{\mathbb{1}_{\mathsf{R}}}{d_{\mathsf{R}}}\right)^{\otimes 2}(\mathbb{1}_{\mathsf{EE}} \otimes \mathrm{SWAP}_{\mathsf{SS}})\right] + \mathrm{tr}\left[\left(\frac{\mathbb{1}_{\mathsf{R}}}{d_{\mathsf{R}}}\right)^{\otimes 2}(\mathrm{SWAP}_{\mathsf{EE}} \otimes \mathbb{1}_{\mathsf{SS}})\right] \\ &= \mathrm{tr}[(\Omega_{\mathsf{S}} \otimes \Omega_{\mathsf{S}})\,\mathrm{SWAP}] + \mathrm{tr}[(\Omega_{\mathsf{E}} \otimes \Omega_{\mathsf{E}})\,\mathrm{SWAP}] \\ &= \mathrm{tr}(\Omega^2_{\mathsf{S}}) + \mathrm{tr}(\Omega^2_{\mathsf{E}}),\end{aligned} \tag{1.73}$$

which neatly renders Eq. (1.69) as

$$\mathbb{E}_{\mathsf{h}}[D\,(\rho'_{\mathsf{S}}, \Omega_{\mathsf{S}})] \leq \frac{1}{2}\sqrt{d_{\mathsf{S}}\,\mathrm{tr}(\Omega^2_{\mathsf{E}})}, \tag{1.74}$$

where $\mathrm{tr}(\Omega^2_{\mathsf{E}}) = d_{\mathsf{R}}^{-2}\,\mathrm{tr}[(\mathrm{tr}_{\mathsf{S}}(\mathbb{1}_{\mathsf{R}}))^2]$ is called an (inverse) effective dimension of the environment, as it measures the dimension of the space in which the environment is most likely to be found. This is also in analogy with the effective dimension introduced in Section 1.2 by means of Eq. (1.14) (there for a probability of occupation of energy eigenstates). This can further be bounded as

$$\mathrm{tr}(\Omega^2_{\mathsf{E}}) \leq d_{\mathsf{S}}/d_{\mathsf{R}}, \tag{1.75}$$

by writing this effective dimension in terms of the eigenvalues λ_i of Ω_{E}, and pulling out the largest one Λ as $\mathrm{tr}(\Omega^2_{\mathsf{E}}) = \sum \lambda_i^2 \leq \Lambda \sum \lambda_i = \Lambda \leq d_{\mathsf{S}}/d_{\mathsf{R}}$, given that $\mathrm{tr}_{\mathsf{S}}(\mathbb{1}_{\mathsf{R}}) = d_{\mathsf{S}}$.

Now, the Lipschitz constant of the trace distance $D(\mathrm{tr}_{\mathsf{E}}[\cdot], \Omega_{\mathsf{S}})$ can be seen to be equal to one, as was done in Eq. (1.65). Thus Levy's lemma can be readily applied to get[18]

Theorem 1.3.4 (General Canonical Principle [65]). For any pure state $\rho_{\mathsf{R}} \in \$(\mathscr{H}_{\mathsf{R}})$ selected uniformly at random, where $\mathscr{H}_{\mathsf{R}} \subseteq \mathscr{H}_{\mathsf{E}} \otimes \mathscr{H}_{\mathsf{S}}$ of dimension d_{R} is a subspace of an SE composite of dimension $d_{\mathsf{ES}} = d_{\mathsf{E}}d_{\mathsf{S}}$, and any $\delta > 0$,

$$\mathbb{P}_{\mathsf{h}}\left[D(\rho_{\mathsf{S}}, \Omega_{\mathsf{S}}) \geq \frac{1}{2}\sqrt{d_{\mathsf{S}}\,\mathrm{tr}(\Omega^2_{\mathsf{E}})} + \delta\right] \leq 2\exp\left(-\frac{2d_{\mathsf{R}}\,\delta^2}{9\,\pi^3}\right), \tag{1.76}$$

where $\Omega_{\mathsf{S}} = \mathrm{tr}_{\mathsf{E}}(\mathbb{1}_{\mathsf{R}})/d_{\mathsf{R}}$ is called the canonical state of the system and with

$$\mathrm{tr}(\Omega^2_{\mathsf{E}}) \leq \frac{d_{\mathsf{S}}}{d_{\mathsf{R}}}, \tag{1.77}$$

satisfied for the inverse dimension of the environment $\mathrm{tr}(\Omega^2_{\mathsf{E}})$.

[18] Notice that in Ref. [65] Levy's lemma is applied to the trace norm as opposed to the trace distance.

This result implies that whenever $d_{\mathsf{S}} \ll 1/\operatorname{tr}(\Omega_{\mathsf{E}}^2)$ and $\delta \ll 1 \ll d_{\mathsf{R}}\delta^2$, most quantum states will be almost canonical with very high probability. The second condition in particular reduces to $d_{\mathsf{R}} \gg 1$ for $\delta = d_{\mathsf{R}}^{-1/3}$. Now, as discussed previously on equilibration in Section 1.2.2, the systems of interest in realistic scenarios are typically much smaller that their environments and in particular we expect $d_{\mathsf{R}} \gg d_{\mathsf{S}}$ as well. This has the implication then that we do not need to invoke the a-priori equal probabilities postulate, but that we rather have typicality of canonical states emerging from their high entanglement with the effective environment. As seen in Ref. [65], when the restriction to the accessible space imposed by R corresponds to the total energy, the canonical state can be seen to correspond to the canonical distribution, or Gibbs so-called state, $\Omega_{\mathsf{S}} \simeq \exp(-\beta H_{\mathsf{S}})/\operatorname{tr}[\exp(-\beta H_{\mathsf{S}})]$ with an inverse temperature β and system Hamiltonian H_{S}, which is further exemplified through the model of a spin chain.

Remark 1.3.5. We notice that this argument can be similarly applied to the expectation value of a given observable $A \in \mathscr{B}(\mathscr{H}_{\mathsf{S}})$. Notice that $\mathbb{E}_{\mathsf{h}}[\operatorname{tr}(A\rho'_{\mathsf{S}})] = \operatorname{tr}[A\Omega_{\mathsf{S}}]$, so we really only need the Lipschitz constant of $\langle A\rangle_{\operatorname{tr}_{\mathsf{E}}(\cdot)} = \operatorname{tr}[A \operatorname{tr}_{\mathsf{E}}(\cdot)]$, which can be obtained similarly as in Eq. (1.66) by letting two $\rho_{\mathsf{R}} = |\phi\rangle\langle\phi|$ and $\rho_{\mathsf{R}} = |\phi\rangle\langle\phi|$ pure states so that

$$\begin{aligned} |\langle A\rangle_{\rho_{\mathsf{R}}-\sigma_{\mathsf{R}}}| &= \frac{1}{2}\,|(\langle\phi| + \langle\psi|)A(|\phi\rangle - |\psi\rangle) + (\langle\phi| - \langle\psi|)A(\langle\phi| + \langle\psi|)| \\ &\leq \|A\|\,||\phi\rangle - |\psi\rangle|\,|\,|\phi\rangle + |\psi\rangle| \\ &\leq 2\|A\|\,||\psi\rangle - |\phi\rangle|\,, \end{aligned} \tag{1.78}$$

and if A only acts in subsystem A we get the same answer given that $\langle A\rangle_{\rho_{\mathsf{S}}} = \langle A\otimes\mathbb{1}\rangle_{\rho_{\mathsf{R}}}$ and $\|A\otimes\mathbb{1}\| = \|A\|$, so we conclude that $\mathscr{L} = 2$, and thus

$$\mathbb{P}_{\mathsf{h}}\left[|\langle A\rangle_{\rho_{\mathsf{R}}-\sigma_{\mathsf{R}}}| \geq \delta\right] \leq 2\exp\left(-\frac{d_{\mathsf{R}}\,\delta^2}{18\pi^3\|A\|^2}\right), \tag{1.79}$$

and here similarly δ can be chosen to a suitable value, e.g. $\delta = d_{\mathsf{R}}^{-1/3}$ so that for a large accessible space, the expectation value of any observable in a random state of the system will be almost the one on the canonical state with high probability.

The approach of typicality is rather complementary to that of dynamical equilibration but it nevertheless provides a further understanding of the quantum foundations of statistical mechanics. One of its features, which can be seen as a drawback, is that it gives a kinematic argument for equilibration (or thermalization) rather than a dynamical one: it just speaks about almost all quantum states looking almost canonical but it does not say anything about how they get there. Another feature of the approach presented here, that will prove challenging in Part II of this thesis, is the drawing of state vectors from the Haar measure: Ref. [35] in Section 6.2 discusses several different approaches with different measures that constrain typicality to a more physically meaningful notion. Finally, it is worth mentioning the approach to equilibration by Ref. [74] employing so-called *unitary designs*, which are distributions of unitary operators reproducing a finite number of moments of the Haar measure. We will formally introduce and employ this concept in Chapter 5.

CHAPTER 2

Quantum Processes

A philosopher once said, "It is necessary for the very existence of science that the same conditions always produce the same results." Well, they don't!

– Richard P. Feynman (*Character of Physical Law*).

In the previous chapter, we already encountered more general transformations of quantum states that are not given simply by the action of an operator. In fact, the whole dynamics, either of the state of a whole Hilbert space (closed system), or of a subspace of a larger composite (open system), can be described as a the action of one of these *superoperators* or *quantum maps*. As we will see, however, when we try and do this quantum map description for an open quantum system composed of a subsystem S and an environment E, we are faced with having to give up an important property known as Complete Positivity, unless we assume that the initial state has no correlations between S and E.

Even if we assume that the initial state has no correlations, the unitary evolution of the whole SE system will lead to a correlated state at a later time t when the system is observed, so that if an experimenter were to intervene in this system at a later time $\tau > t$, they would be faced with the same conundrum as before. A resolution to this can be given by changing the approach to one which considers a dynamical map taking the operations that the experimenter can control, which mathematically are just quantum channels (or more generally, Completely Positive (CP) maps), to output quantum states in what is known as the *superchannel* [75]. As we will see, the superchannel naturally satisfies the complete positivity property and can account for two-time correlations between the initial preparation of the system and a final measurement.

However, as we mentioned, if the experimenter were to intervene on the system a third time, they would require a *bigger* superchannel, i.e. a generalization taking as input more than a couple of interventions and that hence can account for multitime correlations. This is achieved by an object known as the *process tensor*, which has a tensor-like structure in the sense of being a map from multiple CP maps to a quantum state, and similarly satisfies the relevant mathematical properties such as being CP itself.

The process tensor framework then naturally leads to a generalization of the classical notion of stochastic processes as a collection of random variables in time [76, 77]. In particular, this gives a generalization of the concept of Markovianity from classical stochastic processes as that of a dependence in the past to make predictions. While there have been many attempts to do this [78], these have proved inconsistent or insufficient to account for temporal correlations [79]. We will see that the process tensor gives a clear operational Markov condition generalizing the classical one, as well as providing a non-ambiguous measure of non-Markovianity.

Still, assuming no initial correlations and a weak coupling between S and E along evolution is widely applicable and it has been fertile ground for research for many years [8]. Together with Chapter 1, this will bring forward questions that seem almost parallel to those in the foundations of statistical mechanics, namely, why are Markovian processes are so prevalent, when the theory of open quantum systems tells us—as we will see below—that temporal correlations should be the norm, and how does equilibration hold when temporal correlations come into play in multitime processes.

2.1 Quantum channels

The transformations of quantum states that are considered (deterministically) physically realizable are known as quantum channels, and their usage date back from the 1960's with the work of George Sudarshan and collaborators [80, 81] and some years after with Karl Kraus [82].

Quantum channels are linear maps $\Phi : \mathscr{B}(\mathscr{H}_{\mathsf{in}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{out}})$ for any two choice of spaces $\mathscr{H}_{\mathsf{in}}$ and $\mathscr{H}_{\mathsf{out}}$, with the additional conditions of being CP and Trace Preserving (TP), so as to preserve the properties of quantum states. The channel Φ can thus interchangeably be called a Completely Positive Trace Preserving (CPTP) map. In general the input and output spaces can differ; for simplicity and here we will usually assume $\mathscr{H}_{\mathsf{in}} \cong \mathscr{H}_{\mathsf{out}}$, unless stated otherwise, where we distinguish output from input spaces with a prime, e.g. $\mathscr{H}_{\mathsf{A}}$ refers to an input space and $\mathscr{H}_{\mathsf{A}'}$ to an output space. Note also that whenever we refer to a map, we are axiomatically implying throughout this thesis a *linear* map unless stated otherwise.[1]

A map Φ being CP means not only that its action on a positive operator is positive, $\Phi(X) \geq 0$ for any $X \geq 0$, but also that if its domain is only a subspace of a larger space, it will remain positive. That is, let $\mathscr{H} \cong \mathscr{H}_\ell \otimes \mathscr{H}_{\mathsf{in}}$ where $\mathscr{H}_\ell$ is an ℓ-dimensional Hilbert space, and denote by $\mathcal{I}_\ell$ the identity map on $\mathscr{H}_\ell$, i.e. a map acting trivially as $\mathcal{I}_\ell(X) = X$ for any $X \in \mathscr{B}(\mathscr{H})$; then the CP property means

$$(\mathcal{I}_\ell \otimes \Phi)(Y) \geq 0, \tag{2.1}$$

for all positive $Y \in \mathcal{B}(\mathscr{H}_\ell \otimes \mathscr{H}_{\mathsf{in}})$ and all ℓ. In particular if positivity holds only for $\ell \leq \mathrm{L}$, then the map is said to be L-positive. One such example is the map $\Phi_{\mathrm{L}}(X) = (\mathrm{L}-1)\,\mathrm{tr}(X)\mathbb{1}_{\mathrm{L}} - X$ [84]. The CP property is motivated physically to ensure that these map states to states, even when

[1] A discussion can be seen in Ref. [83]; in particular the requirement of linearity does not follow from the linearity of quantum mechanics but rather from linearity of mixing in a statistical theory.

these are correlated with another space. Nevertheless, historically, the CP property hasn't been without contention, in particular for dynamical systems with initial correlations [85–87], where it was argued that in such case either linearity or complete positivity would need to be given up. However as is explained in the following sections, when correctly accounting for correlations in composite systems it is not necessary to give up[2] complete positivity [75, 83, 89, 90].

Secondly, a map Φ is TP whenever $\mathrm{tr}[\Phi(X)] = \mathrm{tr}(X)$ for any $X \in \mathscr{B}(\mathscr{H}_{\mathsf{in}})$, and in particular this means that probabilities are conserved after the action of a TP map on quantum states. As opposed to the CP property, if a quantum map in question fails to be trace preserving but is Trace Non-Increasing (TNI), it remains a physical map, albeit with the interpretation that it can only ever be successfully realised with some probability. There is always a chance that some other transformation could take place. [91].

An example of a quantum channel that is ubiquitous in quantum theory is the partial trace, $\mathrm{tr}_{\mathsf{B}'} : \mathscr{B}(\mathscr{H}_{\mathsf{A}} \otimes \mathscr{H}_{\mathsf{B}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{A}'})$, already introduced in Section 0.1 through Eq. (3). Other standard examples of quantum channels include the identity channel $\mathcal{I}$, with

$$\mathcal{I}(\rho) = \rho, \tag{2.2}$$

the unitary channel $\mathcal{U}$, with

$$\mathcal{U}(\rho) = U\rho U^\dagger, \quad \text{where} \quad UU^\dagger = U^\dagger U = \mathbb{1}, \tag{2.3}$$

and the depolarizing channel Λ_q, with

$$\Lambda_q(\rho) = q\rho + (1-q)\frac{\mathbb{1}_{\mathcal{H}}}{d}, \quad \text{with} \quad 0 \leq q \leq 1. \tag{2.4}$$

Similarly, we have also dealt already with the *dephasing* channel, which as the name suggests, gets rid of the phases of a quantum state, or nullifies the off-diagonal terms of a state with respect to a given basis. Specifically, if we have a system with initial state $\rho \in \$(\mathscr{H})$ and with evolution $U = \exp(-iHt)$ for a Hamiltonian $H = \sum_{n=1}^{\mathfrak{D}} E_n P_n$, with P_n the projector to the eigenspace of energy E_n, then we can write

$$\omega = \mathcal{D}(\rho), \quad \text{where} \quad \mathcal{D}(\cdot) := \sum_{n=1}^{\mathfrak{D}} P_n(\cdot)P_n, \tag{2.5}$$

for the dephasing map $\mathcal{D}$ under the Hamiltonian H. The twirling in Eq. (1.41) as well is another example of a quantum channel.

2.2 Three representations of quantum channels

The definition of a quantum channel can be somewhat abstract and described purely by its inputs and outputs. There are, however, ways to work practically with any quantum channel

[2] While we restrict ourselves to working with CP maps, non-CP maps have also been studied and are not without applications, see e.g. Ref. [88]

through their different representations, some of which can further be extended to any CPTP or Completely Positive Trace Non-Increasing (CPTNI) map or their generalizations. In the following we introduce three of the main representations for quantum channels.

2.2.1 Operator sum representation

The first representation we mention is a decomposition into operators which also gives a condition for the CP property. This representation was first introduced in physics by Karl Kraus in 1971 in Ref. [82] based on work by W. Forrest Stinespring in 1955 [92], and it is thus known as the operator sum representation, the Kraus representation or the Stinespring representation.

Theorem 2.2.1 (Kraus representation [7]). A map $\Phi : \mathscr{B}(\mathscr{H}_{\mathsf{in}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{out}})$ is CP if and only if its action has the form

$$\Phi(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{2.6}$$

where the operators K_i are known as Kraus operators [93].

In particular, if Φ is TP, then

$$\sum_i K_i^\dagger K_i = \mathbb{1}_{\mathsf{in}}, \tag{2.7}$$

whereas if it is such that $\Phi(\mathbb{1}_{\mathsf{in}}) = \mathbb{1}_{\mathsf{out}}$, i.e. so-called *unital*, then

$$\sum_i K_i K_i^\dagger = \mathbb{1}_{\mathsf{out}}. \tag{2.8}$$

The Kraus representation is clearly not unique. The minimal number of Kraus operators is known as the Kraus rank, and satisfies $\kappa \leq d_{\mathsf{in}} d_{\mathsf{out}}$; in particular, there is always a representation with κ orthogonal Kraus operators, i.e. such that $\mathrm{tr}[K_i K_j] = \delta_{ij}$ called the canonical Kraus form.

Example 2.2.2. As a simple example, consider a two-level, qubit system, where any density matrix can be written as $\rho = \frac{1}{2}(\mathbb{1} + \underset{\rightarrow}{r} \cdot \underset{\rightarrow}{\sigma})$, where $\underset{\rightarrow}{r} \in \mathbb{R}^3$ is such that $|\underset{\rightarrow}{r}| \leq 1$, and $\underset{\rightarrow}{\sigma}$ is a vector of the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.9}$$

Then the Kraus operators

$$K_1 = \frac{\sqrt{1+3q}}{2}\mathbb{1}, \quad K_2 = \frac{\sqrt{1-q}}{2}\sigma_x, \quad K_3 = \frac{\sqrt{1-q}}{2}\sigma_y, \quad K_4 = \frac{\sqrt{1-q}}{2}\sigma_z \tag{2.10}$$

describe the action of a qubit depolarizing channel Λ_q in Eq (2.4). This map can equivalently be described as one with an action $\Lambda_Q(\rho) = Q\rho + (1-Q)\sum_i \sigma_i \rho \sigma_i$ through a relation between Q and q [6].



Figure 2.1: **Stinespring dilation of a quantum channel:** The action of a quantum channel $\Phi : \mathscr{B}(\mathscr{H}_{\mathsf{A}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{Aout}})$ on a state ρ can be represented as the reduced state of the outcome of a unitary map $\mathcal{U} : \mathscr{B}(\mathscr{H}_{\mathsf{A}}\mathsf{in} \otimes \mathscr{H}_{\mathsf{B}}\mathsf{in}) \to \mathscr{B}(\mathscr{H}_{\mathsf{A}_{\mathsf{out}}} \otimes \mathscr{H}_{\mathsf{B}_{\mathsf{out}}})$ acting on a joint state $\rho \otimes \beta$. We denote quantum maps by boxes acting on inputs on the left and yielding outputs to the right. Lines denote Hilbert spaces, with no notion of time or direction attached to them; we normally assume $\mathscr{H}_{\mathsf{in}} \cong \mathscr{H}_{\mathsf{out}}$ and omit the output space labels when sufficiently clear.

2.2.2 Dilations

A so-called purification was introduced in Section 0.1 as the fact that every mixed quantum state $\rho \in \mathscr{B}(\mathscr{H}_{\mathsf{A}})$ can be represented by means of a pure quantum state in a larger Hilbert space, $|\psi\rangle\langle\psi| \in \mathscr{B}(\mathscr{H}_{\mathsf{A}} \otimes \mathscr{H}_{\mathsf{B}})$, as the reduced state $\rho = \mathrm{tr}_{\mathsf{B}}(|\psi\rangle\langle\psi|)$. Quantum states, however, can actually be thought of as being a particular case of a quantum channel $\rho : \mathbb{C} \to \mathscr{B}(\mathscr{H})$, so really the idea of purification more generally extends to any quantum channel.

The analogous purification of quantum channels is more precisely referred to as a *Stinespring dilation* and follows from a more general mathematical result known as the Stinespring dilation theorem [92]. In particular, it implies that there exists an ancillary system, say B, such that the action of a CPTP map $\Phi : \mathscr{B}(\mathscr{H}_{\mathsf{A}_{\mathsf{in}}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{A}_{\mathsf{out}}})$ can be written as

$$\Phi(\rho) = \mathrm{tr}_{\mathsf{B}_{\mathsf{out}}}[\mathcal{U}(\rho \otimes \beta)], \tag{2.11}$$

where here $\mathcal{U} : \mathscr{B}(\mathscr{H}_{\mathsf{A}_{\mathsf{in}}} \otimes \mathscr{H}_{\mathsf{B}_{\mathsf{in}}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{A}_{\mathsf{out}}} \otimes \mathscr{H}_{\mathsf{B}_{\mathsf{out}}})$ is a unitary map, and with $\beta \in \mathscr{B}(\mathscr{H}_{\mathsf{B}_{\mathsf{in}}})$ a given quantum state in the ancillary space. Notice $d_{\mathsf{AB}_{\mathsf{in}}} = d_{\mathsf{AB}_{\mathsf{out}}}$, although the individual dimensions may differ. Similarly, the more general Stinespring theorem considers isometries[3] that might not actually be unitary and thus applies in general to CP maps; here we only deal with dilations of the form of Eq. (2.11).

The dilation representation is a powerful one that serves as a cornerstone in open quantum dynamics, as it can be interpreted as all quantum channels arising from a unitary interaction with an environment, as we detail in section 2.3. Notice this implies CPTP maps are precisely the crucial type of maps when it comes to the physical picture; this will be dealt with in depth in the following sections. More generally, the dilation representation is useful when the properties of the global unitary channel can be used to deduce properties of the channel in question, as it can be done with more general maps as in section 2.4; Stinespring theorem assures that a dilation exists for all CP maps, even if it is not unique (more precisely, they can be said to be unique up to a unitary transformation).

[3] An isometry on a Hilbert space is a linear operator that preserves distances; this is equivalent to any linear operator W such that $\|Wv\| = \|v\|$ for all v in such space, with $\|\cdot\|$ a corresponding norm. A unitary operator is a particular case of an isometry.

A practical way to think intuitively about quantum operations is by means of graphical diagrams depicting the relations between maps and states. As we will see, this seemingly simple way of depicting abstract quantities is a handy but powerful tool that allows to deal with more general and complex situations; in essence it is used to represent any quantum computation as a circuit of inputs, quantum channels, which here are the equivalent of logical operations, and outputs. While we do not stick to all the conventions, a thorough description of the graphical calculus often employed in open quantum systems and quantum information can be seen in Ref. [94, 95].

In Fig. 2.1 we depict the dilation of a CP map $\Phi : \mathscr{B}(\mathscr{H}_{\mathsf{A}_{\mathsf{in}}}) \to \mathscr{B}(\mathcal{H}_{\mathsf{A}_{\mathsf{out}}})$ acting on an input state ρ and rendering as outcome a state $\rho_{\mathsf{out}} = \Phi(\rho)$. Conventionally we depict inputs to the left and outputs to the right, with lines representing Hilbert spaces and boxes the quantum maps; we further represent pictorially the partial tracing operation with a trash can (🗑) symbol. Whenever there are parallel lines, these represent a tensor product, as in Fig. 2.1 for $\mathscr{H}_{\mathsf{A}} \otimes \mathscr{H}_{\mathsf{B}}$; in this case the input state is a product state $\rho \otimes \beta$ as well, but in general this need not be the case.

It is instructive to see how the Stinespring dilation relates to the Kraus representation. Let us take Eq. (2.11) assuming that the ancillary space is prepared in a pure state $|\beta\rangle\langle\beta|$; then we can write explicitly the partial trace by introducing a basis $\{|b\rangle\}_{b=1}^{d_{\mathsf{B}_{\mathsf{out}}}}$ for $\mathscr{H}_{\mathsf{B}_{\mathsf{out}}}$ as

$$\begin{aligned}\Phi(\rho) &= \sum_{b=1}^{d_{\mathsf{B}_{\mathsf{out}}}} (\mathbb{1}_{\mathsf{A}_{\mathsf{out}}} \otimes \langle b|)\, U\, (\rho \otimes |\beta\rangle\langle\beta|)\, U^{\dagger} (\mathbb{1}_{\mathsf{A}_{\mathsf{out}}} \otimes |b\rangle) \\ &= \sum_{b=1}^{d_{\mathsf{B}_{\mathsf{out}}}} (\mathbb{1}_{\mathsf{A}_{\mathsf{out}}} \otimes \langle b|)\, U\, (\mathbb{1}_{\mathsf{A}_{\mathsf{in}}} \otimes |\beta\rangle)\, (\rho \otimes \mathbb{1}_{\mathsf{B}_{\mathsf{in}}})\, (\mathbb{1}_{\mathsf{A}_{\mathsf{in}}} \otimes \langle\beta|)\, U^{\dagger} (\mathbb{1}_{\mathsf{A}_{\mathsf{out}}} \otimes |b\rangle) \\ &:= \sum_{b=1}^{d_{\mathsf{B}_{\mathsf{out}}}} K_b\, \rho\, K_b^{\dagger}, \end{aligned} \tag{2.12}$$

where we identified $K_b := (\mathbb{1}_{\mathsf{A}_{\mathsf{out}}} \otimes \langle b|)\, U\, (\mathbb{1}_{\mathsf{A}_{\mathsf{in}}} \otimes |\beta\rangle)$ with the Kraus operators [93]. These clearly satisfy the completeness property in Eq. (2.7). For a canonical Kraus representation this in turn implies a minimal $d_{\mathsf{B}_{\mathsf{out}}} \leq d_{\mathsf{A}_{\mathsf{in}}} d_{\mathsf{A}_{\mathsf{out}}}$. This can then be done similarly if the ancilla is prepared in a mixed state if written as a convex combination [7].

2.2.3 Choi-Jamiołkowski isomorphism

The third main representation that we will deal with in this thesis is based on a (one-to-one) correspondence between quantum channels and quantum states. To begin with, consider a basis $\{|i\rangle\}$ for a d-dimensional Hilbert space $\mathscr{H}$, and define the *vectorization* map by

$$\mathrm{vec}(|i\rangle\langle j|) = |ij\rangle. \tag{2.13}$$

This gives an isomorphism $\mathscr{B}(\mathscr{H}) \cong \mathscr{H} \otimes \mathscr{H}$ in the sense that it is a one to one correspondence between linear bounded operators and vectors. In general we can do this for any operator

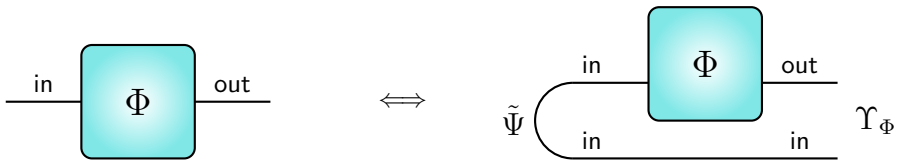


Figure 2.2: **The Choi-Jamiołkowski isomorphism** gives a one-to-one correspondence between quantum maps and states. A map $\Phi : \mathscr{B}(\mathscr{H}_{\mathsf{in}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{out}})$ can be represented as a quantum state $\Upsilon_\Phi \in \mathscr{B}(\mathscr{H}_{\mathsf{out}} \otimes \mathscr{H}_{\mathsf{in}})$, by letting Φ act on half an unnormalized maximally entangled state $\tilde{\Psi} \in \mathscr{B}(\mathscr{H}_{\mathsf{in}} \otimes \mathscr{H}_{\mathsf{in}})$, which is represented by an arch joining the entangled systems.

$A \in \mathscr{B}(\mathscr{H})$ in the basis above with components $\alpha_{ij} = \langle i|A|j\rangle$ as

$$|\phi_A\rangle := \mathrm{vec}(A) = \sum_{i,j=1}^{d} \alpha_{ij}|ij\rangle, \tag{2.14}$$

or similarly for a vector $|\varphi\rangle \in \mathscr{H} \otimes \mathscr{H}$, we can write this in a given basis and turn one of the basis vectors in a covector. Now, how exactly do we get this correspondence? Notice that we may write

$$\begin{aligned} |\phi_A\rangle &= \sum_{i,j=1}^{d} \alpha_{ij}|i\rangle \otimes |j\rangle = \sum_{i,j,k=1}^{d} \alpha_{ij}|k\rangle \otimes |i\rangle\langle j|k\rangle = \sum_{k=1}^{d} |k\rangle \otimes A|k\rangle \\ &= (\mathbb{1} \otimes A) \sum_{i=1}^{d} |ii\rangle, \end{aligned} \tag{2.15}$$

where the state $|\tilde{\psi}\rangle := \sum |ii\rangle$ is precisely the maximally entangled state in Eq. (22) up to a normalization factor. We will usually refer to this as an unnormalized maximally entangled state. It also does not matter if we act with A on the first or the second Hilbert space. This means we can write $|\phi_A\rangle = (\mathbb{1} \otimes A)|\tilde{\psi}\rangle = (A \otimes \mathbb{1})|\tilde{\psi}\rangle$.

Let us label the pair of copies of the spaces with 1 and 2, so that $|\phi_A\rangle \in \mathscr{H}_1 \otimes \mathscr{H}_2$. We can go the opposite way via

$$A = \mathrm{tr}_2\left(|\phi_{A^{\mathrm{T}}}\rangle\langle\tilde{\psi}|\right) = \mathrm{tr}_2\left[\left(\mathbb{1}_1 \otimes A^{\mathrm{T}}\right)|\tilde{\psi}\rangle\langle\tilde{\psi}|\right], \tag{2.16}$$

where T denotes a transpose, i.e. $A^{\mathrm{T}} = \sum \alpha_{ji}|i\rangle\langle j|$.

So what we expect now is to have an analogous correspondence between quantum channels and density matrices; and indeed, if we let

$$\tilde{\Psi} := \sum_{i,j=1}^{d_{\mathsf{A}}} |ii\rangle\langle jj| \in \mathscr{B}(\mathscr{H}_{\mathsf{in}} \otimes \mathscr{H}_{\mathsf{in}}), \tag{2.17}$$

and $\Phi : \mathscr{B}(\mathscr{H}_{\mathsf{in}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{out}})$ a quantum channel, we have

$$\Upsilon_\Phi = (\Phi \otimes \mathcal{I})\tilde{\Psi}, \tag{2.18}$$

and we can recover the action of the map Φ through

$$\Phi(\rho) = \mathrm{tr}_{\mathsf{in}}[\Upsilon_\Phi(\mathbb{1}_{\mathsf{out}} \otimes \rho^{\mathrm{T}})], \tag{2.19}$$

for any $\rho \in \$(\mathscr{H}_{\mathsf{in}})$, and where the partial trace is over the input space. The matrix in Eq. (2.18) defines what is referred to as the Choi-Jamiołkowski isomorphism (CJI) [96, 97], and we will refer to Υ_Φ as the Choi state of Φ. It can also be visualized diagrammatically as in Fig. 2.2.

The CJI, however, as opposed to the Kraus representation or the dilation representation, not only applies to quantum channels or CP maps but generally to any linear map. Here we will consider primarily CP maps, for which Eq. (2.18) implies $\Upsilon_\Phi \geq 0$; indeed a positive Choi state can be decomposed as $\Upsilon_\Phi = \sum_{i=1}^{\mathsf{D}} g_i |\gamma_i\rangle\langle\gamma_i|$ with its $\mathsf{D} = d_{\mathsf{in}} d_{\mathsf{out}}$ nonnegative eigenvalues g_i and eigenvectors $\{|\gamma_i\rangle\}$, then

$$\begin{aligned}
\Phi(X) &= \sum_{i=1}^{\mathsf{D}} g_i \operatorname{tr}_{\mathsf{in}}[|\gamma_i\rangle\langle\gamma_i|(\mathbb{1}_{\mathsf{out}} \otimes X^{\mathrm{T}})] \\
&= \sum_{i=1}^{\mathsf{D}} \sum_{e=1}^{d_{\mathsf{in}}} g_i (\mathbb{1}_{\mathsf{out}} \otimes \langle e_{\mathsf{in}}|)|\gamma_i\rangle\langle\gamma_i|(\mathbb{1}_{\mathsf{out}} \otimes X^{\mathrm{T}}|e_{\mathsf{in}}\rangle) \\
&= \sum_{i=1}^{\mathsf{D}} \sum_{e,\varepsilon=1}^{d_{\mathsf{in}}} g_i (\mathbb{1}_{\mathsf{out}} \otimes \langle e_{\mathsf{in}}|)|\gamma_i\rangle\langle\gamma_i|(\mathbb{1}_{\mathsf{out}} \otimes |\varepsilon_{\mathsf{in}}\rangle)\langle\varepsilon_{\mathsf{in}}|X^{\mathrm{T}}|e_{\mathsf{in}}\rangle \\
&= \sum_{i=1}^{\mathsf{D}} \sum_{e,\varepsilon=1}^{d_{\mathsf{in}}} g_i (\mathbb{1}_{\mathsf{out}} \otimes \langle e_{\mathsf{in}}|)|\gamma_i\rangle\langle e_{\mathsf{in}}|X|\varepsilon_{\mathsf{in}}\rangle\langle\gamma_i|(\mathbb{1}_{\mathsf{out}} \otimes |\varepsilon_{\mathsf{in}}\rangle) \\
&:= \sum_{i=1}^{\mathsf{D}} \mathfrak{G}_i X \mathfrak{G}_i^\dagger,
\end{aligned} \tag{2.20}$$

which gives a Kraus representation with $\mathfrak{G}_i := \sum_{e=1}^{d_{\mathsf{in}}} \sqrt{g_i}(\mathbb{1}_{\mathsf{out}} \otimes \langle e_{\mathsf{in}}|)|\gamma_i\rangle\langle e_{\mathsf{in}}|$ each of the Kraus operators.

If additionally the map is TP as well, we have

$$\operatorname{tr}_{\mathsf{out}}[\Upsilon_\Phi] = \sum_{i,j=1}^{d_{\mathsf{in}}} \operatorname{tr}[\Phi(|i\rangle\langle j|)]|i\rangle\langle j| = \mathbb{1}_{\mathsf{in}}, \tag{2.21}$$

which makes evident as well that the Choi *state* does not have unit trace by construction.

The punchline of the CJI is that linear maps can be represented as matrices in a larger space, and in particular that quantum channels have corresponding positive matrices, also in a larger space. Two straightforward examples come for the identity map, which leads simply to $\Upsilon_{\mathcal{I}} = \tilde{\Psi}$ and the unitary map, which leads also to a maximally entangled state $\Upsilon_{\mathcal{U}} = (\mathcal{U} \otimes \mathcal{I})\tilde{\Psi}$ between the input and output states.

The Choi state in general will be a relevant tool in this thesis, albeit for more general maps, as we will see below.

2.3 Open quantum dynamics

Let us now go back to the physical motivation and consider a generic scenario where an experimenter has access to the subpart S of a larger composite SE, where an environment

E is out of access and control to the experimenter. The whole SE universe is closed so in general it will evolve unitarily. The system S, however, from the perspective of the experimenter, will be described by a CP map taking preparations and yielding outcome states, which suggests a Stinespring representation as in Eq. (2.11).

The standard experimental procedures to reconstruct quantum states and quantum dynamical maps are known as quantum state tomography (QST) and quantum process tomography (QPT), respectively. QST relies on the measurement statistics, $\vec{p} = (\mathrm{tr}[\mathrm{M}_1\rho], \ldots, \mathrm{tr}[\mathrm{M}_\ell\rho])$, of a given measurement $\{\mathrm{M}_i\}$ (which has to form a basis on the space of the system) to reconstruct ρ. QPT, on the other hand, aims to reconstruct the dynamical CP map, Φ, by preparing a set of linearly independent input states $\{\rho_i\}$, sending them through the map, and then collecting the outputs as a linear combination of the inputs, which then allows to reconstruct a pair of Kraus operators of the channel. We will not deal in detail with either procedure but it lets us put into perspective what the real world situation is; more on either can be seen e.g. in Ref. [6].

Let us consider then the following scenario: at some time, which we may set as $t = 0$, the system S is uncorrelated from E so that the experimenter is able to prepare inputs $\rho_{\mathsf{S}}(0)$ and reconstruct the final states by QST at some time $t = \delta t$. We will only consider unitary evolution of the whole under a time-independent Hamiltonian H, so that

$$\mathcal{U}_{\delta t}(\cdot) = \exp(-iH\delta t)(\cdot)\exp(iH\delta t), \tag{2.22}$$

is the unitary map of the evolution during δt. Letting the environment state at $t = 0$ be ε, we have

$$\rho_{\mathsf{S}}(t) = \mathscr{Z}_{\delta t}[\rho_{\mathsf{S}}(0)] = \mathrm{tr}_{\mathsf{E}}[\mathcal{U}_{\delta t}(\rho_{\mathsf{S}}(0)\otimes\varepsilon)], \tag{2.23}$$

where here $\mathscr{Z}_{\delta t} : \$(\mathscr{H}_{\mathsf{S}}) \to \$(\mathscr{H}_{\mathsf{S}})$ is the dynamical map taking S states to S states.

We can allow time to vary for the dynamical map so that $\{\mathscr{Z}_t : t \geq 0\}$ constitutes a one-parameter family of dynamical maps with $\mathscr{Z}_0 = \mathcal{I}$; we can furthermore see that these are CPTP, as we did for the Stinespring dilation in Eq. (2.12). Now, experimentally, QPT can be used to reconstruct the dynamical map.

2.3.1 The Born-Markov approximation

The reconstruction of the dynamical maps $\mathscr{Z}_t$ can be done for numerous physical cases and phenomenological models [8]. Despite this, even ignoring the fact that it is very difficult to prepare a system that is uncoupled to the environment [98], computing the dynamical map is typically unmanageable without making some further simplifying assumptions.

Notice that even if the state of the whole composite, ρ_{SE}, obeys the Schrödinger equation, this does not immediately imply that an analogue differential equation, i.e. a time-local master equation, for ρ_{S} exists. In particular, if ρ_{S} obeys a differential equation, this means that the state of S is local in time [99], i.e. to determine $\rho_{\mathsf{S}}(t+dt)$ we would only need to know $\rho_{\mathsf{S}}(t)$. However, the interaction with the environment makes this generically impossible, as information will be exchanged between S and E and we would need to know the state ρ_{S} at earlier times as well.

The first notion of locality in time is what is commonly understood in open quantum dynamics as *Markovianity*. We will expand on this notion below, however, the general concept is the same: that of *memorylessness*, as opposed to *non-Markovianity*, in which we need to know the past states of a given system to determine its future. Crucially, notice once again that open quantum systems theory is non-Markovian *by definition*, and that Markovian open quantum dynamics are in reality impossible.

Despite this, as is common when doing idealizations in all of physics, approximating open systems as Markovian is effective for a large class of physical scenarios and has a wide applicability [100–103]. The simplest assumptions that can be made to render the dynamics Markovian are known as the Born-Markov approximation, which lead to an important class, albeit the simplest one, of open quantum dynamics and master equations.

Definition 2.3.1 (Born-Markov approximation)**.** These are two assumptions:

i. Weak coupling (Born): The coupling between S and E is sufficiently weak and E is reasonably large, so that ε is practically unaffected and the whole SE state remains approximately in product state at all times: $\rho_{\mathsf{SE}}(t) \approx \rho_{\mathsf{S}}(t) \otimes \varepsilon$, $\forall t \geq 0$.

ii. Forgetful environment (Markov): The self-correlations within E induced by the interaction with S decay rapidly compared to the timescale over which ρ_{S} changes noticeably.

Beginning from the Schrödinger equation, Eq. (8), for a total Hamiltonian taking into account the SE interaction, a master equation by the name of Lindblad-Franke-GKS equation[4] can be derived by assuming the Born-Markov approximation [8]. While the Born-Markov conditions are the underlying physical conditions to render the Hamiltonian dynamics Markovian, the Lindblad-Franke-GKS equation in a general form can be obtained by positing that the dynamical map satisfies

$$\mathscr{Z}_{t_1+t_2} = \mathscr{Z}_{t_1}\mathscr{Z}_{t_2}, \tag{2.24}$$

where $\mathscr{Z}_{t_1}\mathscr{Z}_{t_2} := \mathscr{Z}_{t_1} \circ \mathscr{Z}_{t_2}$ implicitly means a composition of maps. This is known as the divisibility, or semigroup, property, and it is usually employed in the definition of quantum Markovianity, as it resembles a property known by the same name in the theory of classical stochastic processes. We will come back to this issue below, as it will be important for the discussion in this thesis, the point about this property now is that it allows to write $\mathscr{Z}_t = \exp(\mathcal{L}\, t)$, with $\mathcal{L}$ known as the generator of the semigroup, which then can be written in its most general form [8], and which in particular can be derived explicitly using the Born-Markov approximation.

The assumptions in the Born-Markov approximation can be boiled down to positing that the environment is negligibly affected by its interaction with the system and that it is forgetful or dissipative. Both assumptions are widely applicable in physical settings, the Born assumption for example could be taken considered for macroscopic systems or on systems within very large

[4] GKS standing after Gorini, Kossakowski and Sudarshan. The equation is sometimes only referred to after Lindblad. The contributions are: Franke [104], GKS [105] and Lindblad [106] (all from 1976).

environments and the Markov condition could be used if one considers an environment at a high enough temperature.

This already begs the question of how Markovian behaviour comes about with no a-priori assumptions. As expected from the discussion in Chapter 1, we can make a very clear analogy between this question and that of the foundations of statistical mechanics. To do so, let us continue analysing the problem of system-environment correlations in time and how to resolve them.

2.3.2 The initial correlation problem

Let us embrace now the inevitability of the coupling of system S to the environment E, i.e. in a standard experimental setup, the experimenter's preparation of input states might affect the environment as well. The reconstruction of dynamical maps by QPT was suggested in the late nineties [107] and soon after experiments began to be carried out [108–116]. However, non completely positive behaviour was being obtained, which was discarded as experimental noise and hence just not physically valid. On the other hand, despite not having access to the whole SE composite, experiments were also devised to detect initial correlations [117, 118].

As it turns out, in 1994 and 95, Philip Pechukas and Robert Alicki showed that if initial correlations are to be considered, then one either has to give up complete positivity or linearity on the dynamical map [85–87]. A simple argument based on the original by Pechukas and Alicki can be seen in Ref. [83], where a map labelled an *assignment map*, $\zeta : \mathscr{B}(\mathscr{H}_\mathsf{S}) \to \mathscr{B}(\mathscr{H}_\mathsf{E} \otimes \mathscr{H}_\mathsf{S})$, such that it satisfies a *consistency condition* $\mathrm{tr}_\mathsf{E}[\zeta(\rho_\mathsf{S})] = \mathrm{tr}_\mathsf{E}(\rho_\mathsf{SE}) = \rho_\mathsf{S}$ is used to write

$$\mathscr{Z}_t[\rho_\mathsf{S}] = \mathrm{tr}_\mathsf{E}[\mathcal{U}_t\, \zeta(\rho_\mathsf{S})] = \mathrm{tr}_\mathsf{E}[\mathcal{U}_t(\rho_\mathsf{SE})], \tag{2.25}$$

where here $\mathcal{U}_t$ is a unitary map as defined in Eq. (2.22) and where ρ_S and ρ_SE stands for the initial state on either S or SE. Then demanding the CP property on $\mathscr{Z}_t$ implies (as can be seen in detail in Ref. [83]) that $\zeta(\rho_\mathsf{S}) = \rho_\mathsf{S} \otimes \varepsilon$, i.e. the initial state is uncorrelated, or otherwise either the CP property or linearity have to be given up. This conundrum is known as the initial correlation problem.

Now, while the CP property is desirable, it seems tempting to try and go around it, as opposed to try and do the same with linearity, which as we mentioned at the beginning of this chapter is related to the mixing property of any statistical theory, and furthermore, it is a pillar assumption for QPT. In particular, simple examples of Not Completely Positive (NCP) behaviour, i.e. maps that are positive only for some subset $\{\rho_\mathsf{S} : \zeta(\rho_\mathsf{S}) \geq 0\}$ of *compatible* states, can be seen to arise whenever the system and the environment are entangled [119], and more generally for other types of correlations [120, 121]. This led to several arguments for embracing NCP maps [122–125] at the beginning of the current century; the main problem, however, is operationally determining the set of compatible states, i.e. an experimenter does not necessarily know which is this set and determining it experimentally leads to ambiguity, since the very act of trying to identify a

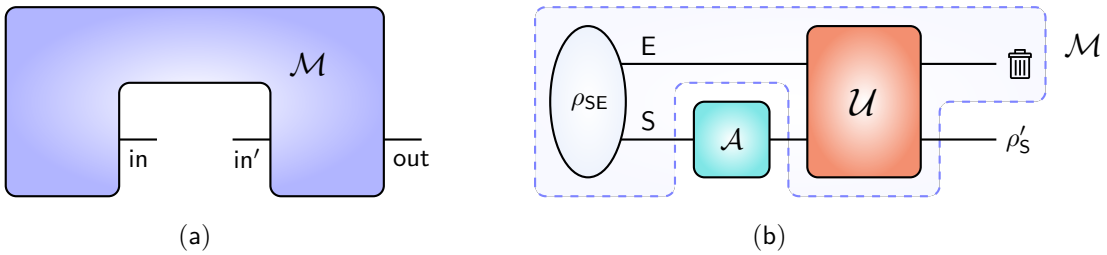


Figure 2.3: **Circuit diagram of the superchannel:** (a) The superchannel $\mathcal{M}$ can be thought of as a two-legged box taking as input a CPTNI map $\mathcal{A} : \mathscr{B}(\mathscr{H}_{\mathsf{in}}) \to \mathscr{B}(\mathscr{H}_{\mathsf{in}'})$ and yielding an output state $\rho'_{\mathsf{S}} \in \$(\mathscr{H}_{\mathsf{out}})$. (b) An initial state ρ_{SE} on the whole SE system, in general correlated, is acted on with a preparation described by $\mathcal{A}$, then the whole evolves unitarily with the unitary channel $\mathcal{U}$, yielding the state $\mathcal{M}[\mathcal{A}] = \rho'_{\mathsf{S}}$. The dashed box displays the contents of $\mathcal{M}$ explicitly as all dynamical content out of the control of the experimenter.

correlated system will disturb the environment [126–128]. A simple example with only two qubits can be seen in detail in Ref. [83].

2.3.3 The superchannel

A resolution to the initial correlation problem comes rather with a change of perspective. As Asher Peres puts it [129]: "*The simple and obvious truth is that quantum phenomena do not occur in a Hilbert space. They occur in a laboratory. (. . .) The experimenter controls the emission process and observes detection events. The theorist's problem is to predict the probability of response of this or that detector, for a given emission procedure*". The ambiguity in the NCP scenario in a sense already gives us a hint: in any experiment the very first step is to prepare some unknown fiducial state into a known input. But if the fiducial state is correlated, the experimenter cannot do this without disturbing the environment, so the prepared states are not really inputs anymore. The crucial insight in Ref. [75] was thus to adopt a description that focuses on the objects that can be controlled operationally: preparations and measurements. Moreover, we will see that we do not need to give up linearity or positivity in any meaningful way.

In general, any experimental intervention can be described by any CPTNI maps $\mathcal{A}_i$. This means that for a preparation $\mathcal{A}$, the reduced output state of a system, S, after a given unitary evolution $\mathcal{U}$ can be described as an object $\mathcal{M}$ taking as input $\mathcal{A}$ and giving as an output a quantum state,

$$\rho'_{\mathsf{S}} = \mathrm{tr}_{\mathsf{E}}[\mathcal{U}\mathcal{A}(\rho_{\mathsf{SE}})] := \mathcal{M}[\mathcal{A}], \tag{2.26}$$

where implicitly we denote $\mathcal{A}$ for $\mathcal{A} \otimes \mathcal{I}_{\mathsf{E}}$, as we will throughout this thesis unless the distinction need to be made explicitly. This can be visualized as in Fig. 2.3. Keep in mind that if the map $\mathcal{A}$ is TNI, then in general the output can be subnormalized, $\mathrm{tr}(\rho'_{\mathsf{S}}) \leq 1$, so despite us referring to it as a quantum state, we are allowing for it to have some missing information whenever the operation $\mathcal{A}$ might fail to be implemented.

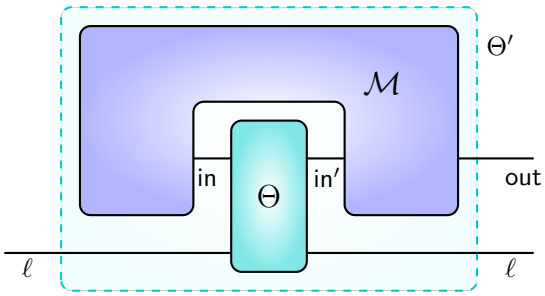


Figure 2.4: **The CP property for a superchannel** means $(\mathcal{M}\otimes\mathcal{I}_\ell)\Theta = \Theta' \geq 0$ for any CP map Θ and any ancillary dimension $\ell \geq 0$; here Θ' can be thought of as a map from an input space of dimension ℓ to a product of such space together with the output space from the action of $\mathcal{M}$.

The map $\mathcal{M}$ is known as the *superchannel*, as it generalizes the idea of a quantum channel as a map from superoperators to quantum states. The superchannel contains all the dynamical information inaccessible to the experimenter. It can be said to be a supersuperoperator $\mathcal{M} : \mathscr{B}(\mathscr{H}_{\mathsf{in}}) \otimes \mathscr{B}(\mathscr{H}_{\mathsf{in}'}) \to \mathscr{B}(\mathscr{H}_{\mathsf{out}})$. Linearity can be seen by inspection of Eq. (2.26), as

$$\mathcal{M}[\sum \alpha_i \mathcal{A}_i] = \sum \alpha_i \mathcal{M}[\mathcal{A}_i], \tag{2.27}$$

and similarly trace preservation can be seen in the sense that $\mathrm{tr}[\mathcal{M}[\mathcal{A}]] = \mathrm{tr}[\mathcal{A}(\rho_{\mathsf{S}})]$, which corresponds to unity when $\mathcal{A}$ is TP, or otherwise to a probability of successfully implementing $\mathcal{A}$.

Complete positivity, however, now means that if we have a channel acting together on an ancillary space of dimension ℓ, i.e. $\Theta : \mathscr{B}(\mathscr{H}_{\mathsf{in}}\otimes\mathscr{H}_\ell) \to \mathscr{B}(\mathscr{H}_{\mathsf{in}'}\otimes\mathscr{H}_\ell)$, then the output corresponds to a positive map $\Theta' : \mathscr{B}(\mathscr{H}_\ell) \to \mathscr{B}(\mathscr{H}_{\mathsf{out}}\otimes\mathscr{H}_\ell)$ for all dimensions ℓ, with

$$(\mathcal{M}\otimes\mathcal{I}_\ell)\Theta = \Theta' \geq 0, \qquad \forall \ell \geq 0, \tag{2.28}$$

where $\mathcal{M}$ acts only on the "in" part of Θ. This is more easily visualized as in Fig. 2.4.

We know that a channel being CP implies a positive Choi state and vice versa. Now, as mentioned before, the superchannel is just a slightly more elaborate channel, so we may similarly obtain its Choi state. Remember that for a channel, the way to obtain the corresponding choi state is simply to act on the input space with half an unnormalized maximally entangled state $\tilde{\Psi} = \sum |ii\rangle\langle jj| \in \mathscr{B}(\mathscr{H}_{\mathsf{in}}\otimes\mathscr{H}_{\mathsf{in}})$ and we obtained an unnormalized state in a larger space including the second copy of the input space. In this case, looking at Fig. 2.3(a), we see that we need to act on $\mathscr{H}_{\mathsf{in}'}$ with half $\tilde{\Psi} \in \mathscr{B}(\mathscr{H}_{\mathsf{in}'}\otimes\mathscr{H}_{\mathsf{in}'})$ and the remaining spaces should go out (almost) intact, giving an unnormalized state in a larger space composed of the input and output spaces.

To act with half the maximally entangled state we can simply swap half of it with the input spaces; let us label the new ancillary spaces by A and B, which really just copies of S and have the same dimension, i.e. $d_{\mathsf{A}} = d_{\mathsf{B}} = d_{\mathsf{S}}$. Then

$$\Upsilon_{\mathcal{M}} := \mathrm{tr}_{\mathsf{E}}[\mathcal{U}\,\mathcal{S}\,(\rho_{\mathsf{SE}}\otimes\tilde{\Psi})], \tag{2.29}$$

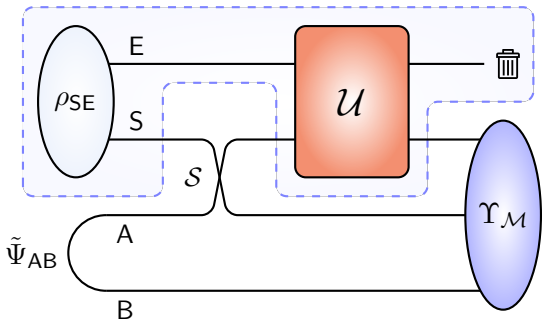


Figure 2.5: **The Choi state $\Upsilon_{\mathcal{M}}$ of the superchannel** can similarly be obtained by letting it act on half an unnormalized maximally entangled state $\tilde{\Psi}_{\mathsf{AB}}$; to do this, half of the ancillary space is swapped with the input on S. Here A and B are labels: both correspond to copies of S.

is the Choi state of the superchannel, where here $\mathcal{U}$ stands for $\mathcal{U} \otimes \mathcal{I}_{\mathsf{AB}}$, and

$$\mathcal{S}(\cdot) := \text{SWAP}_{\mathsf{SA}}(\cdot)\,\text{SWAP}_{\mathsf{SA}} \tag{2.30}$$

$$\text{where here} \quad \text{SWAP}_{\mathsf{SA}} := \mathbb{1}_{\mathsf{E}} \sum_{i,j=1}^{d_{\mathsf{S}}} |ij\rangle\langle ji| \otimes \mathbb{1}_{\mathsf{B}}, \tag{2.31}$$

given that here the input and output space is simply the subsystem S. The swap of course can be done with either A or B. This can also be seen clearly through a circuit diagram as in Fig. 2.5.

Let us define $\mathfrak{S}_{\alpha\beta} := \mathbb{1}_{\mathsf{E}} \otimes |\alpha\rangle\langle\beta|$, so that we can expand the definition of the Choi state as

$$\Upsilon_{\mathcal{M}} = \sum_{\alpha,\ldots,\delta} \text{tr}_{\mathsf{E}}[U\ \mathfrak{S}_{\alpha\beta}\,\rho_{\mathsf{SE}}\,\mathfrak{S}_{\delta\gamma}\,U^\dagger] \otimes |\beta\alpha\rangle\langle\delta\gamma|, \tag{2.32}$$

then to recover the action of the superchannel on a given map $\mathcal{A}$, we need to contract the Choi state of $\mathcal{M}$ with that of $\mathcal{A}$. First notice that we have to write $\Upsilon_{\mathcal{A}} = (\mathbb{1}_{\mathsf{A}} \otimes \mathcal{A})\tilde{\Psi}$, i.e. $\mathcal{A}$ has to act on the B ancillary space (analogous to Eq. (2.19), where we contract with the entangled half that was not acted on in $\Upsilon_{\mathcal{M}}$); this can also be understood easily through diagrams, as in Fig. 2.6. This is, tracing over all inputs

$$\begin{aligned}
\text{tr}_{\mathsf{in}}[\Upsilon_{\mathcal{M}}(\mathbb{1}_{\mathsf{out}} \otimes \Upsilon_{\mathcal{A}}^{\mathrm{T}})] &= \text{tr}_{\mathsf{in}}\left\{\sum_{\alpha,\ldots,\delta} \text{tr}_{\mathsf{E}}[U\,\mathfrak{S}_{\alpha\beta}\,\rho_{\mathsf{SE}}\,\mathfrak{S}_{\delta\gamma}\,U^\dagger] \otimes |\beta\alpha\rangle\langle\delta\gamma|\Upsilon_{\mathcal{A}}^{\mathrm{T}}\right\} \\
&= \sum_{\alpha,\ldots,\delta,i,j} \text{tr}_{\mathsf{E}}[U\,\mathfrak{S}_{\alpha\beta}\,\rho_{\mathsf{SE}}\,\mathfrak{S}_{\delta\gamma}\,U^\dagger]\langle\delta|j\rangle\langle i|\beta\rangle\langle\gamma|\mathcal{A}^{\mathrm{T}}(|i\rangle\langle j|)|\alpha\rangle \\
&= \sum_{\alpha,\ldots,\delta} \text{tr}_{\mathsf{E}}[U(\mathbb{1}_{\mathsf{E}} \otimes |\alpha\rangle\langle\beta|)\rho_{\mathsf{SE}}(\mathbb{1}_{\mathsf{E}} \otimes |\delta\rangle\langle\gamma|)U^\dagger]\langle\alpha|\mathcal{A}(|\beta\rangle\langle\delta|)|\gamma\rangle \\
&= \sum_{\alpha,\ldots,\delta}\sum_{\mu} \text{tr}_{\mathsf{E}}[U(\mathbb{1}_{\mathsf{E}} \otimes |\alpha\rangle\langle\beta|)\rho_{\mathsf{SE}}(\mathbb{1}_{\mathsf{E}} \otimes |\delta\rangle\langle\gamma|)U^\dagger]\langle\alpha|A_\mu|\beta\rangle\langle\delta|A_\mu^\dagger|\gamma\rangle \\
&= \sum_{\mu} \text{tr}_{\mathsf{E}}[U(\mathbb{1}_{\mathsf{E}} \otimes A_\mu)\rho_{\mathsf{SE}}(\mathbb{1}_{\mathsf{E}} \otimes A_\mu^\dagger)U^\dagger] \\
&= \text{tr}_{\mathsf{E}}[\mathcal{U}\,\mathcal{A}\,(\rho_{\mathsf{SE}})],
\end{aligned} \tag{2.33}$$

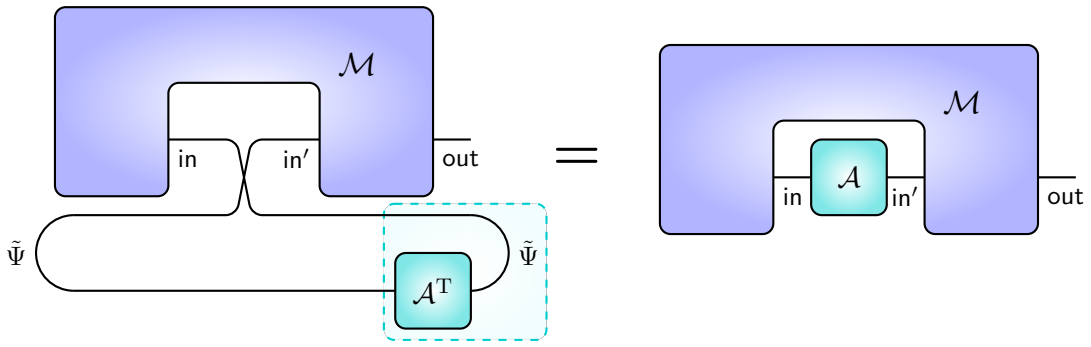


Figure 2.6: **The action of a superchannel** $\mathcal{M}$ on a CP map $\mathcal{A}$ can be given as the contraction of the respective Choi states as $\mathcal{M}[\mathcal{A}] = \mathrm{tr}_{\mathsf{in}(\prime)}[\Upsilon_{\mathcal{M}}(\mathbb{1}_{\mathsf{out}} \otimes \Upsilon_{\mathcal{A}}^{\mathrm{T}})]$, where $\Upsilon_{\mathcal{A}}^{\mathrm{T}} = (\mathbb{1}_{\mathsf{in}} \otimes \mathcal{A}^{\mathrm{T}})\tilde{\Psi}$.

where in the fifth line we used the fact that $\mathcal{A}$ is CP so that $\mathcal{A}(\cdot) = \sum A_\mu(\cdot)A_\mu^\dagger$.

The Choi state is manifestly positive, given that everything in its definition is positive; this in turn implies that $\mathcal{M}$ is CP, as expected. We may now think of this operationally as a situation in which the experimenter brings along an ancillary system of dimension ℓ and performs an entangling operation $\Theta_{\mathsf{S}\ell}$, with the subsequent SE dynamics occurring; then we are assured that the output $\Theta'_{\mathsf{S}\ell}$ will be positive. The Kraus operators of $\mathcal{M}$ can be obtained as well, as done in Ref. [83, 126]. This shows conclusively that the dynamics of initially correlated systems are indeed completely positive, we were only looking in the wrong place. Notice as well that the CPTP property for the superchannel was a natural consequence rather than an a-priori condition.

Now, the superchannel, just as in the case of quantum channels, can be reconstructed experimentally through QPT without any a-priori knowledge about the initial state or the SE dynamics. In a similar fashion to standard QPT, now the experimenter requires d_{S}^4 linearly independent CP maps $\{\mathcal{A}_i\}$ and measuring the corresponding outputs $\rho_{\mathsf{S}}^{(i)}$, with which e.g. the Choi state can be reconstructed. This procedure, just as any QPT or QST procedure requires a relatively great amount of resources, however it was reported already in Ref. [130].

Going back to the discussion of NCP dynamical maps, notice that the case $\mathcal{M}[\mathcal{I}]$ gives the usual dynamical map scenario, which however is operationally void without determining the superchannel. Notice that the initial uncorrelated state case with a dynamical map is contained as a particular one under the superchannel, i.e. for $\rho_{\mathsf{SE}} = \rho_{\mathsf{S}} \otimes \varepsilon$, we have

$$\mathcal{M}[\mathcal{A}] = \mathrm{tr}_{\mathsf{E}}[\mathcal{U}_t\,(\mathcal{A}(\rho_{\mathsf{S}}) \otimes \varepsilon)] = \mathscr{Z}_t[\sigma_{\mathsf{S}}], \tag{2.34}$$

where $\sigma_{\mathsf{S}} = \mathcal{A}(\rho_{\mathsf{S}})$. In such case the dynamical map description is sufficient, however, in general the superchannel can be employed to quantify initial correlations e.g. by letting $\rho_{\mathsf{SE}} = \rho_{\mathsf{S}} \otimes \varepsilon + \chi_{\mathsf{SE}}$ [126]. As we will see below, this also turns out to be a special case of the general notion of a Markovian process.

2.4 Multiple time-steps

One of the crucial features of the superchannel is that it allows to diagnose the presence of initial correlations playing a role in the reduced subsystem dynamics. This is already some form of *memory* or non-Markovian effect. This, however, only accounts for correlations between the initial preparation and the final measurement; if we were to fully describe non-Markovianity, we would look for an extension to any dependence on any given time in the past, analogous to the classical definition for non-Markovianity.

This is crucial because if we are to deal consistently with non-Markovianity, we first require a non-ambiguous characterization that genuinely deals with temporal correlations across an arbitrary number of points. Far from being just an interesting theoretical aspect, this is a tremendously practical question, given the ever-increasing technological capability and interest to deal with noisy systems over several time-steps [131–140].

Let us then first revisit the classical notion of a stochastic process and of Markovianity.

2.4.1 Classical stochastic processes

A classical stochastic process is a probabilistic process in time, i.e. it can be described mathematically as a collection of random variables indexed by time. As these are mainly used to model situations in which only partial knowledge or certainty is available, it is no surprise that classical stochastic processes are ubiquitous in our attempt to describe anything from coin tossing to financial markets or disease spread. We have already appealed to some of the notions needed to discuss stochastic processes, however, we can formalize these a bit more to make these concepts concrete. Most of these concepts can be consulted in full detail in standard textbooks such as [141] or in Ref. [8] related to open systems or furthermore in particular in the context related to this thesis in Ref. [89, 90].

It is worth mentioning that some of the notation we use here is only for local purposes i.e. it does not necessarily have any relation with the symbols on previous sections unless explicitly stated.

First we need to introduce a *probability space*, which sets the scene to model probabilistically a given class of situations: this is a triple (Ω, Σ, μ) where

i. Ω is called a *sample space*; mathematically, this is an arbitrary nonempty set. It stands for the set of all (discrete or continuous) outcomes, i.e. single realisations of the model.

ii. Σ is called an *event space* and is a collection of subsets of the sample space called events. The event space Σ is a σ-algebra: this means that, denoting by $\pi(\Omega)$ the power set (set of all subsets) of the sample space, $\Sigma \subseteq \pi(\Omega)$ satisfies[5]
   - $\Omega \in \Sigma$.
   - If $\alpha \in \Sigma$, then also $(\Omega \setminus \alpha) \in \Sigma$.

[5] We use standard set notation with $\setminus$ meaning set subtraction, $\cup$ union of sets and $\cap$ intersection of sets.

- If $\alpha_i \in \Sigma$ for $i = 1, 2, \ldots$, then also $\cup_i \alpha_i \in \Sigma$.
- If $\alpha_i \in \Sigma$ for $i = 1, 2, \ldots$, then also $\cap_i \alpha_i \in \Sigma$.

Together, the pair (Ω, Σ) is called a *measurable space*.

iii. μ is called a *probability measure*. It is a function $\mu : \Sigma \to [0, 1]$ such that

- $\mu(\Omega) = 1$
- $\mu(\cup_i \alpha_i) = \sum_i \mu(\alpha_i)$ for any countable collection of pairwise disjoint[6] subsets $\{\alpha_i\} \in \Sigma$.

Example 2.4.1. Perhaps the simplest example is a coin toss: if a fair coin is tossed three times, then $\Omega = \{\texttt{HHH}, \texttt{HHT}, \texttt{HTT}, \texttt{HTH}, \texttt{THT}, \texttt{THH}, \texttt{TTH}, \texttt{TTT}\}$ is the sample space, where $\texttt{H}$ and $\texttt{T}$ stand for 'heads' and 'tails', respectively, and the order in each element stands for the first, second and third toss. The power set $\pi(\Omega)$ has $2^{|\Omega|} = 2^8$ events. If the experiment concerns the event "at least two heads occur", i.e. $\omega = \{\texttt{HHH}, \texttt{HHT}, \texttt{HTH}, \texttt{THH}\}$, then the event space is $\Sigma = \{\emptyset, \varkappa, \omega, \Omega\}$, where $\varkappa = \Omega \setminus \omega$, so the elements correspond to the experiment not being performed, not more than one head occurred, at least two head occurred and the experiment being performed. The probability measure gives $\mu(\emptyset) = 0$, $\mu(\omega) = \mu(\varkappa) = 0.5$, $\mu(\Omega) = 1$.

Now, we may define a *random variable* X on the probability space (Ω, Σ, μ) as a function $X : \Omega \to \mathbb{R}$; in general these can correspond to functions from the sample space to any other measurable space, e.g. in Section 1.3.1 we considered random unitary matrices as random variables. Intuitively, random variables are real-valued quantities that can be measured from outcomes of random trials. This point of view is useful simply because an experimenter often cares about some function of the outcomes of the experiment rather than the outcomes themselves. This allows to ask about the probability that the random variable takes a value within a subset $s \in \mathtt{S}$, where $\mathtt{S} \subseteq \mathbb{R}$ is a collection of open subsets[7] of the real numbers, i.e. in a slight abuse of notation we may define

$$\mathbb{P}[X \in s] := \mathbb{P}[\{\omega \in \Omega : X(\omega) \in s\}], \tag{2.35}$$

where $\mathbb{P}$ is a probability distribution $\mathbb{P} : \mathtt{S} \to [0, 1]$, related to the probability measure μ via $\mathbb{P}[s] = \mu[X^{-1}(s)]$, with $X^{-1}(s) \in \Sigma$. This is so because $\mathtt{S}$ is a σ-algebra and $(\mathbb{R}, \mathtt{S})$ a measurable space. We notice that given a function $f : \mathbb{R} \to \mathbb{R}$ such that $f^{-1}(s) \in \mathtt{S}$, then any other Y defined by $Y(\omega) = f(X(\omega))$ for an event ω is also a random variable. In general we will refer to the value $X(\omega)$ as a *realisation*.

A *stochastic process* then concerns $\mathtt{S}$-valued random variables X in a parameter $t \geq 0$, usually standing for time. That is, a stochastic process is a random variable $X : \Omega \times \mathbb{R}_0^+ \to \mathbb{R}$, where $\mathbb{R}_0^+$ stands for the set of non-negative real numbers. We can then think of a stochastic process in two ways. For a fixed event $\omega \in \Omega$, the function $X : \mathbb{R}_0^+ \to \mathbb{R}$ defines a so-called trajectory of the stochastic process. Conversely, $X : \Omega \to \mathbb{R}$ for a fixed $t \geq 0$ denotes a collection of random variables at a given time.

In practice we usually care about the probability distribution of a given trajectory and a

[6] That is, $\alpha_i \cap \alpha_j = \emptyset, \forall i \neq j$.
[7] Formally these have to be so-called Borel subsets [141].

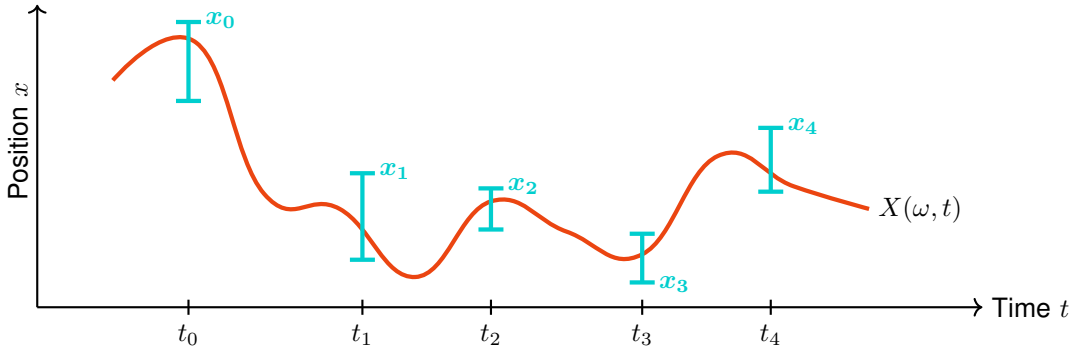


Figure 2.7: **An example of a stochastic trajectory** for a particle in a one-dimensional random walk: for given times $\mathscr{T}_5 = \{t_0, t_1, \ldots, t_4\}$, a classical stochastic process is determined by the joint probability distribution $\mathbb{P}_{\mathscr{T}_5}(x_0, x_1, \ldots, x_4)$ to find the particle in the region x_i at time t_i.

sensible way to describe it is through a discrete number of k times

$$\mathscr{T}_k = \{t_0, t_1, , \ldots, t_{k-1}\}, \tag{2.36}$$

and whenever we explicitly assume that times are ordered as $t_0 < t_1 < \ldots < t_n$, we will denote this by $n : 0$. We can then construct a vector of random variables $\mathbf{X}_{\mathscr{T}_k} = (X_0(t_0), X_1(t_1), \ldots, X_{k-1}(t_{k-1}))$ for each time in $\mathscr{T}_k$ with each $X_i(\cdot, t_i) : \Omega \to \mathbb{R}$. This then defines the joint probability distribution

$$\mathbb{P}(\vec{x}_{\mathscr{T}_k}) = \mu\left[\mathbf{X}^{-1}_{\mathscr{T}_k}(\vec{x}_{\mathscr{T}_k})\right], \tag{2.37}$$

for realisations $X_i(\omega, t_i)$ of an event $\omega \in \Omega$ to lie within x_i at time t_i, where we also defined $\vec{x}_{\mathscr{T}_k} = (x_0, x_1, \ldots, x_{k-1})$, i.e. the expression in Eq. (2.37) explicitly means, for example, for a two time-step process,

$$\mathbb{P}(x_0, x_1) = \mathbb{P}[\{\omega \in \Omega : X_0(\omega, t_0) \in x_0 \text{ and } X_1(\omega, t_1) \in x_1\}]. \tag{2.38}$$

This is enough to statistically characterize the stochastic process with finite time-steps as the probability for a given trajectory to fall within sets x_i at each time t_i, together with k-point correlations between the random variables at each time-step.

The choice of $\mathscr{T}_k$ is arbitrary, so this already tell us that if obtain realisations for some but not all such times, say $\mathscr{T}_\ell$, with $|\mathscr{T}_\ell| < |\mathscr{T}_k|$, then the joint probability distribution $\mathbb{P}(\vec{x}_{\mathscr{T}_\ell})$ should be contained in $\mathbb{P}_{\mathscr{T}_k}(\vec{x}_{\mathscr{T}_k})$. Indeed, we can marginalise, i.e. sum out the probabilities for the extra random variables in the larger set of times to recover the joint distribution in the smaller one,

$$\mathbb{P}(\vec{x}_{\mathscr{T}_\ell}) = \sum_{\mathscr{T}_k \setminus \mathscr{T}_\ell} \mathbb{P}(\vec{x}_{\mathscr{T}_k}), \tag{2.39}$$

where the sum runs over realisations for the excessive times in $\mathscr{T}_k$ but not in $\mathscr{T}_\ell$.

This containment property generalizes to the fact that there exists an infinite joint probability distribution that contains all the finite ones: this is reconciled through the so-called Kolmogorov

extension theorem [8, 142, 143], which binds the realistic scenario of a discrete number of observations with the firm mathematical footing of a general stochastic process. More specifically, it gives the consistency conditions for a family of joint probabilities to guarantee the existence of an underlying continuous stochastic process.

2.4.2 The classical Markov condition

Within any classical stochastic process we can consider different particular cases for the joint distribution in Eq. (2.7). In particular, a Markov process as introduced for open quantum systems as being *memoryless* implies that, relative to a given time-step, the probabilities for future steps should only depend on such time-step[8] and not on the past ones.

Definition 2.4.2 (Classical Markov condition)**.** A classical stochastic process is called Markovian if for any ordered times $t_0 < t_1 < \cdots < t_k$ we have

$$\mathbb{P}(x_k|\vec{x}_{k-1:0}) = \mathbb{P}(x_k|x_{k-1}), \tag{2.40}$$

where here $\vec{x}_{j:0} = (x_0, \ldots, x_j)$ and $\mathbb{P}(A|B)$ refers to the *conditional probability* of event A given event B, defined by $\mathbb{P}(A|B) = \mathbb{P}(A, B)/\mathbb{P}(B)$, i.e.

$$\frac{\mathbb{P}(\vec{x}_{k:0})}{\mathbb{P}(\vec{x}_{k-1:0})} = \mathbb{P}(x_k|\vec{x}_{k-1:0}) \overset{\text{Markov}}{=} \mathbb{P}(x_k|x_{k-1}) = \frac{\mathbb{P}(x_k, x_{k-1})}{\mathbb{P}(x_{k-1})}. \tag{2.41}$$

Conversely, any classical process that does not satisfy Eq. (2.40) is called non-Markovian.

Physically, a conditional probability $\mathbb{P}(x_j|x_i)$ can be understood as a probability for a system to go from a state x_i to state x_j and is often referred to as a transition probability or a propagator. The Markov condition is a very significant simplification, since normally we would need all $(k+1)$-point correlations to compute the following joint probability distribution. Notice that in general, by repeatedly applying the definition of the conditional probability,

$$\mathbb{P}(\vec{x}_{k:0}) = \mathbb{P}(x_k|\vec{x}_{k-1:0})\mathbb{P}(x_{k-1}|\vec{x}_{k-2:0}) \cdots \mathbb{P}(x_1|x_0)\mathbb{P}(x_0), \tag{2.42}$$

so that to determine the full joint probability distribution we require an increasing number of transition probabilities. However, if the the Markov condition holds, a Markovian joint distribution satisfies

$$\mathbb{P}(\vec{x}_{k:0}) = \mathbb{P}(x_k|x_{k-1})\mathbb{P}(x_{k-1}|x_{k-2}) \cdots \mathbb{P}(x_1|x_0)\mathbb{P}(x_0), \tag{2.43}$$

requiring only 2-point correlations to determine the future joint distribution.

Specifically, suppose we have a classical system with n possible outcomes (as in the case of a coin flip, $n = 2$, for example). Then to determine the distribution $\mathbb{P}(x_0)$ we require $n - 1$ probabilities, to determine $\mathbb{P}(x_1|x_0)\mathbb{P}(x_0)$ we require $n - 1 + n(n - 1) = n^2 - 1$ probabilities, and so on until we reach $n^{k+1} - 1$ to determine $\mathbb{P}(\vec{x}_{k:0})$, i.e. we get an exponential growth in

[8] Strictly speaking, outright *memoryless* should refer to no dependence even in the present state.

the number of time-steps k. However, if the process is Markovian, we require $(n-1)(1+kn)$ probabilities, so just a linear growing number of terms in the number of time-steps k.

Notice that the Markov condition is a statement regarding multiple time-steps, or equivalently a statement regarding the observation of multiple events. That is, to have a general notion of Markovianity or memorylessness we must consider correlations across several events in time.

Now, continuing with the case of n possible outcomes, notice that we can store the initial probability distribution $\mathbb{P}(x_0)$ as an n-dimensional vector $\mathbb{P}_0$ with entries corresponding to the probability of each outcome so that

$$\mathbb{P}_k = \mathfrak{P}_{k:0}\mathbb{P}_0, \tag{2.44}$$

is the vector corresponding to $\mathbb{P}(x_k)$, where $\mathfrak{P}_{k:0}$ is an $n \times n$ matrix containing the propagator for each outcome from time-step t_0 to time-step t_k. Explicitly, this means a propagator matrix $\mathfrak{P}_{j:i}$ from time-step t_i to step $t_j > t_i$, with random variables taking given values $X_i = u_a$ and $X_j = v_b$, where $1 \leq a, b \leq n$, will have the probability entries

$$\left(\mathfrak{P}_{j:i}\right)_{ab} = \mathbb{P}[X_i = u_a | X_j = v_b], \tag{2.45}$$

which are such that

$$\sum_{b=1}^{n} \left(\mathfrak{P}_{j:i}\right)_{ab} = 1, \tag{2.46}$$

and are called *stochastic matrices*. This property can be thought as justifying the label of these matrices as propagators, as they would give an evolution of a probability vector $\mathbb{P}_i$ at time t_i to a legitimate probability vector $\mathbb{P}_j$ at time t_j.

A Markovian process then implies that the full propagator can be broken up in the intermediate steps as

$$\mathfrak{P}_{k:0} = \mathfrak{P}_{k:k-1}\mathfrak{P}_{k-1:k-2}\cdots\mathfrak{P}_{1:0}, \tag{2.47}$$

and we can further marginalise to break it up in any two propagator matrices as

$$\mathfrak{P}_{k:0} = \mathfrak{P}_{k:j}\mathfrak{P}_{j:0}, \qquad \forall t_k > t_j > t_0, \tag{2.48}$$

which is known as a divisibility property, and which is the motivation for the analogous property on quantum dynamical maps to define Markovianity as in Eq. (2.24).

The divisibility property on two contiguous time-steps can equivalently be expressed as

$$\mathbb{P}(x_{\ell+1}|x_{\ell-1}) = \sum_{t_\ell} \mathbb{P}(x_{\ell+1}|x_\ell)\mathbb{P}(x_\ell|x_{\ell-1}), \tag{2.49}$$

where the time-steps are ordered $t_{\ell-1} < t_\ell < t_{\ell+1}$ for any $\ell > 0$, which usually in this form is known as the Chapman-Kolmogorov equation [78]. The intuitive explanation of this equation is that the probability of transitioning from a state $x_{\ell-1}$ to another $x_{\ell+1}$ can be obtained by multiplying the transition probabilities to and from an intermediate state x_ℓ and summing over all the possible of these.

Both divisibility in Eq. (2.48) and the Chapman-Kolmogorov equation in Eq. (2.49), however, do not say anything about higher point correlations, i.e. divisibility is insufficient to characterize Markovianity and it is possible to find non-Markovian processes which can satisfy divisibility [144–146], where however, the propagator would not correspond to a conditional transition probability matrix [147].

Clearly the Markov condition is a massive mathematical simplification: because of this it has been largely studied and has a wide reach and applicability [148, 149]. If we turn to physics, however, not only in the more general quantum case is it an idealization but also in the classical case as soon as we start considering fully realistic scenarios we realise that Markovian processes are rather exceptional [150].

A simple example contrasting both kinds of processes is that of drawing a marble from a bag full of marbles with a handful of colors: if we were to replace or put back each marble as we draw it from the bag, the process will be Markovian and we do not need to remember anything to update our predictions of what color the next marble will have; if, however, we discard each marble once we take it out, by knowing the full amount of marbles we have to keep track of all the colors as they come out to update our predictions accordingly. Other examples for realistic scenarios can be seen in Ref. [150], in particular for cases where Markovianity becomes a reasonable approximation within certain limits in which the initial configuration is forgotten.

In a sense, with the previous Chapter 1 we have seen that equilibration on average is indeed a process where the initial state is effectively forgotten. Let us thus go back to approaching the case of non-Markovianity and stochastic processes in quantum mechanics.

2.4.3 The process tensor

To consider the generalization of a stochastic process in open quantum systems we first need to consider multiple interventions. Given the operational scenario of the superchannel, we see that what we need is to extend this picture to not only a preparation and a measurement but to an arbitrary number of interventions. Crucially, notice that in the discussion of classical stochastic processes, there was an implicit assumption that realisations do not alter in any way the subsequent states, and thus such assumption will not hold in the quantum extension anymore.

The physical scenario is now the following: an experimenter prepares a fiducial quantum state ρ_{SE} on system S of a joint SE composite through an operation $\mathcal{A}_0$, which in general is an $\mathtt{CPTNI}$ map; subsequently the whole composite evolves unitarily through a unitary map $\mathcal{U}_1$, after which an operation $\mathcal{A}_1$ is performed on S, then the whole evolves unitarily under a unitary map $\mathcal{U}_2$, and so on, until an intervention $\mathcal{A}_{k-1}$, followed finally by a unitary map $\mathcal{U}_k$. This means the final state in system S will be given by

$$\rho_{\mathsf{S}}^{(k)} = \mathrm{tr}_{\mathsf{E}}\left[\mathcal{U}_k \, \mathcal{A}_{k-1} \, \mathcal{U}_{k-1} \, \mathcal{A}_{k-1} \cdots \mathcal{U}_1 \mathcal{A}_0 \left(\rho_{\mathsf{SE}}\right)\right], \tag{2.50}$$

where here we also implicitly write $\mathcal{A}_\ell$ for $\mathcal{A}_\ell \otimes \mathcal{I}_{\mathsf{E}}$. As for the case of the superchannel, keep in mind that $\rho_{\mathsf{S}}^{(k)}$ can be subnormalized (even if we refer to it as a quantum state) if the operations

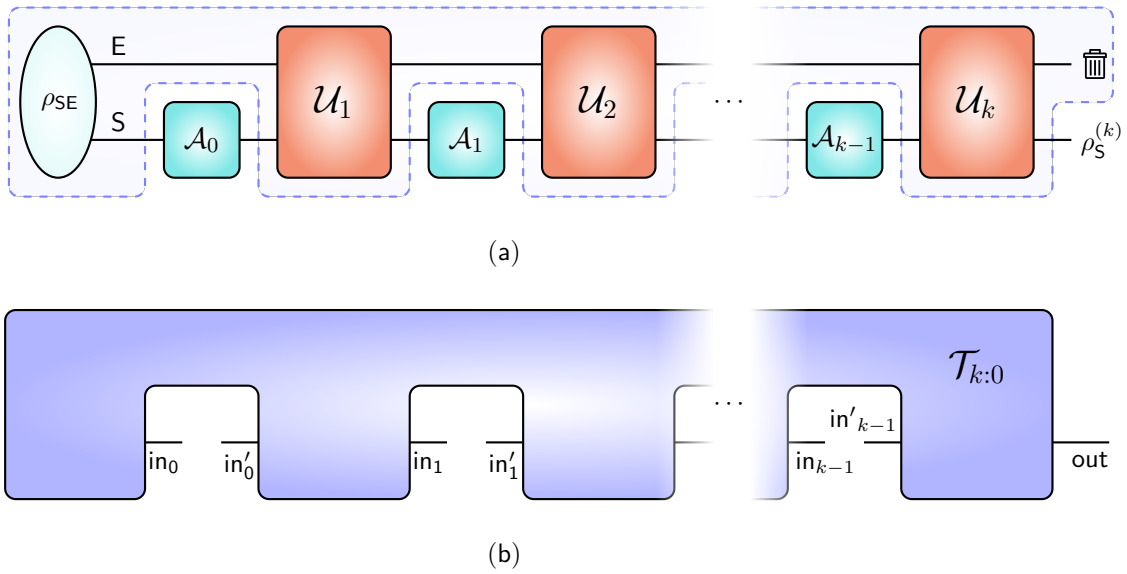


Figure 2.8: **A k-step process tensor is a map from k operations (CPTNI maps) to a quantum state:** (a) A joint system-environment in a fiducial state ρ_{SE} is prepared with an operation $\mathcal{A}_0$ and subsequently undergoes a joint unitary evolution $\mathcal{U}_1$, until an intervention with an operation $\mathcal{A}_1$ is made with subsequent evolution $\mathcal{U}_1$, and so on until a time-step k, where the state of the system S is given by $\rho_{\mathsf{S}}^{(k)}$. (b) The process tensor $\mathcal{T}_{k:0}$ is all of the content out of direct control to the experimenter, and contains all the dynamical information about the system, including SE correlations across all points in time.

are in general TNI but not necessarily TP.

Now we see that the generalization of the superchannel to k time-steps must be a map $\mathcal{T}_{k:0} : \mathscr{B}(\mathscr{H}_{\mathsf{S}})^{\otimes 2k} \to \mathscr{B}(\mathscr{H}_{\mathsf{S}})$, taking k CPTNI maps as arguments and giving a quantum state as output at time-step k, i.e.

$$\mathcal{T}_{k:0}[\vec{\mathcal{A}}_{k-1:0}] = \rho_{\mathsf{S}}^{(k)}, \tag{2.51}$$

where $\vec{\mathcal{A}}_{k-1:0} = (\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_{k-1})$. This can depicted diagramatically as in Fig. 2.8.

Such a map, $\mathcal{T}_{k:0}$ is called a *process tensor* [151, 152] and we will refer to it as a k-step process tensor, or often simply as a k-step process. In this sense, a 0-step process is a joint fiducial state, which might have undergone a unitary evolution, and yielding an S quantum state, e.g. as is considered for equilibration on average. A superchannel is the particular case $\mathcal{M} = \mathcal{T}_{1:0}$, i.e. a 1-step process, taking a preparation and yielding a reduced S quantum state. The name of $\mathcal{T}_{k:0}$ makes reference to the multilinear structure of such map, which can easily be seen by insertion through Eq. (2.50), as $\mathcal{T}_{k:0}[\alpha\vec{\mathcal{A}}_{k:0}] + \beta\vec{\mathcal{B}}_{k:0}] = \alpha\mathcal{T}_{k:0}[\vec{\mathcal{A}}_{k:0} + \beta\mathcal{T}_{k:0}[\vec{\mathcal{B}}_{k:0}]$ for any two sets of operations $\{\mathcal{A}_i\}$ and $\{\mathcal{B}_i\}$ and scalars α, β. This implies that the process tensor can be reconstructed through a relevant tomographic scheme [153]. Henceforth, whenever we consider a general k-step process tensor, we will simply denote it by $\mathcal{T}$ unless necessary otherwise.

Now, we know that the superchannel is CP, however, we can similarly check that this will be

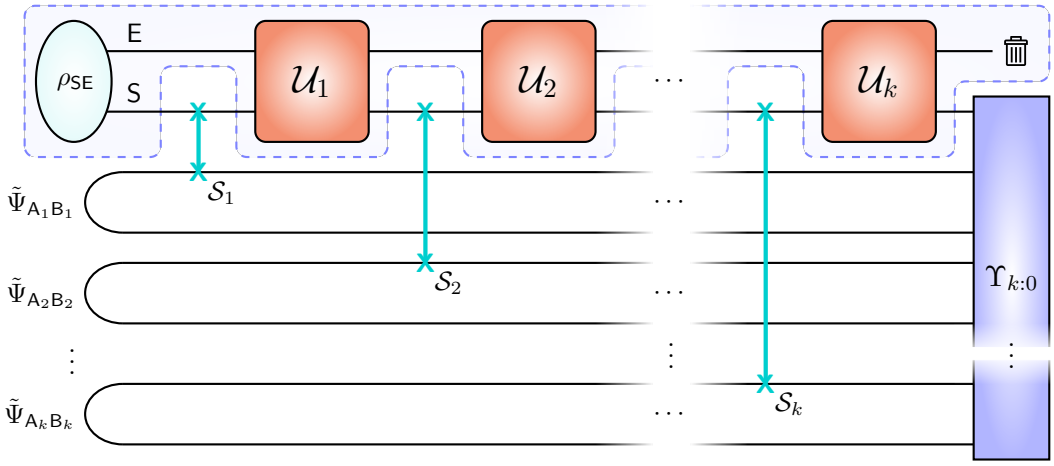


Figure 2.9: **The Choi state representation of a k-step process tensor**, denoted $\Upsilon_{k:0}$, can be obtained by swapping out the system, $\mathcal{S}_i$ with half a maximally entangled state, $\Psi_{\mathsf{A}_i\mathsf{B}_i}$, at each step i. The final state is an unnormalized many-body state acting on a d_{S}^{2k+1} dimensional system.

the case for an arbitrary number of time-steps also by constructing its Choi state and checking that is is positive. Knowing the procedure for the superchannel, depicted in Fig. 2.5, we can see that the generalization to an arbitrary number of k steps follows by introducing k maximally entangled states $\tilde{\Psi}_{\mathsf{A}_i\mathsf{B}_i} \in \mathscr{B}(\mathscr{H}_{\mathsf{A}_i} \otimes \mathscr{H}_{\mathsf{B}_i})$, where $\mathscr{H}_{\mathsf{A}_i} \cong \mathscr{H}_{\mathsf{S}}$ and similarly for B, and letting half of each act as an input at every step by swapping the input spaces with the corresponding ancilla. This is more clearly illustrated in Fig. 2.9.

Specifically, the Choi state of the process tensor takes the form

$$\Upsilon_{k:0} = \mathrm{tr}_{\mathsf{E}}\left[\mathcal{U}_k \mathcal{S}_k \mathcal{U}_{k-1} \mathcal{U}_{k-1} \cdots \mathcal{U}_1 \mathcal{S}_1 \left(\rho_{\mathsf{SE}} \otimes \tilde{\Psi}^{\otimes k}\right)\right], \tag{2.52}$$

where here we are implicitly writing $\mathcal{U}_i$ for $\mathcal{U}_i \otimes \mathcal{I}_{\mathsf{A}_1\mathsf{B}_1\cdots\mathsf{A}_k\mathsf{B}_k}$ and $\tilde{\Psi}^{\otimes k} = \tilde{\Psi}_{\mathsf{A}_1\mathsf{B}_1} \otimes \cdots \otimes \tilde{\Psi}_{\mathsf{A}_k\mathsf{B}_k}$. The generalized swap between system S and an ancilla A_i at time-step i is defined by

$$\mathcal{S}_i(\cdot) := \mathrm{SWAP}_{\mathsf{S}\mathsf{A}_i}(\cdot)\,\mathrm{SWAP}_{\mathsf{S}\mathsf{A}_i} \tag{2.53}$$

where here

$$\mathrm{SWAP}_{\mathsf{S}\mathsf{A}_i} := \sum_{i,j=1}^{d_{\mathsf{S}}} \mathcal{I}_{\mathsf{E}} \otimes |i\rangle\langle j| \otimes \mathcal{I}_{\mathsf{A}_1\mathsf{B}_1\cdots\mathsf{A}_{i-1}\mathsf{B}_{i-1}} \otimes |j\rangle\langle i| \otimes \mathcal{I}_{\mathsf{B}_i\mathsf{A}_{i+1}\mathsf{B}_{i+1}\cdots\mathsf{A}_k\mathsf{B}_k}, \tag{2.54}$$

and as in the case of the superchannel, the Choi state of the process tensor is manifestly positive by definition.

Similar to the case of the process tensor, we will commonly denote simply by Υ a k-step process tensor unless it is relevant to denote explicitly the number of time-steps.

To see that this reproduces correctly the action of the process tensor, we can contract with the Choi state of the collection of time-ordered operations $\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_{k-1}$: these are uncorrelated, so the full Choi state is simply a tensor product of Choi states.

We can illustrate this for $k = 2$ and the generalization follows trivially. First let us expand the swaps and the maximally entangled states as we did in Eq. (2.32) for the superchannel, i.e.

$$\Upsilon_{2:0} = \sum_{\alpha,\ldots,\delta} \mathrm{tr}_{\mathsf{E}} \left[U_2 \, \mathfrak{S}_{\alpha_2\beta_2} \, U_1 \, \mathfrak{S}_{\alpha_1\beta_1} \, \rho_{\mathsf{SE}} \, \mathfrak{S}_{\delta_1\gamma_1} \, U_1^\dagger \, \mathfrak{S}_{\delta_2\gamma_2} \, U_2^\dagger \right] \otimes |\beta_1\alpha_1\beta_2\alpha_2\rangle\langle\delta_1\gamma_1\delta_2\gamma_2|, \tag{2.55}$$

so that, contracting with $\Upsilon^{\mathrm{T}}_{\vec{\mathcal{A}}_{1:0}} := \Upsilon^{\mathrm{T}}_{\mathcal{A}_0} \otimes \Upsilon^{\mathrm{T}}_{\mathcal{A}_1}$, where here similarly the single operation Choi states are defined by $\Upsilon_{\mathcal{A}_i} = (\mathbb{1}_{\mathsf{A}_i} \otimes \mathcal{A}_i)\tilde{\Psi}$, then

$$\begin{aligned}
&\mathrm{tr}_{\mathsf{in}}[\Upsilon_{2:0}(\mathbb{1}_{\mathsf{out}} \otimes \Upsilon^{\mathrm{T}}_{\vec{\mathcal{A}}_{1:0}})] \\
&= \sum_{\alpha,\ldots,\delta} \mathrm{tr}_{\mathsf{E}} \left[U_2 \, \mathfrak{S}_{\alpha_2\beta_2} \, U_1 \, \mathfrak{S}_{\alpha_1\beta_1} \, \rho_{\mathsf{SE}} \, \mathfrak{S}_{\delta_1\gamma_1} \, U_1^\dagger \, \mathfrak{S}_{\delta_2\gamma_2} \, U_2^\dagger \right] \langle\gamma_1|\mathcal{A}_0^{\mathrm{T}}(|\beta_1\rangle\langle\delta_1|)|\alpha_1\rangle\langle\gamma_2|\mathcal{A}_1^{\mathrm{T}}(|\beta_2\rangle\langle\delta_2|)|\alpha_2\rangle \\
&= \sum_{\alpha,\ldots,\delta} \mathrm{tr}_{\mathsf{E}} \left[U_2 \, \mathfrak{S}_{\alpha_2\beta_2} \, U_1 \, \mathfrak{S}_{\alpha_1\beta_1} \, \rho_{\mathsf{SE}} \, \mathfrak{S}_{\delta_1\gamma_1} \, U_1^\dagger \, \mathfrak{S}_{\delta_2\gamma_2} \, U_2^\dagger \right] \langle\alpha_1|\mathcal{A}_0(|\beta_1\rangle\langle\delta_1|)|\gamma_1\rangle\langle\alpha_2|\mathcal{A}_1(|\beta_2\rangle\langle\delta_2|)|\gamma_2\rangle \\
&= \sum_{\mu} \mathrm{tr}_{\mathsf{E}} \left[U_2 A_{\mu_1} U_1 A_{\mu_0} \rho_{\mathsf{SE}} A_{\mu_0}^\dagger U_1^\dagger A_{\mu_1}^\dagger U_2^\dagger \right] \\
&= \mathrm{tr}_{\mathsf{E}} \left[\mathcal{U}_2 \, \mathcal{A}_1 \, \mathcal{U}_1 \, \mathcal{A}_0(\rho_{\mathsf{SE}}) \right] \\
&= \mathcal{T}_{2:0}[\vec{\mathcal{A}}_{1:0}],
\end{aligned} \tag{2.56}$$

where the trace is over all input spaces, as labelled in Fig. 2.8, and where we used the fact that the operations $\mathcal{A}_i$ are CP by decomposing $\mathcal{A}_i(\cdot) = \sum_{\mu_i} A_{\mu_i}(\cdot)A_{\mu_i}^\dagger$. This can be checked similarly for any number of time-steps k, so that

$$\mathcal{T}[\vec{\mathcal{A}}_{k-1:0}] = \mathrm{tr}_{\mathsf{in}} \left[\Upsilon \left(\mathbb{1}_{\mathsf{out}} \otimes \Upsilon^{\mathrm{T}}_{\vec{\mathcal{A}}_{k-1:0}} \right) \right], \tag{2.57}$$

for any k-step process $\mathcal{T}$ with Choi state Υ.

Now, with the Choi state at hand we can also readily verify that the process tensor is TP, i.e.

$$\begin{aligned}
\mathrm{tr}_{\mathsf{out}}[\Upsilon] &= \sum_{\alpha,\ldots,\delta} \mathrm{tr} \left[U_k \, \mathfrak{S}_{\alpha_k\beta_k} \cdots U_1 \, \mathfrak{S}_{\alpha_1\beta_1} \, \rho_{\mathsf{SE}} \, \mathfrak{S}_{\delta_1\gamma_1} \, U_1^\dagger \cdots \mathfrak{S}_{\delta_k\gamma_k} \, U_k^\dagger \right] |\beta_1\alpha_1 \cdots \beta_k\alpha_k\rangle\langle\delta_1\gamma_1 \cdots \delta_k\gamma_k| \\
&= \sum_{\alpha,\beta} |\beta_1\alpha_1 \cdots \beta_k\alpha_k\rangle\langle\beta_1\alpha_1 \cdots \beta_k\alpha_k| \\
&= \mathbb{1}_{\mathsf{A}_1\mathsf{B}_1\cdots\mathsf{A}_k\mathsf{B}_k} \\
&= \mathbb{1}_{\mathsf{in}},
\end{aligned} \tag{2.58}$$

where the trace is over the output in system S, thus whenever the operations $\mathcal{A}_i$ are TP, the action of the process tensor on these will yield a properly normalized quantum state.

The final ingredient we require from the process tensor is a containment property (which amounts to a *causality* property [89]) and more generally a consistency condition. Now, in principle this is akin to doing a marginalization, however, when we did this for classical processes we implicitly made the assumption that there was a single way to probe the system in question and that this probing had no influence at all in the process. So, even if classically summing over all events amounts to doing nothing, we cannot do the same in the quantum case. So, for example,

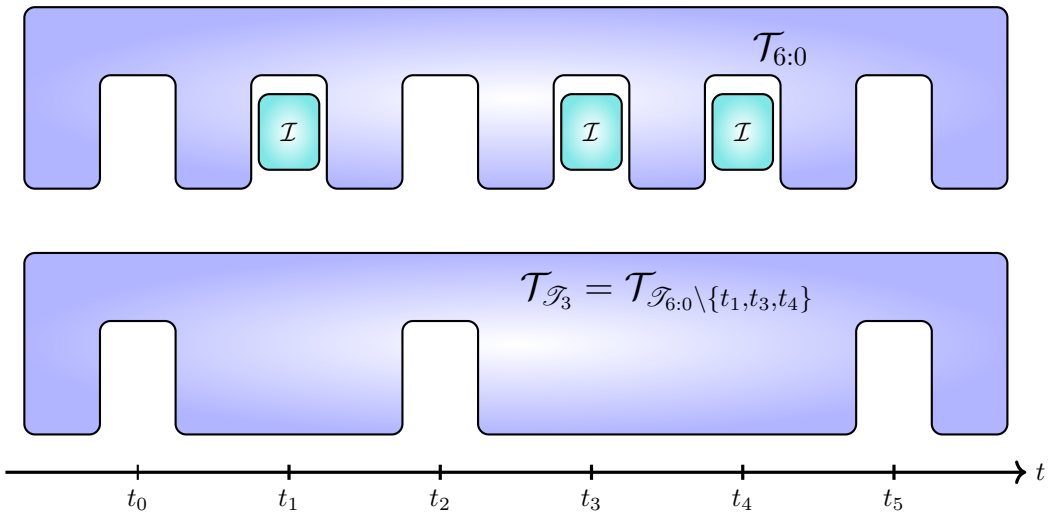


Figure 2.10: **Process tensor containment property (example):** A 6-step process tensor $\mathcal{T}_{6:0}$ contains the 3-step process tensor $\mathcal{T}_{\mathscr{T}_3}$ where $\mathscr{T}_3 = \{t_0, t_2, t_5\}$; this can be seen by applying identity operations $\mathcal{I}$ at timesteps t_1, t_3, t_4. The overarching idea is that there exists a unique maximal description containing any process on a reduced number of time-steps [154].

the sum over outcomes of a measurement $\mathcal{B}$ in a basis $\{|\beta\rangle\}$ is given by $\mathcal{B}(\rho) = \sum_\beta |\beta\rangle\langle\beta|\rho|\beta\rangle\langle\beta|$, thus a state being classical corresponds to it being diagonal in the basis $\{|\beta\rangle\}$ and $\mathcal{B}(\rho) = \rho$, however when ρ contains non-diagonal non-zero elements this stops being true.

The containment property clearly holds for the process tensor by replacing the relevant operations with identities, however, it was recently proved that consistency in general holds not only for the process tensor as well, but for any stochastic theory [154]. This result serves to define quantum stochastic processes by means of the process tensor in an unambiguous way. For our purposes, we should notice that indeed smaller process tensors are contained in bigger ones in the sense that

$$\mathcal{T}_{\mathscr{T}_\ell}\left[\vec{A}_{\mathscr{T}_\ell}\right] = \mathcal{T}_{\mathscr{T}_k}\left[\vec{A}_{\mathscr{T}_\ell} \cup \mathcal{I}_{\mathscr{T}_k\setminus\mathscr{T}_\ell}\right], \qquad \forall \mathscr{T}_\ell, \mathscr{T}_k : |\mathscr{T}_\ell| < |\mathscr{T}_\ell|, \tag{2.59}$$

where $\mathscr{T}_n$ is a discrete set of times as defined in Eq. (2.36). This can also be visualized easily through Fig. 2.10.

Finally, let us draw a more precise picture in relation with the discussion of the classical case. A quantum event x_i at the i^{th} time-step corresponds to an outcome, i.e. a random variable, of the corresponding intervention, which is a `CPTNI` map $\mathcal{A}^{(x_i)}(\cdot) := \sum_\nu A_\nu^{(x_i)}(\cdot)A_\nu^{(x_i)\,\dagger}$ with Kraus operators $\{A_\nu^{(x_i)}\}$ satisfying $\sum_\nu A_\nu^{(x_i)}A_\nu^{(x_i)\,\dagger} \leq \mathbb{1}$. More generally, an intervention corresponds to the action of an *instrument* $\mathcal{J} = \{\mathcal{A}^{(x_i)}\}$, which overall yields a `CPTP` map when summed over possible outcomes $\sum_{x_i} \mathcal{A}^{(x_i)} := \mathcal{A}^{\mathcal{J}}$ [155], naturally generalizing the concept of a `POVM`.

Thus, if an experimenter applies the sequence of `CP` maps $\mathcal{A}_0^{(x_0)}, \mathcal{A}_1^{(x_1)}, \ldots, \mathcal{A}_{k-1}^{(x_{k-1})}$, each being an element of a corresponding instrument $\mathcal{J}_0, \mathcal{J}_1, \ldots, \mathcal{J}_{k-1}$, and measures the final state $\rho_{\mathsf{S}}^{(k)}$ with a `POVM` through an instrument $\mathcal{J}_k = \left\{\mathcal{A}_k^{(x_k)} = \mathrm{M}_k^{(x_k)}\right\}$, the probability to observe a sequence of

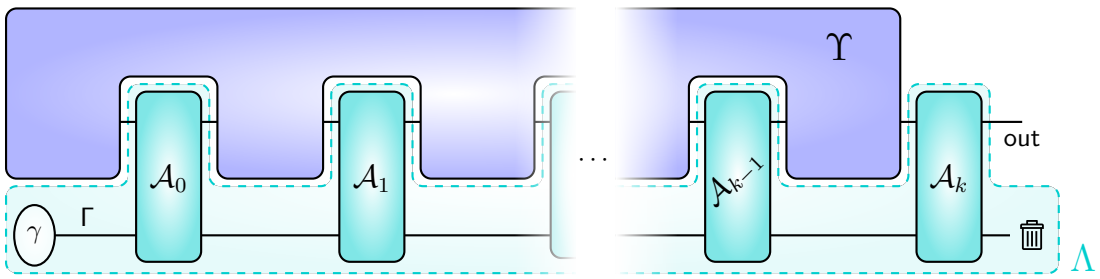


Figure 2.11: **A set of correlated operations** $\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_{k-1}\}$ followed by a POVM element $\mathcal{A}_k = \mathrm{M}_k$ acting on a joint $\mathsf{S}\Gamma$ system, where Γ is an ancillary system initialized in state γ, can be described as a process-tensor like object that is CP with Choi state Λ. Here we equivalently denote the process tensor as Υ.

quantum events is given by

$$\begin{aligned}\mathbb{P}\left(\vec{x}_{k:0}|\vec{\mathcal{J}}_{k:0}\right) &= \mathrm{tr}\left[\mathcal{A}_k^{(x_k)}\mathcal{U}_k\mathcal{A}_{k-1}^{(x_{k-1})}\mathcal{U}_{k-1}\ \ldots\mathcal{A}_1^{x_1}\mathcal{U}_1\,\mathcal{A}_0^{x_0}(\rho_{\mathsf{SE}})\right]\\ &= \mathrm{tr}\left\{\mathcal{A}_k^{(x_k)}\,\mathcal{T}[\vec{\mathcal{A}}_{k-1:0}^{(x_{k-1:0})}]\right\},\end{aligned} \tag{2.60}$$

where here $\vec{\mathcal{J}}_{k:0} = \{\mathcal{J}_0, \mathcal{J}_1, \ldots, \mathcal{J}_k\}$, so that the process tensor contains the probability distributions for all possible measurements, naturally generalising the classical setting described above.

This can be neatly rewritten by means of the corresponding Choi states, clearly separating the influence of the environment from that of the interventions, in a multitime generalization of the Born rule [132, 156, 157]:

$$\mathbb{P}\left(\vec{x}_{k:0}|\vec{\mathcal{J}}_{k:0}\right) = \mathrm{tr}\left[\,\Upsilon\Lambda^{\mathrm{T}}\,\right], \tag{2.61}$$

where we have now defined Λ as the Choi state of the interventions $\left\{\mathcal{A}_i^{(x_i)}\right\}_{i=0}^{k}$. Notice that Λ itself can be thought of having a process-tensor-like structure, here with a set of uncorrelated operations between time-steps. So in fact, we can generalize this to a situation where the experimental interventions are correlated with one another by means of an ancillary space Γ, equivalent to an environment, albeit one that the experiment has access to. We can visualize this as a comb contracting with a process tensor as in Fig. 2.11.

Finally, we can more generally consider the operations Λ as an element of a so-called *tester* [158], which is a set of operation combs that overall lead to a deterministic operation. That is, explicitly denoting $\Lambda_{k:0}^{(x_{k:0})} = \Upsilon_{\vec{\mathcal{A}}_{k:0}^{(x_{k:0})}}$, then a tester is given by the set $\left\{\Lambda_{k:0}^{(x_{k:0})}\right\}$, where each element is positive and where the sum, $\Lambda_{k:0}^{\vec{\mathcal{J}}_{k:0}} := \sum_{x_i}\Lambda_{k:0}^{(x_{k:0})}$, adds up to a CPTP map, $\mathrm{tr}_{\mathsf{in}}[\Lambda_{k:0}^{\vec{\mathcal{J}}_{k:0}}] = \mathbb{1}_{\mathsf{out}}$. In a nutshell, this also generalizes the concept of instrument, i.e. what a tester element is to a tester, a CP map is to an instrument and a POVM element is to a POVM.

The topic of quantum stochastic processes is rapidly developing and by no means are we making an exhaustive presentation in this thesis. While many questions date back to several

decades, with noticeable approaches by Accardi [76, 77] and Lindblad [155], similar to the case of the foundations of statistical mechanics, the topic has experienced a renewed interest from different perspectives (and names), with a great deal of advancement in recent years [156–169]. As we will see, in a sense, with this thesis we have contributed to this program from a front that we can directly relate with equilibration and typicality. To draw this connection let us then discuss how the process tensor naturally generalizes the concept of Markovianity and accounts for a non-ambiguous measure of non-Markovianity.

2.4.4 The quantum Markov condition

As we have now argued, the standard approach to open quantum dynamics by means of dynamical maps offers a necessary but not sufficient condition for Markovianity. Attempts to define what a quantum Markovian process is have been made [78, 170] and several *measures* of non-Markovianity have been proposed e.g. based on divisibility of dynamical maps [171, 172], positivity [173–176], or on trace distance, positing that this distinguishability for two initial quantum states must be monotonically decreasing over time, essentially implying that if one can distinguish better these two states it is because some information is flowing back from the environment [177]. All of these are valid witnesses, however, they have been seen to be inconsistent with each other [178], disagreeing on whether a process is Markovian and/or on the degree of non-Markovianity.

As we have stressed before, at its core a quantum Markov condition has to consider multitime correlations and approaches based on dynamical maps or generally quantum channels capture by definition only two-point correlations.

Let us begin by writing the conditional probability of getting an outcome x_k given outcomes $\vec{x}_{k-1:0}$, all with respective instruments $\vec{\mathcal{J}}_{k:0}$ analogous to Eq. (2.41), i.e.

$$\mathbb{P}(x_k|\vec{x}_{k-1:0};\vec{\mathcal{J}}_{k:0}) = \frac{\mathbb{P}(\vec{x}_{k:0}|\vec{\mathcal{J}}_{k:0})}{\mathbb{P}(\vec{x}_{k-1:0}|\vec{\mathcal{J}}_{k-1:0})}, \tag{2.62}$$

where the numerator is given by Eq. (2.60) and similarly for the denominator, $\mathbb{P}(\vec{x}_{k-1:0}|\vec{\mathcal{J}}_{k-1:0}) = \mathrm{tr}\left\{\mathcal{A}_{k-1}^{(x_{k-1})}\mathcal{T}_{k-1:0}\left[\vec{\mathcal{A}}_{k-2:0}^{(\vec{x}_{k-2:0})}\right]\right\}$. The main issue for defining a Markov condition lies in the invasiveness the operations $\mathcal{A}_i^{(x_i)}$. Notice that the denominator of Eq. (2.62) won't necessarily capture all of the information of the output state after interrogation at time-step $k-1$, while this will affect the numerator at time-step k.

The main idea in Ref. [152] to resolve this is called a *causal break*. Essentially, the history dependence of a process at time-step k can be checked by fixing its state at such time-step and analyzing its future. This can be described as follows: an experimenter performs $k-1$ operations, and at time-step k performs a measurement, given by a POVM $\{\mathrm{M}^{(m_k)}\}$, and independently reprepares the system in an independent known state, $\sigma_{\mathsf{S}}^{(s_k)}$. Here the labels m_k and s_k just distinguish the measured outcome and the input for the fresh state. What this achieves is to break the information flow on subsystem S, while leaving it in a known state, thus allowing to condition future statistics on this state. Put differently, after a causal break the SE state will be

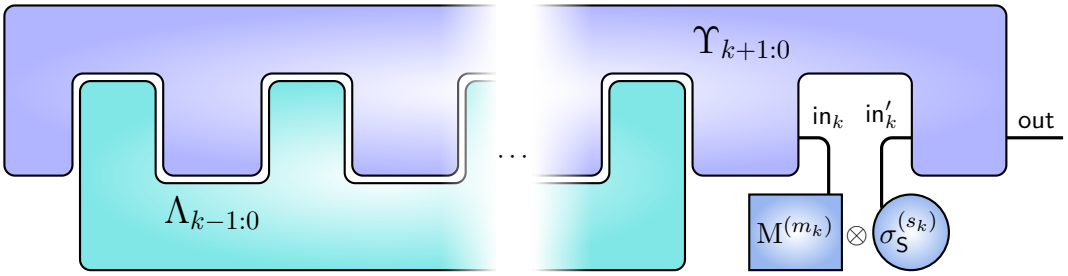


Figure 2.12: **Causal break** by means of an instrument $\mathcal{J}_k^{\mathcal{B}} = \left\{\mathrm{M}_k^{(m_k)} \otimes \sigma_{\mathsf{S}}^{(s_k)}\right\}$ with $\mathrm{M}_k^{(m_k)}$ a POVM element with outcome m_k and $\sigma_{\mathsf{S}}^{(s_k)}$ an independent input state labelled by s_k, acting at time-step k on a process $\Upsilon_{k+1:0}$ which previously acts on an operation process $\Lambda_{k-1:0}$. Quantum Markovianity refers to an independence at any future time-step $k+1$ from the previous operations $\Lambda_{k-1:0}$ and the outcome m_k; any dependence on these will point to information being carried through the environment on $\Upsilon_{k+1:0}$ and in such case the process is non-Markovian.

in a product form, and independent of the previous state of S, however, the E part an explicitly depend on the state $\rho_{\mathsf{SE}}^{(k)}$ before the causal break, and as such on all operations $\vec{\mathcal{A}}_{k-1:0}^{(\vec{x}_{k-1:0})}$ on S before the causal break.

Specifically, in terms of an instrument, this means a causal break at time-step k is given by

$$\mathcal{J}_k^{\mathcal{B}} := \left\{\mathcal{B}_k^{(x_k)} := \mathrm{M}^{(m_k)} \otimes \sigma_{\mathsf{S}}^{(s_k)}\right\}, \tag{2.63}$$

which can be depicted as in Fig. 2.12. In general any operation with an input independent of its output will constitute a causal break, with the respective Choi states being uncorrelated. After a causal break the joint SE state is rendered in a product form, with the state S only depending on the label s_k but not on the previous $\rho_{\mathsf{SE}}^{(k)}$. That is, the causal break acts on the SE state as

$$\left[\mathcal{B}_k^{(x_k)} \otimes \mathcal{I}_{\mathsf{E}}\right] \rho_{\mathsf{SE}}^{(k)} = \sigma_{\mathsf{S}}^{(s_k)} \otimes \mathrm{tr}_{\mathsf{S}}\left[\rho_{\mathsf{SE}}^{(k)} \left(\mathrm{M}^{(m_k)} \otimes \mathbb{1}_{\mathsf{E}}\right)^{\mathrm{T}}\right], \tag{2.64}$$

so the environment, however, can still depend on $\rho_{\mathsf{SE}}^{(k)}$ and the full operation process $\Lambda_{k-1:0}$, thus if we are able to distinguish any two distinct operation tensors, $\Lambda_{k-1:0} \neq \Lambda'_{k-1:0}$ and two different POVM outcomes $m_k \neq m'_k$, the environment must have carried the memory allowing us to do so.

We can then introduce the following:

Definition 2.4.3 (Quantum Markov condition [152, 157])**.** A $(k+1)$-step quantum process $\Upsilon_{k+1:0}$ is Markovian if the statistics with respect to any measurement $\mathcal{J}_{k+1}$ after a causal break $\mathcal{J}_k^{\mathcal{B}} := \left\{\mathrm{M}^{(m_k)} \otimes \sigma_{\mathsf{S}}^{(s_k)}\right\}$ are independent of outcome m_k and all historic outcomes $\vec{x}_{k-1:0}$ with all possible historic instruments $\vec{\mathcal{J}}_{k-1:0}$,

$$\mathbb{P}\left(x_{k+1}|m_k, s_k, \vec{x}_{k-1:0}; \mathcal{J}_{k+1}, \mathcal{J}_k^{\mathcal{B}}, \vec{\mathcal{J}}_{k-1:0}\right) = \mathbb{P}\left(x_{k+1}|s_k; \mathcal{J}_{k+1}, \mathcal{J}_k^{\mathcal{B}}\right), \tag{2.65}$$

for any given state $\sigma_{\mathsf{S}}^{(s_k)}$ prepared in the causal break.

Now analogous to the classical case, if Eq. (2.65) is not satisfied, the quantum process is called non-Markovian. This provides an unambiguous method to witness memory effects. Importantly, the *history* of a process, given by the operation process $\Lambda_{k-1:0}$ with instruments $\vec{\mathcal{J}}_{k-1:0}$ and the measurement associated to $\mathrm{M}^{(m_k)}$ is the generalization of a *trajectory* for a quantum process [179].

Now, if we fix the choice of instruments at each time-step and these consist of causal breaks, $\mathcal{J}_i^{\mathcal{B}} = \left\{\mathcal{B}_i^{(x_i)}\right\}$ for $i = 0, 1, \ldots, k$, then Eq. (2.65) implies the classical Markov condition $\mathbb{P}\left(x_{k+1}|\vec{x}_{k:0}\right) = \mathbb{P}(x_{k+1}|x_k)$ as in Eq. (2.40). Two other consequences that are shown in Ref. [152] are that Markovian dynamics are divisible, within the time-steps where they are defined, as in Eq. (2.24) (with the converse, as discussed for the classical case, not being true), and that whenever the definitions e.g. in Ref. [171–178] predict non-Markovianity, Eq. (2.65) will also predict non-Markovianity, while the converse is not true. As stressed before, many of the shortcomings in other approaches boil down to having necessary but not sufficient conditions for Markovianity.

Finally, the concept of a causal break simply gives us an operationally non-ambiguous way to characterize a quantum process as Markovian but whether this is the case should not depend on such concept; precisely, the process tensor was conceived around the idea of separating experimentally accessible quantities from those intrinsic to the dynamical process. As it turns out, the process tensor gives a straightforward way to classify Markovian processes: given that these have no temporal correlations and are divisible, this implies that their Choi state is simply a tensor product of dynamical maps connecting adjacent time-steps. More specifically, notice that pairs of subsystems of the Choi state Υ in Fig. 2.9 correspond to the different in and in' spaces at each time-steps of $\mathcal{T}$, so that the Choi state encodes temporal correlations as spatial correlations.

This means that all Markovian processes have the form [83, 152]

$$\Upsilon^{(\mathrm{M})} = \mathcal{Z}_{k:k-1} \otimes \mathcal{Z}_{k-1:k-2} \otimes \cdots \mathcal{Z}_{1:0} \otimes \rho_{\mathsf{S}}^{(0)}, \tag{2.66}$$

for dynamical CPTP map Choi states $\mathcal{Z}_{j:i}$ connecting adjacent time-steps i to j. A formal proof, showing that both Eq. (2.66) implies Eq. (2.65) and vice-versa, can be seen in Ref. [90].

In particular for dynamical maps $\mathscr{Z}_{j:i}(\cdot) = \mathrm{tr}_{\mathsf{E}}[\mathcal{U}_{j:i}(\cdot \otimes \varepsilon_j)]$, with unitary maps $\mathcal{U}_{j:i}$ dictating evolution from time t_i to time t_j, and fiducial E states ε_j, this implies that strictly Markovian dynamics in open quantum systems are those in which the environment absolutely forgets at every step, i.e. when the environment is completely discarded, or traced out, and the joint dynamics at each state has no memory whatsoever of previous E states. We can visualize this as in Fig. 2.13.

2.4.5 An unambiguous measure for non-Markovianity

The previous discussion makes it clear that strictly Markovian processes cannot be physically realised by continuous open quantum system dynamics, or, in other words, all open quantum evolutions generated by a time-independent SE Hamiltonians are non-Markovian according to the criterion in Eq. (2.65).

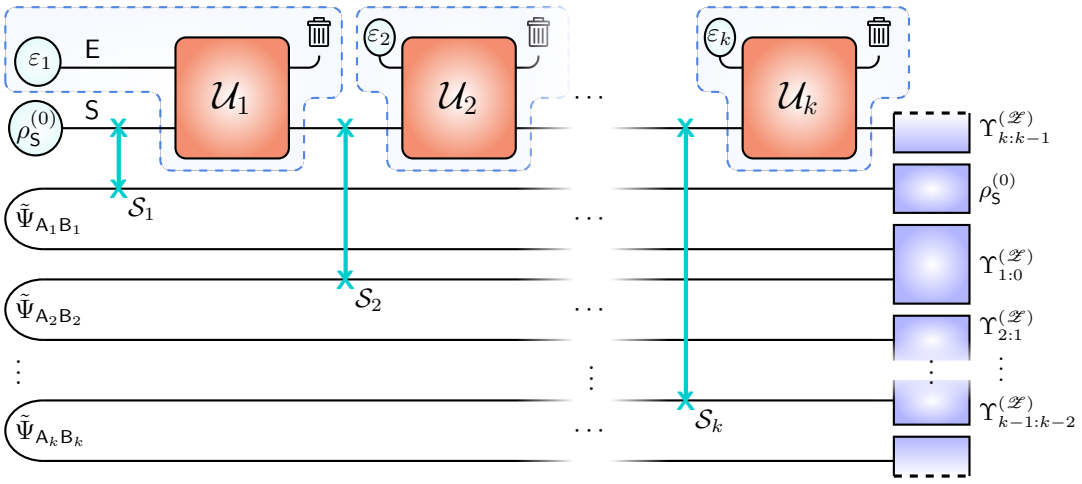


Figure 2.13: **The process tensor for a Markov process** has no temporal correlations and its Choi state takes a tensor product form of maps connecting adjacent time-steps. For an open quantum evolution through a dynamical map $\mathscr{Z}_{j:i}(\cdot) = \mathrm{tr}_{\mathsf{E}}[\mathcal{U}_{j:i}(\cdot \otimes \varepsilon_j)]$ from time-step i to j, it takes a form equivalent to $\Upsilon_{k:0} = \Upsilon_{k:k-1}^{(\mathscr{Z})} \otimes \cdots \otimes \Upsilon_{1:0}^{(\mathscr{Z})} \otimes \rho_{\mathsf{S}}^{(0)}$, where $\Upsilon_{j:i}^{(\mathscr{Z})}$ is the Choi state of $\mathscr{Z}_{j:i}$. The dashed lines represent correlation between the pair of spaces where $\Upsilon_{k:k-1}^{(\mathscr{Z})}$ acts on.

Far from this feature rendering the process tensor framework useless, it provides a clear and non-ambiguous way to measure non-Markovianity. Consider any relevant distinguishability measure Δ satisfying Def. 0.4.1. Then we can quantify the non-Markovianity of any process Υ by measuring *how far* it is from the closest Markovian process,

$$\mathcal{N}_\Delta := \min_{\Upsilon^{(\mathrm{M})}} \Delta\left(\Upsilon,\, \Upsilon^{(\mathrm{M})}\right), \tag{2.67}$$

where the minimum is taken to represent *the nearest* Markovian process.

The choice of the distance measure Δ will depend on the problem at hand, the operational meaning that one is looking for, or simply computational convenience which e.g. can then be used to place bounds on a different measure of interest. Furthermore, this choice can be relaxed to be any *pseudo-distance* [152] respecting only condition *i.* of Def. 0.4.1 and being CP-contractive, which means $\Delta(\Phi(\rho), \Phi(\sigma)) \leq \Delta(\rho, \sigma)$ for any CP map.

Example 2.4.4. To exemplify, one such pseudo-distance is given by the relative entropy,

$$S(\rho \,\|\, \sigma) = \mathrm{tr}[\rho\,(\log\rho - \log\sigma)], \tag{2.68}$$

which is read as *the relative entropy of ρ with respect to σ*, which satisfies positivity but not the symmetry nor triangle inequality properties, but it is nevertheless a valid measure of the separation of a pair of states.

The relative entropy has several useful properties and has been widely studied, having a close connection as a generalization from classical entropy measures [6, 7]; in particular, the classical relative entropy of two probability distributions is related to the probability of distinguishing

these after a large (finite) number of independent samples. This gives a large deviations bound known as Sanov's theorem [180, 181], which can be generalized to the quantum case [7, 182, 183] as the probability of confusing two quantum states ρ, σ after a large number of realizations n as

$$\mathbb{P}_{\text{confusion}} = \exp\left[-n\, S(\rho \,\|\, \sigma)\right]. \tag{2.69}$$

This means that with the relative entropy, the measure $\mathcal{N}_S = \min_{\Upsilon^{(\mathrm{M})}} S(\Upsilon \| \Upsilon^{(\mathrm{M})})$ gives an operational meaning through the probability of confusing a quantum process for a Markovian one decreasing exponentially in the number of realizations of the process, $\mathbb{P}^{(\mathrm{M})}_{\text{confusion}} = \exp[-n\,\mathcal{N}_S]$.

Furthermore, the Markovian process minimizer of the relative entropy is given simply by the marginals (i.e. the reduced components) of the process tensor. We can see this as follows. First, given that we can write $\mathcal{N}_S = \min_{\Upsilon^{(\mathrm{M})}} \{-\operatorname{tr}[\,\Upsilon\, \log\, \Upsilon^{(\mathrm{M})}] - S(\Upsilon)\}$ with $S(\Upsilon)$ the von Neumann entropy of Υ, we have

$$\begin{aligned}
-\operatorname{tr}[\,\Upsilon\, \log\, \Upsilon^{(\mathrm{M})}] &= -\operatorname{tr}\left\{ \Upsilon\, \log\left[\bigotimes_{\ell=1}^{k} \mathcal{Z}_{\ell:\ell-1} \otimes \rho_{\mathsf{S}}^{(0)} \right] \right\} \\
&= -\sum_{i=1}^{k} \operatorname{tr}\left\{ \Upsilon\, (\log[\mathcal{Z}_{i:i-1} \otimes \mathbb{1}]) \right\} - \operatorname{tr}\left\{ \Upsilon \left(\log[\mathbb{1} \otimes \rho_{\mathsf{S}}^{(0)}] \right) \right\} \\
&= -\sum_{i=1}^{k} \operatorname{tr}\left(\Upsilon_{i:i-1} \log \mathcal{Z}_{i:i-1} \right) - \operatorname{tr}\left(\Upsilon_0 \log \rho_{\mathsf{S}}^{(0)} \right),
\end{aligned} \tag{2.70}$$

where in the second line the identities are in all remaining spaces and in the third line $\Upsilon_{\ell:\ell-1} = \operatorname{tr}_{\overline{\ell:\ell-1}}(\Upsilon)$ is the reduced state of Υ obtained by tracing all subsystems except the ones corresponding to dynamics from time-step $\ell-1$ to time-step ℓ, and similarly for $\Upsilon_{\overline{0}} = \operatorname{tr}_{\overline{0}}(\Upsilon)$ just over the zero$^{\text{th}}$ time-step; these are what we refer to as the *marginals* of Υ.

So the minimization in $\mathcal{N}_S$ amounts to minimizing the individual relative entropies, as $-\operatorname{tr}[\rho \log \sigma] = S(\rho\|\sigma) + S(\rho)$, with the minimum occurring for $\rho = \sigma$; thus we can conclude

$$\Upsilon_S^{(\mathrm{M})} = \Upsilon_{k:k-1} \otimes \Upsilon_{k-1:k-2} \otimes \cdots \otimes \Upsilon_{1:0} \otimes \rho_{\mathsf{S}}^{(0)}, \tag{2.71}$$

is the closest Markovian process under relative entropy.

For other measures the minimization can potentially be non-trivial given that the Choi state is a d_{S}^{2k+1} dimensional many-body state. Despite this, bounds and different relations are usually accessible between different distinguishability measures. In the following we will see an extensive use and application of the measure of non-Markovianity in Eq. (2.67) to derive the main results of this PhD. Determining the degree of Markovianity of general quantum processes is not only a foundational task, with an ever-increasing interest and relevance in determining the breakdown of the Markovian approximation in modern experiments [118, 130, 134, 135].

We now proceed to presenting these results.

PART II

Main Results

CHAPTER 3

Equilibration in Quantum Processes across Multiple Points in Time

We all behave like Maxwell's demon [...] *We disturb the tendency toward equilibrium.*

– James Gleick (*The Information: A History, a Theory, a Flood*)

As we described in Chapter 1, the seemingly paradoxical nature of the unitary evolution of quantum systems towards equilibrium can be resolved by means of the concept of equilibration on average. We saw that in essence, the mechanism behind equilibration is that of dephasing, as illustrated in Fig. 1.1. In particular, if we consider an open system S, we have seen in Chapter 2 that it might be the case that correlations in time or memory, manifesting in the statistics of sequential observations, maintain information about the initial perturbation. That is, put simply, it is unclear whether quantum stochastic processes equilibrate in an analogous way, i.e. whether they are most often found close to some average value.

The first problem we face when trying to draw a connection with equilibration on average on multitime processes is trying to pin down what exactly we mean by a process equilibrating, or introducing a notion of an *equilibrium process* and a proper multitime *timescale* on which a process would equilibrate on average within finite time-windows. This is not an easy task but here we elaborate on the notion we gave in Ref. [1]. In our work, we derived sufficient conditions for general multitime observations to relax close to their corresponding equilibrium values at each time-step when the corresponding operations are implemented with an imperfect, *fuzzy* clock, or equivalently, on a system with uniformly fluctuating energies.

While our approach to equilibration is entirely general, ultimately the idea is that there might be a similar connection between equilibration in quantum processes and the dynamical emergence of Markovianity, just as there is between equilibration on average of quantum systems and thermalization. In a sense it is expected that a Markovianization is a stronger condition (i.e. more constrained) than multitime process equilibration, just as thermalization is to system equilibration. Making this connection explicit (or for that matter making a different connection) will have to be done in the future, but as we stated before, it is important both for foundational

and practical reasons knowing how and when does quantum processes dynamically Markovianize and it is mainly in this spirit that we explore how they equilibrate.

3.1 Equilibration due to finite temporal resolution

The approach of the results studied in Section 1.2.3 focuses on the temporal fluctuations of the expectation values of A around equilibrium, in essence with a small variance relating to the expectation value of A concentrating around its mean. As we saw, this gives a statistically meaningful characterization of how equilibration is dynamically achieved.

To explore equilibration on multitime processes, we took a similar approach but one focusing on procedures with a clear operational meaning. Precisely, one can picture a situation where an experimenter can implement operations with some finite temporal resolution only, i.e. where they have some uncertainty in the readings of their clock. In the words of the previous chapter, this *fuzziness* can be associated to the instruments describing a set of operations at a given time in a process, with the fuzziness described by a given probability distribution. Thus we are asking how different an evolving quantum state appears from equilibrium when measured at a time that can vary in each realisation, being randomly drawn from some distribution that quantifies the fuzziness associated with finite temporal resolution.

To precisely define what we mean fuzziness, let us first define a *probability density function* (PDF), which will let us generalize the idea of finite temporal resolution. We now treat the time variables t_i, i.e. the waiting time between the i^{th} and the $(i+1)^{\text{th}}$ interventions, as a non-negative real random variable, which now, however, are continuous rather than discrete (as opposed to the presentation in Section 2.4.1). Let us focus on a single time t, analogous to the standard equilibration case, for now. This means that our sample space is the whole non-negative real line $\mathbb{R}_0^+$, with events corresponding to intervals within it. In the discrete setting we employed random variables and their distributions, however the question now is how precisely do we define the distribution of a continuous random variable? A way to do this is to specify probabilities that the random variable will be within a given interval, rather than that of it taking a possible value. We can achieve this with the following.

Definition 3.1.1 (Probability Density Function)**.** A probability density function (PDF) for a real random variable X is a function $\mathscr{P} : \mathbb{R} \to \mathbb{R}_0^+$ such that

$$\mathbb{P}(a \leq X \leq b) = \int_a^b \mathscr{P}(x)\, dx, \tag{3.1}$$

for all $a \leq b$, and

$$\mathbb{P}(X \in \mathbb{R}) = \int_{-\infty}^{\infty} \mathscr{P}(x)\, dx = 1. \tag{3.2}$$

This definition of course can be generalized; in the case of finite distributions, for example, the corresponding quantity is the probability mass function.

Thus specifically by a finite temporal resolution observation we mean an observable A (either on subsystem S or acting coarsely on SE) measured after a time $t > 0$ sampled from a one-parameter family of probability distributions with density function $\mathscr{P}_T$, i.e. which is such that $\int_0^\infty dt\, \mathscr{P}_T(t) = 1$. The parameter T represents the uncertainty or fuzziness of the distribution; for example, it could be associated with the variance of the distribution. With this definition, we may generalize the time-average over a time-window T of a given time-dependent quantity f by

$$\overline{f}^{\mathscr{P}_T} := \int_0^\infty dt\, \mathscr{P}_T(t)\, f(t), \tag{3.3}$$

so that the uniform average considered in Ref. 1.2.3 corresponds to the case $\mathscr{P}_T = T^{-1}$.

Let us then consider a dynamical setup as in standard equilibration, focusing on the dynamics of a d_S-dimensional subsystem S of a $d_\mathsf{S} d_\mathsf{E}$-dimensional composite system SE, and refering to subsystem equilibration as the relaxation of S towards some steady state, while the whole SE evolves unitarily via $U = \exp(-iHt)$, with a general time-independent Hamiltonian $H = \sum_{n=1}^{\mathfrak{D}} E_n P_n$. For simplicity, we denote the full SE initial state as ρ.

Then we can define the time average of the initial state over a finite-interval of width T as the fuzzy average

$$\overline{\rho}^{\mathscr{P}_T} := \int_0^\infty dt\, \mathscr{P}_T(t)\, \rho(t) = \int_0^\infty dt\, \mathscr{P}_T(t)\, \mathrm{e}^{-it(E_n - E_m)}\, P_n\, \rho\, P_m, \tag{3.4}$$

so we now need to make sense of the integral of the exponential factor. To do this, we know that in the infinite-time limit the average state *should*, by definition, correspond to the uniform time-averaged state ω. That is, we require

$$\omega := \lim_{T\to\infty} \overline{\rho}^{\mathscr{P}_T} = \mathcal{D}(\rho), \tag{3.5}$$

where $\mathcal{D}$ is the dephasing map with respect to H,

$$\mathcal{D}(\rho) = \sum_{n=1}^{\mathfrak{D}} P_n\, \rho\, P_n, \tag{3.6}$$

which we defined briefly before in Eq. (2.5). This means ω is independent of the choice of the PDF $\mathscr{P}_T$. We can further generalize this dephasing map to a finite-time one as

$$\mathcal{G}_T(\rho) := G^{(T)}_{nm} P_n\, \rho\, P_m, \quad \text{where} \quad G^{(T)}_{nm} := \overline{\mathrm{e}^{-it(E_n - E_m)}}^{\mathscr{P}_T}, \tag{3.7}$$

which we similarly used briefly before only to simplify notation in Fig. 1.1. This means that we can make sense of the integral in Eq. (3.4) by requiring $\mathscr{P}_T$ to be such that

$$\lim_{T\to\infty} G^{(T)}_{nm} = \delta_{nm}, \tag{3.8}$$

which otherwise is entirely general.

We consider then the average distinguishability by means of an observable A between the equilibrium state ω, and the non-equilibrium, fuzzy state $\overline{\rho}^{\mathscr{P}_T}$, which can be quantified as the

difference of expectation values between these states as $\left|\langle A\rangle_{\overline{\rho}^{\mathscr{P}_T}-\omega}\right| = |\operatorname{tr}[A(\overline{\rho}^{\mathscr{P}_T}-\omega)]|$. Given the finite time and fully dephasing operators, we can bound this as

$$\left|\langle A\rangle_{\overline{\rho}^{\mathscr{P}_T}-\omega}\right| = |\operatorname{tr}[A(\mathcal{G}_T-\mathcal{D})(\rho)]| \leq \|A\|\,\|(\mathcal{G}_T-\mathcal{D})(\rho)\|_2\,, \tag{3.9}$$

where we used $|\operatorname{tr}[X\sigma] \leq \|X\|\|\sigma\|_2$, with the Schatten norms defined as in Eq. (19), which can be seen to follow from Hölder's inequality in Eq. (1.11) and the hierarchy of the Schatten norms in Eq. (20). Now we can bound the right-hand-side as

$$\begin{aligned}\|(\mathcal{G}_T-\mathcal{D})(\rho)\|_2^2 =& \operatorname{tr}\left|\sum_{n\neq m} G_{nm}P_n\rho P_m\right|^2 = \sum_{n\neq m}|G^{(T)}_{nm}|^2\operatorname{tr}\left[P_n\rho P_m\rho\right]\\ \leq& \max_{n\neq m}|G^{(T)}_{nm}|^2\sum_{n\neq m}\operatorname{tr}[P_n\rho P_m\rho] = \max_{n\neq m}|G^{(T)}_{nm}|^2\operatorname{tr}(\rho^2-\omega^2)\\ =& \|\rho-\omega\|_2^2\ \max_{n\neq m}|G^{(T)}_{nm}|^2,\end{aligned} \tag{3.10}$$

where in the second equality we implicitly used $P_nP_{n'} = \delta_{nn'}$, and in the last line we used $\operatorname{tr}(\rho^2-\omega^2) = \|\rho-\omega\|_2^2$, given that $\operatorname{tr}(\rho\,\omega) = \operatorname{tr}(\omega^2)$.

Thus it follows that

$$\left|\langle A\rangle_{\overline{\rho}^{\mathscr{P}_T}-\omega}\right| \leq \mathscr{S}_T\,\|A\|\|\rho-\omega\|_2, \quad \text{where} \quad \mathscr{S}_T := \max_{n\neq m}|G^{(T)}_{nm}|, \tag{3.11}$$

with the rate of convergence to zero determined by $\mathscr{S}_T$, which essentially tell us the off-diagonal term that will die off at the slowest rate on average with respect to the fuzzy clock.

In particular, when the fuzziness T corresponds to that of the uniform distribution over an interval of width T as we described in Section 1.2.3, the PDF is $\mathscr{P}_T = T^{-1}$ and we get $|G^{(T)}_{nm}| = |\sin(T\mathcal{E}_{nm})/T\mathcal{E}_{nm}|$ with $\mathcal{E}_{nm} := (E_n-E_m)/2$. The bound in Eq. (3.11) then tells us that the evolved state $\rho(t)$ will differ from the equilibrium ω when measured at a given time with a temporal-resolution T at most with proportion $|T\mathcal{E}_{nm}|^{-1}$ for the smallest energy gap $\mathcal{E}_{nm}$, with a scale set by the size of the observable A and how different the initial state ρ is from the equilibrium ω. Notice as well that in general $\|\omega\|_2^2 \leq d_{\text{eff}}^{-1}(\rho)$, where the inverse effective dimension is defined in Eq. (1.14), with equality both for pure ρ or when the Hamiltonian is non-degenerate; both quantities relate to how spread the initial state ρ is in the energy eigenbasis.

This initial fuzziness can be interpreted as the observer not knowing exactly when the process actually began. However, one question we can ask is whether we are able to overcome the fuzziness of the initial interval by making a sequence of measurements. As we described in Section 2, these operations can correspond to any possible experimental intervention, which can furthermore be correlated with each other through an ancillary system. In this case temporal correlations within the dynamics itself can propagate through the environment and similarly the disturbance introduced by the experimental operations might become relevant. Moreover, any fuzziness in the subsequent measurements also has to be accounted for.

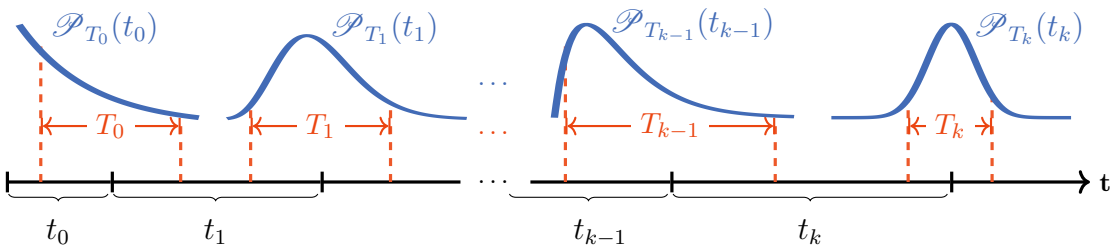


Figure 3.1: **Finite temporal resolution in a quantum process:** between interventions, each Hamiltonian evolution is time-averaged over the waiting times between interventions t_0, $t_1, \cdots, t_k$ with corresponding average waiting times τ_0, $\tau_1, \cdots, \tau_k$, over a probability distribution with PDF given by $\mathscr{P}_{T_i}$, with T_i having a suitable uncertainty parameter role.

3.2 Multitime equilibration due to finite temporal resolution

Of course one might not stop at a single observation but continue gathering data to assess how close the system remains to equilibrium with respect to a set of possible operations, $\{\mathcal{A}_i\}$, given by weighted CPTNI maps $\mathcal{A}_\ell(\cdot) := \sum_\mu a_{\ell_\mu} K_{\ell_\mu}(\cdot) K_{\ell_\mu}^\dagger$, with $\sum_\mu K_{\ell_\mu}^\dagger K_{\ell_\mu} \leq \mathbb{1}$ and $a_{\ell_\mu} \in \mathbb{R}$ being the corresponding outcome weights. More specifically, the scenario is the following: an initial state ρ in the full SE composite evolves unitarily through a time-independent Hamiltonian dynamics until, at time t_0, an operation $\mathcal{A}_0$ is performed jointly on S along with an ancilla Γ, which is initially uncorrelated in state γ. We now denote the full initial state by

$$\varrho := \rho \otimes \gamma, \tag{3.12}$$

and after the first operation, SE evolve unitarily again for a time t_1 until another operation $\mathcal{A}_1$ is made on System-Ancilla (SΓ), and so on for k time-steps. The unitary evolution at each step is given by the map

$$\mathcal{U}_\ell(\cdot) = \exp\{-iH_\ell t_\ell\}(\cdot)\exp\{iH_\ell t_\ell\}, \tag{3.13}$$

acting on SE. The time-independent Hamiltonians H_ℓ can in general be different at each step. The ancillary space Γ can be interpreted as a quantum memory device, and might carry information about previous interactions with the system. As done in Chapter 2.4, we denote the Choi state of the operations $\{\mathcal{A}_i\}$ by Λ and the underlying dynamical process by Υ (with the number of time-steps implicit).

The joint multitime expectation of these set of operations is thus given by

$$\langle\Lambda\rangle_\Upsilon = \mathrm{tr}[\mathcal{A}_k \mathcal{U}_k \cdots \mathcal{A}_0 \mathcal{U}_0(\varrho)], \tag{3.14}$$

where implicitly $\mathcal{A}_i$'s act only on SΓ, while the unitaries $\mathcal{U}_i$ act on SE. For simplicity, we consider a fixed set of projectors $\{P_n\}$ for all Hamiltonians such that $H_\ell = \sum P_n E_{n_\ell}$ at each step ℓ, with P_n projecting onto the energy eigenspaces of H_i with energy E_{n_i}. Also, we denote simply by $\cdot$ the composition of superoperators when clear by context.

We denote the time intervals of the free evolutions as t_j, which is preceded by the j^{th} measurement and followed by $(j+1)^{\text{th}}$ measurement. In other words, t_j is the waiting time

between j^{th} and $(j+1)^{\text{th}}$ measurements. To avoid cluttering of notation we will denote the multitime probability distribution as $\mathscr{P}_{\mathbf{T}}(\mathbf{t}) = \prod_{i=0}^{k} \mathscr{P}_{T_i}(t_i)$, where now we use the labels

$$\mathbf{t} := (t_0, t_1, \ldots, t_k) \quad \text{and} \quad \mathbf{T} = (T_0, T_1, \ldots, T_k) \tag{3.15}$$

for waiting times and the fuzziness parameters for each time interval, respectively. We can now define a finite temporal resolution process as

$$\overline{\Upsilon}^{\mathscr{P}_{\mathbf{T}}} := \int_0^\infty dt_k \cdots \int_0^\infty dt_1 \int_0^\infty dt_0 \, \mathscr{P}_{\mathbf{T}}(\mathbf{t}) \, \Upsilon, \tag{3.16}$$

and we will denote the average waiting time between each pair of measurements by

$$\tau_i := \int_0^\infty dt_i \; t_i \; \mathscr{P}_{T_i}(t_i), \tag{3.17}$$

and we can visualize these concepts pictorially as in Fig. 3.1.

We are interested in quantifying how different this out-of-equilibrium process, where time intervals are fuzzy, looks from an equilibrium process. To define the equilibrium process we follow the lead of earlier results, i.e. the initial state relaxes to the equilibrium state

$$\varpi := \varpi_0 = \lim_{T_0 \to \infty} \int_0^\infty \mathcal{U}_0(\varrho) \, \mathscr{P}_{T_0}(t_0) \, dt_0 = \mathcal{D}(\varrho) = \omega \otimes \gamma, \tag{3.18}$$

until an operation $\mathcal{A}_0$ is performed, and subsequently the system relaxes again to an equilibrium state $\varpi_1 = \mathcal{D}\mathcal{A}_0(\varpi)$ until an operation $\mathcal{A}_1$ is made, and so on for k-time-steps. Then we can define

$$\varpi_i := \lim_{T_i \to \infty} \int_0^\infty \mathcal{U}_i \mathcal{A}_{i-1}(\varpi_{i-1}) \, \mathscr{P}_{T_i}(t_i) \, dt_i = \mathcal{D}\mathcal{A}_{i-1} \cdots \mathcal{A}_0 \mathcal{D}(\varrho), \tag{3.19}$$

for any $i = 0, \cdots, k$ as the equilibrium states after each intervention up to $\mathcal{A}_{i-1}$.

This is a sensible definition for the intermediate equilibrium states, which, however, is dependent on each operation $\mathcal{A}_j$. We can, however, define the equilibrium quantum process independently of the operations as

$$\Omega := \lim_{\mathbf{T} \to \infty} \overline{\Upsilon}^{\mathscr{P}_{\mathbf{T}}}, \tag{3.20}$$

which is depicted in Fig. 3.2 as a set of dephasing maps $\mathcal{D}$ at each timestep. This means then that we can write

$$\langle \Lambda \rangle_\Omega = \mathrm{tr}[\Lambda \Omega] = \mathrm{tr}[\mathcal{A}_k \varpi_k] = \mathrm{tr}[\mathcal{A}_k \mathcal{D} \mathcal{A}_{k-1} \mathcal{D} \cdots \mathcal{A}_0 \mathcal{D}(\varrho)], \tag{3.21}$$

for the expectation of a sequence of operations $\{\mathcal{A}_i\}$ on the equilibrium process Ω.

Since we can also express each finite averaging in the energy eigenbasis using the partial dephasing maps $\mathcal{G}_{T_i}$, defined in Eq. (3.7), we can similarly write $\langle \Lambda \rangle_{\overline{\Upsilon}^{\mathscr{P}_{\mathbf{T}}}} = \mathrm{tr}[\mathcal{A}_k \varrho_k]$, where we now define

$$\varrho_i := \mathcal{G}_{T_i} \mathcal{A}_{i-1} \cdots \mathcal{A}_0 \mathcal{G}_{T_0}(\varrho), \quad \text{for} \quad i = 0, 1, \cdots, k, \tag{3.22}$$

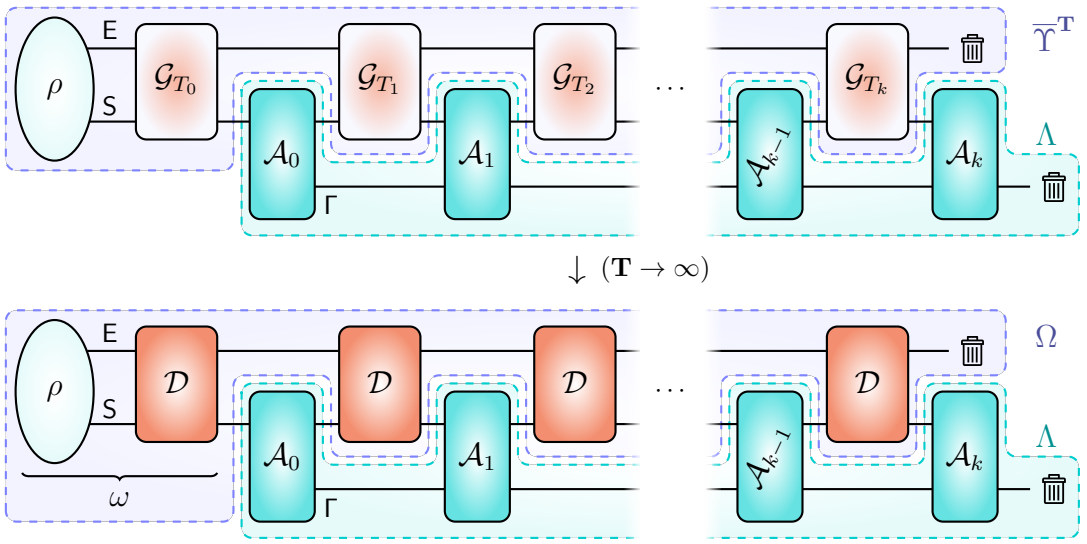


Figure 3.2: **Equilibration of quantum processes by finite temporal resolution:** This refers to a k-step process $\overline{\Upsilon}^{\mathbf{T}}$ time-averaged over each Hamiltonian evolution ($\mathcal{G}$) within time-windows of width $\mathbf{T} = (T_0, T_1, \ldots, T_k)$ remaining close to an equilibrium process Ω. We define an equilibrium process as one dephased ($\mathcal{D}$) with respect to the corresponding Hamiltonian at each time step. Equilibration is determined according to a set of operations $\{\mathcal{A}_i\}$ on a subsystem S, which can be correlated in time through an ancillary space Γ, and is represented by a single tensor Λ. By definition, equality is attained in the limit of all $\mathbf{T} \to \infty$.

as intermediate finite-time-averaged states after each intervention up to $\mathcal{A}_{i-1}$. As by definition $\lim_{T_i \to \infty} \mathcal{G}_{T_i} = \mathcal{D}$, the infinite-time limits $\mathbf{T} \to \infty$ make Υ indistinguishable from Ω. We also depict this in Fig. 3.2.

We may now generalize the left hand side of Eq. (3.11) with $|\langle\Lambda\rangle_{\overline{\Upsilon}^{\mathbf{T}} - \Omega}|$, asking how different the statistics of a set of operations $\{\mathcal{A}_i\}$ can be on a fuzzy clock process, $\overline{\Upsilon}^{\mathbf{T}}$, as opposed to those in the equilibrium one Ω. For clarity, let us present the case $H_i = H = \sum E_n P_n$, i.e. with a fixed Hamiltonian for all time-steps, and with $T_i = T_j = T$, i.e. with a clock with the same amount of fuzziness at all steps, as in Ref. [1]; we will then show how the general case reduces to this particular one.

Theorem 3.2.1 (Multitime equilibration due to finite temporal resolution [1]). Given an environment-system-ancilla ($\mathsf{SE}\Gamma$) composite with initial state $\varrho = \rho \otimes \gamma$ and initial equilibrium state $\varpi = \omega \otimes \gamma$, for any k-step process Υ with an evolution generated by a time-independent Hamiltonian on SE at each step, and for any fuzzy multitime observable Λ corresponding to a sequence of temporally local operations $\{\mathcal{A}_i\}_{i=0}^{k}$, each with fuzziness T acting on the joint $\mathsf{S}\Gamma$ system,

$$\left|\langle\Lambda\rangle_{\overline{\Upsilon}^{\mathscr{P}_{\mathbf{T}}} - \Omega}\right| \leq \mathbb{A}_k + \sum_{\ell=0}^{k-1} \|\mathcal{A}_{k:\ell+1}\| \, (\mathbb{B}_\ell + \mathbb{C}_\ell) \quad \text{with} \quad \mathbb{A}_k := \mathscr{S}_T^{k+1} \|\mathcal{A}_{k:0}\| \, \|\varrho - \varpi\|_2, \tag{3.23}$$

where here $\mathscr{S}_T := \max_{n \neq m} |G_{nm}^{(T)}|$ and $\mathcal{A}_{j:i} := \mathcal{A}_j \cdots \mathcal{A}_i$ is a composition of operations; the norm

$\|\cdot\|$ here stands for the norm on superoperators induced by the 2-norm, $\|\mathcal{X}\| = \sup_{\|\sigma\|_2=1} \|\mathcal{X}(\sigma)\|_2$; the first term is a single-time equilibration contribution, whereas the second term contains k multitime contributions where

$$\mathbb{B}_\ell := \|[\mathcal{G}_T^{k-\ell} - \mathcal{D}, \mathcal{A}_\ell]\varrho_\ell\|_2, \qquad \mathbb{C}_\ell := \|[\mathcal{D}, \mathcal{A}_\ell](\varrho_\ell - \varpi_\ell)\|_2, \tag{3.24}$$

with ϱ_i and ϖ_i intermediate finite-time averaged and equilibrium states at step i defined in Eq. (3.22) and Eq. (3.19), and where $[\cdot,\cdot]$ denotes a commutator of superoperators.

Proof. The main idea to bound the difference $\left|\langle\Lambda\rangle_{\overline{\Upsilon}^{\mathscr{P}_\mathbf{T}}-\Omega}\right|$ is to rearrange terms by commutation to obtain a single-time contribution with added correction terms. Let us slightly simplify notation by writing $\mathcal{G}_i = \mathcal{G}_{T_i}$ and only labelling $\mathcal{D}$ whenever it is relevant to know at which step this map is acting on. Let us first get this idea through with the $k = 1$ case,

$$\begin{aligned}
\left|\langle\Lambda\rangle_{\overline{\Upsilon}^{\mathscr{P}_\mathbf{T}}-\Omega}\right| &= \left|\mathrm{tr}\left[\left(\bigcirc_{j=0}^{1}\mathcal{A}_j\mathcal{G}_j - \bigcirc_{j=0}^{1}\mathcal{A}_j\mathcal{D}_j\right)(\varrho)\right]\right| \\
&= \left|\mathrm{tr}\left[\mathcal{A}_1\mathcal{A}_0\left(\mathcal{G}_1\mathcal{G}_0 - \mathcal{D}_1\mathcal{D}\right)(\varrho)\right] + \mathrm{tr}\left[\mathcal{A}_1\Big([\mathcal{G}_1,\mathcal{A}_0]\mathcal{G}_0 - [\mathcal{D}_1,\mathcal{A}_0]\mathcal{D}\Big)(\varrho)\right]\right| \\
&\leq \left|\mathrm{tr}\left[\mathcal{A}_{1:0}\left(\mathcal{G}_{1:0} - \mathcal{D}\right)(\varrho)\right]\right| + \left|\mathrm{tr}\left\{\mathcal{A}_1[\mathcal{G}_1 - \mathcal{D}, \mathcal{A}_0]\mathcal{G}_0(\varrho)\right\}\right| + \left|\mathrm{tr}\left\{\mathcal{A}_1[\mathcal{D}, \mathcal{A}_0](\mathcal{G}_0 - \mathcal{D})(\varrho)\right\}\right|,
\end{aligned} \tag{3.25}$$

where the third line follows by the triangle inequality ($|a - c| \leq |a - b| + |b - c|$, here with $b = \mathrm{tr}\{\mathcal{A}_1[\mathcal{D}, \mathcal{A}_0]\mathcal{G}_0(\varrho)\}$). We now adopt the notation $\mathcal{A}_{j:i} := \mathcal{A}_i \circ \cdots \circ \mathcal{A}_j$ and from now on ommit the $\circ$ symbol to denote composition of contiguous superoperators.

Then similarly we can do this for an arbitrary number of time-steps k,

$$\begin{aligned}
\left|\langle\Lambda\rangle_{\overline{\Upsilon}^{\mathscr{P}_\mathbf{T}}-\Omega}\right| &= \left|\mathrm{tr}\left[\left(\bigcirc_{j=0}^{k}\mathcal{A}_j\mathcal{G}_j - \bigcirc_{j=0}^{k}\mathcal{A}_j\mathcal{D}_j\right)(\varrho)\right]\right| \\
&= \left|\mathrm{tr}\left[\mathcal{A}_{k:0}\left(\mathcal{G}_{k:0} - \mathcal{D}\right)(\varrho)\right] + \sum_{\ell=0}^{k-1}\mathrm{tr}\left[\mathcal{A}_{k:\ell+1}\Big([\mathcal{G}_{k:\ell+1},\mathcal{A}_\ell]\mathcal{G}_\ell\bigcirc_{j=0}^{\ell-1}\mathcal{A}_j\mathcal{G}_j - [\mathcal{D},\mathcal{A}_\ell]\mathcal{D}_\ell\bigcirc_{j=0}^{\ell-1}\mathcal{A}_j\mathcal{D}_j\Big)(\varrho)\right]\right| \\
&\leq \left|\mathrm{tr}\left[\mathcal{A}_{k:0}\left(\mathcal{G}_{k:0} - \mathcal{D}\right)(\varrho)\right]\right| + \sum_{\ell=0}^{k-1}\left|\mathrm{tr}\left[\mathcal{A}_{k:\ell+1}[\mathcal{G}_{k:\ell+1} - \mathcal{D}, \mathcal{A}_\ell](\varrho_\ell)\right]\right| + \sum_{\ell=0}^{k-1}\left|\mathrm{tr}\left[\mathcal{A}_{k:\ell+1}[\mathcal{D}, \mathcal{A}_\ell](\varrho_\ell - \varpi_\ell)\right]\right|,
\end{aligned} \tag{3.26}$$

where $\varrho_\ell := \mathcal{G}_\ell \bigcirc_{j=0}^{\ell-1}\mathcal{A}_j\mathcal{G}_j(\varrho)$. Using $|\mathrm{tr}\,\mathcal{X}(\varrho)| \leq \|\mathcal{X}\|\|\varrho\|_2$, where for simplicity, here $\|\cdot\|$ stands for the induced 2-norm defined as $\|\mathcal{X}\| := \sup_{\|\sigma\|_2=1}\|\mathcal{X}(\sigma)\|_2$, which is a possible generalization of the operator norm for superoperators.[1] Then we further can bound

$$\begin{aligned}
\left|\langle\Lambda\rangle_{\overline{\Upsilon}^{\mathscr{P}_\mathbf{T}}-\Omega}\right| \leq{}& \|\mathcal{A}_{k:0}\|\ \|(\mathcal{G}_{k:0} - \mathcal{D})(\varrho)\|_2 + \sum_{\ell=0}^{k-1}\|\mathcal{A}_{k:\ell+1}\|\ \|[\mathcal{G}_{k:\ell+1} - \mathcal{D}, \mathcal{A}_\ell](\varrho_\ell)\|_2 \\
&+ \sum_{\ell=0}^{k-1}\|\mathcal{A}_{k:\ell+1}\|\ \|[\mathcal{D}, \mathcal{A}_\ell](\varrho_\ell - \varpi_\ell)\|_2,
\end{aligned} \tag{3.27}$$

[1] We can write the action of any bounded linear map $\mathcal{X}$ (not necessarily CP) as $\mathcal{X}(\cdot) = \sum_\mu L_\mu(\cdot)R_\mu^\dagger$ as a generalization of the Kraus representation in Section 0.1 [83]. The inequality can then be seen to follow with Hölder's inequality and the hierarchy of Schatten-norms. We use 2-norm for convenience in calculation.

which contains the terms $\mathbb{B}$ and $\mathbb{C}$ of Eq. (3.24) in the second and third terms of the inequality, and where the first term generalizes Eq. (3.10) as

$$\begin{aligned}
\|(\mathcal{G}_{k:0} - \mathcal{D})(\varrho)\|_2^2 &= \mathrm{tr}\left|\sum_{n\neq m} G^{(k)}_{n_k m_k}\cdots G^{(0)}_{n_0 m_0} P_n(\varrho)\, P_m\right|^2 \\
&= \sum_{\substack{n\neq m\\ n'\neq m'}} \prod_{j=0}^{k} G^{(j)}_{n_j m_j} G^{(j)}_{m'_j n'_j}\, \mathrm{tr}\left[P_n \varrho\, P_m P_{m'}\, \varrho\, P_{n'}\right] \\
&= \sum_{n\neq m} \prod_{j=0}^{k} |G^{(j)}_{n_j m_j}|^2\, \mathrm{tr}\left[P_n \varrho P_m \varrho\right] \\
&\leq \prod_{j=0}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2 \left\{\sum_{n,m} \mathrm{tr}[P_n \varrho P_m \varrho] - \sum_n \mathrm{tr}[P_n \varrho P_n \varrho]\right\} \\
&= \prod_{j=0}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2\, \mathrm{tr}(\varrho^2 - \varpi^2) \\
&= \|\varrho - \varpi\|_2^2 \prod_{j=0}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2. \qquad (3.28)
\end{aligned}$$

where similarly in the last line $\mathrm{tr}(\varrho^2 - \varpi^2) = \|\varrho - \varpi\|_2^2$, because $\mathrm{tr}(\varrho\,\varpi) = \mathrm{tr}(\varpi^2)$. ■

Remark 3.2.2. In general, by definition the term $\mathscr{S}_T$, which depends on the waiting time distribution $\mathscr{P}_T$, converges to zero in increasing T, with the rate of convergence depending on the specific distribution. In particular, as we exemplified for the single-time case, for the uniform distribution on all time-steps as standard equilibration in Section 1.2.3, we average over a time-window of width T around each τ_i for all time-steps, with $\mathscr{P}_T = T^{-1}$ in the interval $[\tau_i - T/2, \tau_i + T/2]$, and $\mathscr{P}_T = 0$ outside it. This yields

$$\left|G^{(T)}_{mn}\right| = |\sin(T\mathcal{E}_{mn})/T\mathcal{E}_{mn}|, \qquad (3.29)$$

where the term $\mathscr{S}_T$ then picks the smallest non-zero energy gap in the Hamiltonian.

Similarly, if the fuzziness corresponds to that of a half-normal distribution with variance T, then overall $\mathscr{S}_T$ decays exponentially with

$$\left|G^{(T)}_{mn}\right| \sim \exp(-T\mathcal{E}^2_{mn})\left|1 - \mathrm{erf}(i\sqrt{T}\mathcal{E}_{mn})\right|, \qquad (3.30)$$

where erf is the error function and $E_m - E_n = 2\mathcal{E}_{mn}$. For both cases, if T is small, $\mathscr{S}_T$ will also be vanishingly small whenever the energy gap $\mathcal{E}_{nm}$ that maximizes $|G^{(T)}_{nm}|$ is large enough, i.e. $\mathcal{E}_{nm} \gg T$. This property holds in general, since distributions $\mathscr{P}_T$ can be approximated as uniform for small T or because the gaps $\mathcal{E}_{nm}$ can be seen as a rescaling factor on T in the definition of $G^{(T)}_{nm}$.

The term $\mathbb{A}_k$ in Eq. (3.23) neglects temporal correlations and the operations $\{\mathcal{A}_i\}$ are all composed as a single operation $\mathcal{A}_{k:0} = \mathcal{A}_k \cdots \mathcal{A}_0$. This is essence can be interpreted as a single-time contribution to equilibration. The two-norm distance satisfies $\|\varrho - \varpi\|_2^2 \leq 1 - (d_E d_S)^{-1}$ as

the ancillary input γ can be taken to be pure. As discussed above, this term is suppressed through the $\mathscr{S}_T$ contributions when i) the averaging window, or equivalently the fuzziness of the clock T is large enough and ii) for small T whenever the energy gap maximizing the time-averaging $\left|G^{(T)}_{nm}\right|$ factor is large with respect to T.

Now, we can bound further the terms $\mathbb{B}_\ell$ and $\mathbb{C}_\ell$ in Theorem 3.2.1, which contain genuine multitime contributions relating to how well the intermediate states at step ℓ equilibrate. Continuing from Eq. (3.27), we have

$$\mathbb{B}_\ell = \|[\mathcal{G}_{k:\ell+1} - \mathcal{D}, \mathcal{A}_\ell](\varrho_\ell)\|_2 \leq \|\mathcal{A}_\ell\| \, \|(\mathcal{G}_{k:\ell+1} - \mathcal{D})(\varrho_\ell)\|_2 + \|(\mathcal{G}_{k:\ell+1} - \mathcal{D})\mathcal{A}_\ell(\varrho_\ell)\|_2 \,, \tag{3.31}$$

where we used the triangle inequality on the commutator. Now from this inequality we have, similarly, for the first term,

$$\begin{aligned}
\|(\mathcal{G}_{k:\ell+1} - \mathcal{D})\,(\varrho_\ell)\|_2^2 &\leq \prod_{j=\ell+1}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2 \left\{ \sum_{n,m} \mathrm{tr}[P_n \varrho_\ell P_m \varrho_\ell] - \sum_n \mathrm{tr}[P_n \varrho_\ell P_n \varrho_\ell] \right\} \\
&= \prod_{j=\ell+1}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2 \, \mathrm{tr}[\varrho_\ell^2 - \mathcal{D}(\varrho_\ell)\varrho_\ell] \\
&= \|\varrho_\ell - \mathcal{D}(\varrho_\ell)\|_2^2 \prod_{j=\ell+1}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2,
\end{aligned} \tag{3.32}$$

as $\mathrm{tr}[(\mathcal{D}(\varrho_\ell))^2] = \mathrm{tr}[\mathcal{D}(\varrho_\ell)\varrho_\ell]$, whilst for the second term, with $\varrho'_\ell = \mathcal{A}_\ell(\varrho_\ell)$,

$$\begin{aligned}
\|(\mathcal{G}_{k:\ell+1} - \mathcal{D})\,(\varrho'_\ell)\|_2^2 &\leq \prod_{j=\ell+1}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2 \left\{ \sum_{n,m} \mathrm{tr}[P_n(\varrho'_\ell) P_m(\varrho'_\ell)] - \sum_n \mathrm{tr}[P_n(\varrho'_\ell) P_n(\varrho'_\ell)] \right\} \\
&= \prod_{j=\ell+1}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2 \, \mathrm{tr}[\varrho'^2_\ell - \mathcal{D}(\varrho'_\ell)\varrho'_\ell] \\
&= \|\mathcal{A}_\ell(\varrho_\ell) - \mathcal{D}\mathcal{A}_\ell(\varrho_\ell)\|_2^2 \prod_{j=\ell+1}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}|^2,
\end{aligned} \tag{3.33}$$

so putting these together,

$$\mathbb{B}_\ell \leq \prod_{j=\ell+1}^{k} \max_{n\neq m} |G^{(j)}_{n_j m_j}| \left\{ \|\mathcal{A}_\ell\| \|\varrho_\ell - \mathcal{D}(\varrho_\ell)\|_2 + \|\mathcal{A}_\ell(\varrho_\ell) - \mathcal{D}\mathcal{A}_\ell(\varrho_\ell)\|_2 \right\}. \tag{3.34}$$

This means that in the particular case of same evolution $H_i = H_j = H$ and same fuzziness, $T_i = T_j = T$, we have

$$\mathbb{B}_\ell \lesssim \mathscr{S}_T^{k-\ell}, \tag{3.35}$$

so that, crucially, this term is suppressed overall in the width of the time-window T.

Finally, for $\mathbb{C}_\ell$, notice that we can further simplify the last term of Eq. (3.37) as

$$\begin{aligned}
\mathbb{C}_\ell &= \|[\mathcal{D}, \mathcal{A}_\ell](\varrho_\ell - \varpi_\ell)\|_2 \leq \|\mathcal{D}(\varrho_{\ell+1})\|_2 + \|\varpi_{\ell+1}\|_2 + \|\mathcal{A}_\ell\|(\|\mathcal{D}(\varrho_\ell)\|_2 + \|\varpi_\ell\|_2) \\
&\leq \|\mathcal{D}(\varrho_{\ell+1}) - \varpi_{\ell+1}\|_2 + \|\mathcal{A}_\ell\| \|\mathcal{D}(\varrho_\ell) - \varpi_\ell\|_2,
\end{aligned} \tag{3.36}$$

and each term is the purity of a dephased state, which will decay as the inverse effective dimension of that state. This follows as in general, $\mathrm{tr}\left[(\mathcal{D}(\sigma))^2\right] \leq d_{\mathrm{eff}}^{-1}(\sigma)$ for any state σ, with equality for either pure states or non-degenerate Hamiltonians. On the other hand, when the control operations from 0 to ℓ succeed in driving ϱ_ℓ so that the action of the commutator does not dephase it significantly, the purity of ϱ_ℓ may be large and thus $\mathbb{C}_\ell$ may become trivial (i.e. it approaches 1).

More concretely, the operations $\mathcal{A}_j$ interleaved within the intermediate states ϱ_ℓ and ϖ_ℓ will relate in the multitime correction terms in Eq. (3.24) to how greatly they disturb either the finite-time averaged ϱ_{j-1} or the equilibrated ϖ_{j-1}. This is most evident in the term $\mathbb{C}_\ell$, which can be bounded as well as $\mathbb{C}_\ell \leq \|[\mathcal{D}, \mathcal{A}_\ell]\| \|\varrho_\ell - \varpi_\ell\|_2$. The norm of the commutator can be written in terms of both the capacity of the operations $\mathcal{A}_\ell$ to generate coherences between different energy eigenspaces from equilibrium and the degree to which the operations can turn such coherences into populations. Environments in physical systems are typically much larger than the subsystems that can be probed, and, keeping in mind that the operations $\mathcal{A}_j$ act only on subsystem S and the ancilla $\mathsf{\Gamma}$, the ability to generate and detect energy coherences should be severely limited in many physically relevant cases.

A more general version of the bound in Theorem 3.2.1 can thus be given as follows. Let us denote $\mathfrak{S}_{b:a} := \prod_{j=a}^{b} \max_{n \neq m} |G^{(j)}_{n_j m_j}|$, then,

$$\begin{aligned}\left|\langle \Lambda \rangle_{\overline{\Upsilon}^{\mathscr{P}_{\mathbf{T}}} - \Omega}\right| \leq{}& \mathfrak{S}_{k:0} \, \|\mathcal{A}_{k:0}\| \, \|\varrho - \varpi\|_2 \\ &+ \sum_{\ell=0}^{k-1} \mathfrak{S}_{k:\ell+1} \, \|\mathcal{A}_{k:\ell+1}\| \left\{ \|\mathcal{A}_\ell\| \|\varrho_\ell - \mathcal{D}(\varrho_\ell)\|_2 + \|\mathcal{A}_\ell(\varrho_\ell) - \mathcal{D}\mathcal{A}_\ell(\varrho_\ell)\|_2 \right\} \\ &+ \sum_{\ell=0}^{k-1} \|\mathcal{A}_{k:\ell+1}\| \left\{ \|\mathcal{D}(\varrho_{\ell+1}) - \varpi_{\ell+1}\|_2 + \|\mathcal{A}_\ell\| \|\mathcal{D}(\varrho_\ell) - \varpi_\ell\|_2 \right\}, \end{aligned} \tag{3.37}$$

so if we now take the Hamiltonian at each time-step to be a fixed $H = \sum E_n P_n$ and we fix the fuziness to be the same at each step $T_i = T$, we now have $\mathfrak{S}_{b:a} = \mathscr{S}_T^{b-a}$ where $\mathscr{S}_T := \max_{n \neq m} \left|G^{(T)}_{nm}\right|$ as in Eq. (3.11).

To summarize, our result in Theorem 3.2.1 shows that either subsystems or global coarse properties of a closed time-independent Hamiltonian system will display equilibration for multiple sequential operations with a temporal uncertainty or fuzziness provided:

i. Both the initial and intermediate states have a significant overlap with the energy eigenstates.

ii. The temporal fuzziness is large enough relative to the average measurement time or, equivalently, the energy gaps in the Hamiltonian are large enough with respect to the temporal fuzziness.

iii. The disturbance by the operations on intermediate states is small.

3.3 Genuine multitime equilibration

In the previous section we have stressed the multitime nature of the bound in Theorem 3.2.1, however, how can we be sure that this is not simply an elaborated example of the results of Ref. [54] (described in Section 1.2.3)? Why would it not be possible to write the joint expectation $|\langle\Lambda\rangle_{\Upsilon}|$ as a single Heisenberg picture operator acting on the initial state, i.e. to simply group all of the time evolutions and measurements into a single Hermitian operator $\mathfrak{F}$ representing the measurement procedure acting on the initial state? This indeed suggests that a single-time equilibration bound, like the one in Eq. (1.27) would suffice to study equilibration in general quantum processes.

Let us then give a simple example that demonstrates that Theorem 3.2.1 indeed captures genuine multitime phenomena.

Example 3.3.1. For simplicity, consider only two CPTNI interventions acting on SΓ of the form

$$\mathcal{A}(\cdot) = \sum_{\mu} a_{\mu} A_{\mu}(\cdot) A_{\mu}^{\dagger}, \qquad \mathcal{B}(\cdot) = \sum_{\mu} b_{\mu} B_{\mu}(\cdot) B_{\mu}^{\dagger}, \tag{3.38}$$

then the joint expectation for these operations interleaved with evolutions over time intervals δt_0 and δt_1 is

$$\langle\Lambda\rangle_{\Upsilon} = \mathrm{tr}[\mathcal{B}\mathcal{U}_1\mathcal{A}\mathcal{U}_0(\varrho)] = \sum a_{\mu} b_{\nu}\, \mathrm{tr}[B_{\nu} U_1 A_{\mu} \mathcal{U}_0(\varrho) A_{\mu}^{\dagger} U_1^{\dagger} B_{\nu}^{\dagger}], \tag{3.39}$$

where as above, $\mathcal{U}_{\ell} = U_{\ell}(\cdot)U_{\ell}^{\dagger}$, with $U_{\ell} := \exp(-iH_{\ell}t_{\ell})$ with H_{ℓ} the Hamiltonian at time-step ℓ and $\varrho = \rho \otimes \gamma$ the full initial SEΓ state. Let us fix the basis for the Hamiltonians so that $H_{\ell} = \sum E_{n_{\ell}} P_n$ as above. We can now move terms around using cyclicity of trace to get

$$\langle\Lambda\rangle_{\Upsilon} = \sum a_{\mu} b_{\nu}\, \mathrm{tr}[A_{\mu}^{\dagger} U_1^{\dagger} B_{\nu}^{\dagger} B_{\nu} U_1 A_{\mu} \mathcal{U}_0(\varrho)] = \mathrm{tr}[\mathcal{A}\mathcal{U}_1^{\star}(\mathsf{B})\,\mathcal{U}_0(\varrho)], \tag{3.40}$$

where we define where $\mathcal{U}_{\ell}^{\star}(\cdot) := U_{\ell}^{\dagger}(\cdot)U_{\ell}$ with $\mathsf{B} := \sum b_{\mu} B_{\mu}^{\dagger} B_{\mu}$. The argument we refer to is that we can write this as $\langle\mathfrak{F}\rangle_{\varrho} = \mathrm{tr}[\mathfrak{F}\varrho(\delta t_0)]$, where $\varrho(\delta t_0) := \mathcal{U}_0(\varrho)$ and

$$\mathfrak{F} := \mathcal{A}\mathcal{U}_1^{\star}(\mathsf{B}), \tag{3.41}$$

and obtain that $|\langle\Lambda\rangle_{\overline{\Upsilon}^{\mathscr{P}_{\mathbf{T}}}-\Omega}| = |\langle\mathfrak{F}\rangle_{\overline{\varrho}^{\mathscr{P}_{T_0}}-\omega}|$.

This happens to be the case if the interval δt_1 is fixed, allowing fuzziness only in the first evolution time. In such case indeed we can simply apply the single-time result in Eq. (3.10), given that $\langle\Lambda\rangle_{\Omega} = \langle\mathfrak{F}\rangle_{\varpi_0}$ with $\varpi_0 = \mathcal{D}(\varrho)$ as defined in Eq. (3.18).

However, when the fuzziness of the clock is present for each intervention, i.e. in both evolution times, we have

$$\langle\Lambda\rangle_{\overline{\Upsilon}^{\mathscr{P}_{\mathbf{T}}}} = \mathrm{tr}[\mathcal{B}\,\mathcal{G}_1\mathcal{A}\,\mathcal{G}_0(\varrho)], \tag{3.42}$$

where $\mathcal{G}_{\ell} = \mathcal{G}_{T_{\ell}}$, as we use in the proof of Theorem 3.2.1 and defined in Eq. (3.7), is the time-evolution superoperator finite-time averaged with respect to the probability distributions $\mathscr{P}_{T_{\ell}}$

with a characteristic temporal fuzziness T_ℓ. As defined in Eq. (3.21), the joint expectation with respect to the equilibrium process is

$$\langle \Lambda \rangle_\Omega = \mathrm{tr}[\mathcal{B}\,\mathcal{D}\,\mathcal{A}\,\mathcal{D}(\varrho)]. \tag{3.43}$$

Now due to the double time average, it is impossible to write the difference of both quantities in terms of an operator expectation value, since

$$\begin{aligned} \left| \langle \Lambda \rangle_{\overline{\Upsilon}^{\mathscr{P}_T} - \Omega} \right| &= \mathrm{tr}[\mathcal{B}\,\mathcal{G}_1\,\mathcal{A}\,\mathcal{G}_0(\varrho) - \mathcal{B}\,\mathcal{D}\,\mathcal{A}\,\mathcal{D}(\varrho)] = \mathrm{tr}[\mathcal{B}\,\mathcal{G}_1\,\mathcal{A}\,\overline{\varrho}^{\mathscr{P}^{T_0}} - \mathcal{B}\,\mathcal{D}\mathcal{A}\,\varpi_0] \\ &\neq \left| \langle \mathfrak{E} \rangle_{\overline{\varrho}^{\mathscr{P}^{T_0}} - \varpi_0} \right|, \end{aligned} \tag{3.44}$$

for any operator $\mathfrak{E}$. Exceptions occur when we take the $T_1 \to \infty$ limit (so that $\mathcal{G}_1 \to \mathcal{D}$), or if there is no fuzziness for a fixed δt_1 (so that $\mathcal{G}_1$ and $\mathcal{D}$ are replaced by a fixed $\mathcal{U}_{\delta t_1}$) as argued above. This means that in general the result in Theorem 3.2.1 constitutes an equilibration result that cannot be reduced to a single-time one.

3.4 Conclusions

Let us finally note that our approach to describe the operations that can act on the process is general in the sense that these are CP maps which can be correlated between time-steps and propagate information from their interactions with the subsystem S through the ancillary space Γ. While these set a scale in all terms of the right-hand side of Eq. (3.11), they can also contribute to loosening it, potentially allowing to distinguish the fuzzy process from the equilibrium one within a finite time. It is not entirely clear, however, if a departure from equilibration is more readily accessible with a larger ancillary space Γ, and, for long time fuzziness T, the upper-bound in Eq. (3.11) should remain close to zero.

Similar to the single-time standard case described in Section 1.2.3, equilibration over multiple observations in open systems is expected intuitively through decoherence arguments [184]. The interplay with memory effects, through both the E and Γ in the interventions is as yet not entirely clear, e.g., under which circumstances finite temporal resolution equilibration can occur without the dynamics being Markovian, i.e., memoryless, or if the temporal correlations among interventions through the ancillary space can display a departure from equilibration within a finite-time.

In the following Chapters we will see more clearly how the questions on the foundations of statistical mechanics can be posed in direct analogy in the context of quantum processes with respect to Markovianity. Whilst here we approached the question of equilibration in quantum processes somewhat pragmatically, we can conjecture a bridge between the characterization of the equilibration process and Markovianity akin to that between the generic, time-average equilibrium state and the Gibbs state for thermal equilibrium. The relationship between the two properties is as yet, however, not so transparent but we can certainly expect some progress in this direction in the near future.

CHAPTER 4

Markovian Typicality

Il n'y a de nouveau que ce qui est oublié.[1]

– Rose Bertin

As we saw in Chapter 1, the quest towards understanding how thermodynamics emerges purely from quantum mechanical laws has seen a great deal of progress in recent years. Most prominently, equilibration on average deals with the dynamical explanation of how reversible and recurrent Schrödinger dynamics lead to irreversible reduced dynamics which converge and revolve around an equilibrium state, specifically telling us that time-dependent quantum properties evolve towards a certain fixed equilibrium value and stay close to it for most times. Moreover, we have seen that this dynamical convergence towards and around equilibrium holds more generally for subparts of closed systems under general quantum stochastic processes.

Whenever equilibration on average holds, it implies that the dynamics erases the information contained in the initial state of the respective system, however, there may still be non-Markovian memory of the initial state encoded in the temporal correlations between observables. Moreover, as we highlighted in Chapter 2, all open quantum evolutions generated by a time-independent SE Hamiltonian are non-Markovian, i.e. we know that, as far as nature is concerned, non-Markovianity is the rule and Markovianity is the exception which at best is an idealization.

How can we explain then this apparent contradiction? Furthermore, the Born-Markov approximation has proven to be extremely fertile over the years, being applicable to a wide class of physical models and situations [8, 100–103, 148, 149].

There is now an evident parallel flow of ideas that we can draw from the emergence of statistical mechanics discussed in Chapter 1; quantum systems dynamically equilibrate despite non-equilibrium being generic, in turn fulfilling the second law of thermodynamics, whose emergence can be explained from first principles such as entanglement, rather than from the equal a-priori probabilities postulate. Now we know that quantum processes satisfy an analogous form of dynamical equilibration and we can similarly ask if the emergence of forgetful processes

[1] *There is nothing new except what has been forgotten.* The quote is sometimes attributed to Marie Antoinette, of whom Bertin was the dressmaker.

can arise fundamentally rather than from ad-hoc assumptions or approximations such as the Born-Markov condition.

In Ref. [2] without resorting to the Born-Markov assumption or any other approximation, we formally proved that quantum processes are close to Markovian ones, when the subsystem S is sufficiently small compared to the whole SE, with a probability that tends to unity exponentially in the size of the latter. That is, we showed that Markovian processes are typical when these occur in small subsystems, with generic processes obeying a concentration of measure around Markovian ones. We also showed that, for a fixed global system size, it may not be possible to neglect non-Markovian effects when the process is allowed to continue for long enough, although detecting non-Markovianity for such processes would usually require non-trivial entangling resources. These results give birth to *almost* Markovian processes from closed dynamics analogous to the way in which entanglement supersedes the fundamental postulate of statistical mechanics.

4.1 Random quantum processes

The main approach we take to formally prove that Markovian processes are typical is generally speaking the one described in Section 1.3; here as well we want to explore the statistical properties of quantum processes and study how in this case Markovian processes turn out to be exceptional. As we will see, however, we will need some additional mathematical concepts when it comes to the moments of the unitary group.

Similar to the case of quantum states, to approach the question of what sampling a random quantum process means, we require a probability measure that assigns non-vanishing probabilities to mathematically generic unitary dynamics on the closed SE composite. We similarly we achieved this by sampling the evolution from the unitarily invariant Haar measure, introduced in Section 1.3.1. As we saw, this has the additional advantage of allowing employ random matrix theory techniques [185–191] and leads to the relatively straightforward application of concentration of measure results [70, 192, 193].

Consider then a k-step quantum process Υ on a SE composite of dimension $d_{\mathsf{SE}} = d_{\mathsf{S}} d_{\mathsf{E}}$, with initial state ρ and with unitary evolution given by unitary maps $\mathcal{U}_i = U_i(\cdot)\,U_i^\dagger$ acting on the full SE at the i^{th} timestep. As we now are dealing with unitary evolution at several time-steps, we use the Haar measure to sample two distinct types of unitary SE evolution.

Definition 4.1.1. We refer to these two ways of sampling as:

i. Random interaction: All U_i independently chosen.
ii. Constant interaction: $U_i = U_j, \quad \forall \leq i \neq j \leq k$.

We now depict this in Fig. 4.1(a). The entire set of unitaries enters into the process tensor as in the definition of the Choi state in Eq. (2.52). In the first case, the global system will quickly explore its entire (pure) state space for any initial state. The second case corresponds more closely to what one might expect for a truly closed system, where the Hamiltonian remains the

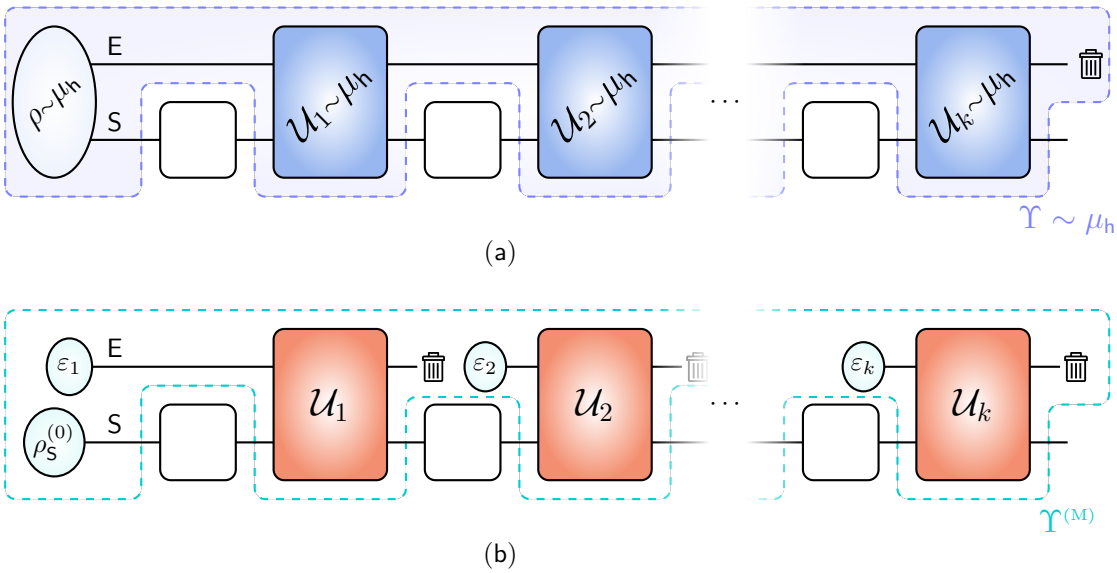


Figure 4.1: **Sampling a quantum process at random and quantum Markovian processes:** (a) By the sampling of a random quantum process we mean a k-step process with a unitary evolution sampled from the Haar measure μ_h either according to a random interaction (independent $\mathcal{U}_i \neq \mathcal{U}_j$) or a constant interaction (all $\mathcal{U}_i = \mathcal{U}_j$) as per Definition 4.1.1. (b) We consider how distinguishable is a generic process sampled at uniformly at random from a Markovian one, which is mathematically equivalent to one where SE is initially uncorrelated, $\rho = \rho_S^{(0)} \otimes \varepsilon_1$, with E being subsequently discarded and replaced afresh at every time-step.

same throughout the process. These correspond to two extremes; more generally, the dynamics from step to step may be related but not identical.

Two main features of this approach now stand out. The first is that we are directly sampling unitary dynamics from the Haar measure which are not generally given as a time-independent Hamiltonian dynamics; this is slightly different from sampling a pure quantum state at random, as done in Section 1.3, as in this case the type of dynamics generated by the Haar measure will be relevant. The second is precisely that the interaction or information flow between all parts of the whole SE composite will be relevant, i.e. no parts of the environment dimension are superfluous and in this sense we can think of having a strong interaction among all of the SE composite degrees of freedom.

4.2 The moments of the unitary group & average processes

As a k-step process tensor generically involves k unitary maps $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_k$. The fiducial state ρ can be taken to have undergone some evolution $\mathcal{U}_0$, which will let us interpret it as a random state. Back in Section 1.3.1, when we distributed SE states according to the Haar measure, we only required the first and second moments of the unitary group to obtain a concentration of measure result. In this case, however, we are generically considering quantities involving $k+1$

unitary maps, either all the same or all different and independently sampled.

We showed in Eq. (1.48) that the average quantum state drawn from the Haar measure is the maximally mixed state. For a random process, with all $\mathcal{U}_i \neq \mathcal{U}_j$ independently chosen, we can simply apply the average independently over each unitary. First let us rewrite the definition of the Choi state representation of the process tensor,

$$\begin{aligned}\Upsilon &= \mathrm{tr}_{\mathsf{E}}\left[\mathcal{U}_k\,\mathcal{S}_k\,\mathcal{U}_{k-1}\mathcal{S}_{k-1}\cdots\mathcal{U}_1\mathcal{S}_1\left(\rho\otimes\tilde{\Psi}^{\otimes k}\right)\right]\\ &= \sum_{\alpha,\ldots,\delta}\mathrm{tr}_{\mathsf{E}}\left[U_k\,\mathfrak{S}_{\alpha_k\beta_k}\cdots U_1\,\mathfrak{S}_{\alpha_1\beta_1}\,\rho\,\mathfrak{S}_{\delta_1\gamma_1}\,U_1^\dagger\cdots\mathfrak{S}_{\delta_k\gamma_k}\,U_k^\dagger\right]\otimes|\beta_1\alpha_1\cdots\beta_k\alpha_k\rangle\langle\delta_1\gamma_1\cdots\delta_k\gamma_k|, \quad (4.1)\end{aligned}$$

where the sum runs over all Greek indices from 1 to d_{S}, with $\tilde{\Psi}$ being an unnormalized maximally entangled state acting in the respective d_{S}-dimensional ancillary spaces $\mathsf{A}_i\mathsf{B}_i$, where $\mathcal{S}_i$ are swaps between S and ancillary system A_i at time-step i, and where $\mathfrak{S}_{\alpha\beta} = \mathbb{1}_{\mathsf{E}}\otimes|\alpha\rangle\langle\beta|$. Full detail can be revisited around the definition in Eq. (2.52). Then let us denote as

$$\mathbb{E}_{\mathfrak{h}}[f(U_0,U_1,\ldots,U_k)] = \int_{\mathbb{U}(d)} f(U_0,U_1,\ldots,U_k)\,d\mu_{\mathfrak{h}}(U_0)\,d\mu_{\mathfrak{h}}(U_1)\cdots d\mu_{\mathfrak{h}}(U_k), \quad (4.2)$$

the integration, or averaging, over the Haar measure independently over all time-steps, implicitly being over the different unitaries of the argument, so that we obtain

$$\begin{aligned}\mathbb{E}_{\mathfrak{h}}[\Upsilon] &= \sum_{\alpha,\ldots,\gamma}\mathrm{tr}_{\mathsf{E}}\left\{\mathbb{E}_{\mathfrak{h}}\left[U_k\,\mathfrak{S}_{\alpha_k\beta_k}\cdots U_1\,\mathfrak{S}_{\alpha_1\beta_1}\,U_0\rho\,U_0^\dagger\,\mathfrak{S}_{\delta_1\gamma_1}\,U_1^\dagger\cdots\mathfrak{S}_{\delta_k\gamma_k}\,U_k^\dagger\right]\right\}\\ &\qquad\otimes|\beta_1\alpha_1\cdots\beta_k\alpha_k\rangle\langle\delta_1\gamma_1\cdots\delta_k\gamma_k|\\ &= \sum_{\alpha,\ldots,\gamma,\epsilon}\frac{\mathbb{1}_{\mathsf{S}}}{d_{\mathsf{S}}}\langle\epsilon\beta_k|\mathbb{E}_{\mathfrak{h}}\left[U_{k-1}\,\mathfrak{S}_{\alpha_{k-1}\beta_{k-1}}\cdots U_1\,\mathfrak{S}_{\alpha_1\beta_1}\,U_0\rho\,U_0^\dagger\,\mathfrak{S}_{\delta_1\gamma_1}\,U_1^\dagger\cdots\mathfrak{S}_{\delta_{k-1}\gamma_{k-1}}\,U_{k-1}^\dagger\right]|\epsilon\delta_k\rangle\\ &\qquad\otimes|\beta_1\alpha_1\cdots\beta_k\alpha_k\rangle\langle\delta_1\gamma_1\cdots\delta_k\alpha_k|\\ &= \sum_{\alpha,\ldots,\gamma,\epsilon}\frac{\mathbb{1}_{\mathsf{S}}}{d_{\mathsf{S}}^2}\langle\epsilon\beta_{k-1}|\mathbb{E}_{\mathfrak{h}}\left[U_{k-2}\,\mathfrak{S}_{\alpha_{k-2}\beta_{k-2}}\cdots U_1\,\mathfrak{S}_{\alpha_1\beta_1}\,U_0\rho\,U_0^\dagger\,\mathfrak{S}_{\delta_1\gamma_1}\,U_1^\dagger\cdots\mathfrak{S}_{\delta_{k-2}\gamma_{k-2}}\,U_{k-2}^\dagger\right]|\epsilon\delta_{k-1}\rangle\\ &\qquad\otimes|\beta_1\alpha_1\cdots\beta_{k-1}\alpha_{k-1}\rangle\langle\delta_1\gamma_1\cdots\delta_{k-1}\alpha_{k-1}|\otimes 1_{\mathsf{A}_k\mathsf{B}_k}\\ &\vdots\\ &= \frac{\mathbb{1}_{\mathsf{S}\mathsf{A}_1\mathsf{B}_1\cdots\mathsf{A}_k\mathsf{B}_k}}{d_{\mathsf{S}}^{k+1}}, \quad (4.3)\end{aligned}$$

where we introduced $\{|\epsilon\rangle\}$ as a basis for E to perform each trace. This is, up to normalization of the maximally entangled states, a maximally mixed state in the full system-ancillary space, implying that the average process tensor is *maximally noisy*.

Example 4.2.1. As a simple example consider $k = 2$. Then the action of the average process tensor, say $\mathcal{T}^{\mathfrak{h}}_{2:0}$, on a pair of CPTNI operations $\mathcal{A}_0$ and $\mathcal{A}_1$ is given by

$$\begin{aligned}\mathcal{T}^{\mathfrak{h}}_{2:0}[\{\mathcal{A}_i\}] &= \mathrm{tr}_{\mathsf{in}}\left\{\mathbb{E}_{\mathfrak{h}}[\Upsilon_{2:0}]\left[(\mathbb{1}_{\mathsf{A}_1}\otimes\mathcal{A}_0^{\mathrm{T}})\tilde{\Psi}\otimes(\mathbb{1}_{\mathsf{A}_2}\otimes\mathcal{A}_1^{\mathrm{T}})\tilde{\Psi}\right]\right\} = \frac{\mathrm{tr}[\mathcal{A}_0(\mathbb{1})]\,\mathrm{tr}[\mathcal{A}_1(\mathbb{1})]}{d_{\mathsf{S}}^3}\,\mathbb{1}_{\mathsf{S}}\\ &\leq \frac{\mathbb{1}_{\mathsf{S}}}{d_{\mathsf{S}}}, \quad (4.4)\end{aligned}$$

with equality for TP maps, i.e. with $\mathrm{tr}[\mathcal{A}_i(\mathbb{1})] = \mathrm{tr}(\mathbb{1}) = d_{\mathsf{S}}$.

This generalizes similarly for any k, implying that the Haar average process tensor is maximally noisy, or analogous to quantum channel terms, completely depolarizing. Notice as well that this average process is Markovian, as the Choi state has a tensor product structure.

This fact relies upon the independent sampling of each unitary; however, if we consider a constant interaction, with all $\mathcal{U}_i = \mathcal{U}_j$, we require a single integral equivalent to the $(k+1)$-moment of the unitary group, as defined in Eq. (1.44). In Section 1.3.1 we were able to compute the first and second moments by means of the twirl map and the Schur-Weyl duality. However, for higher-order moments this is not practical anymore, as per by the Schur-Weyl duality we need to consider all permutations and then relate back the twirl map with the moments of the unitary group. We can, however, capture in an abstract form the behaviour of the moments of the unitary group by means of an object called the *Weingarten function* (which in a sense also relies on the Schur-Weyl duality); this will let us, in particular, understand the asymptotic behaviour in dimensions and time-steps for random processes, as will become clear below.

The Weingarten function can be defined in different ways; it is a fairly complicated function to evaluate explicitly [187, 194] but in any case, for the n-moments of the unitary group, it only depends on a given permutation of n and it gives a rational number in the dimension of the unitary in question. Tables with particular cases are often cited in the literature, which are helpful for computing lower moments, see e.g. Ref. [69, 187]. An alternative is to perform calculations numerically [189, 195]). For our purposes, we keep the definition of the Weingaten function as follows.

Definition 4.2.2 (Weingarten function [187])**.** Let $\mathfrak{G}_n$ be the symmetric group on $\{1, 2, \ldots, n\}$, and let $n \leq d$ and $\sigma \in \mathfrak{G}_n$, then the (unitary) Weingarten function,[2] Wg, is defined by

$$\mathsf{Wg}(\sigma, d) = \mathbb{E}_{\mathsf{h}} \left[\prod_{i=1}^{n} U_{ii} U^*_{i\sigma(i)} \right], \tag{4.5}$$

where $U \sim \mu_{\mathsf{h}}$ is a Haar distributed $d \times d$ unitary matrix.

More explicit expressions of Wg, such as the one derived in Ref. [186], are outside the scope of this thesis, however, it is important to point out that Wg depends not only on the particular permutation σ but on its cycle structure, that is, for example, for $n = 3$, the cyclic permutations $\sigma = (1, 2, 3)$ and $\sigma' = (1, 3, 2)$, meaning $\sigma : 1 \to 2 \to 3 \to 1$ and $\sigma' : 1 \to 3 \to 2 \to 1$, will evaluate to the same value on Wg, i.e. $\mathsf{Wg}(\sigma, d) = \mathsf{Wg}(\sigma', d)$. We denote this cycle type with brackets as follows: given a set of positive integers $\{m_i\}$ such that $m_1 + m_2 + \ldots + m_\ell = n$, then $[m_1, m_2, \ldots, m_\ell]$ means cyclic permutations replacing m_1, m_2 and m_ℓ elements. For the previous example then the cycle type of σ and σ' is $[3]$, similarly for the permutations $(1, 2)(3)$, $(1, 3)(2)$ and $(2, 3)(1)$ it will be $[2, 1]$, whilst for $(1)(2)(3)$ we denote it as $\mathbf{1}^3 := [1, 1, 1]$.

[2] Wg is named after Donald Weingarten [185], who first studied asymptotic properties of the n-moments of $\mathbb{U}(d)$. An explicit expression in terms of characters of symmetric groups and Schur functions was first derived by Benoît Collins [186] (later expanded in Ref. [188]), who coined the term.

The similarity with the expression for the n-moments of $\mathbb{U}(d)$ in Eq. (1.44) is now quite evident, and it implies the following.

Theorem 4.2.3 (The n-moments of the unitary group [186, 187]). Let $U \in \mathbb{U}(d)$ with $U \sim \mu_{\mathsf{h}}$ be a a $d \times d$ Haar random distributed unitary and $n \leq d$, then

$$\mathbb{E}_{\mathsf{h}}\left[\prod_{\ell=1}^{n} U_{i_\ell j_\ell} U^*_{i'_\ell j'_\ell}\right] = \sum_{\sigma,\tau\in\mathfrak{G}_n} \prod_{\ell=1}^{n} \delta_{i_\ell i'_{\sigma(\ell)}} \delta_{j_\ell j'_{\tau(\ell)}} \operatorname{Wg}(\tau\sigma^{-1}, d), \tag{4.6}$$

where U_{ij} is the ij^{th} entry of U.

A property that follows from this result, together with the invariance of the Haar measure, is that $\mathbb{E}_{\mathsf{h}}\left[U_{i_1 j_1}\cdots U_{i_n j_n} U^*_{i'_1 j'_1}\cdots U^*_{i'_{n'} j'_{n'}}\right] = 0$, i.e. there has to be the same amount of U^*_{ij} as there are U_{ij} [187]. As we have already computed the first and second moments in Eq. (1.48) and in Eq. (1.50) (together with Eq. (1.53)), albeit indirectly through the twirl map, we can nevertheless compare with the values in the table of Ref. [187] for the corresponding Weingarten functions,

$$\mathsf{Wg}([1], d) = \frac{1}{d}, \qquad \mathsf{Wg}([\mathbf{1}^2], d) = \frac{1}{d^2-1}, \qquad \mathsf{Wg}([2], d) = -\frac{1}{d(d^2-1)}. \tag{4.7}$$

For the first moment, the Weingarten function corresponds to the only coefficient of the 1-twirl, whilst for the second moment the Weingarten functions appear as coefficients in the α, β functions of the 2-twirl in Eq. (1.53), that is, in fact we can write the 1-twirl $\Xi^{(2)}$ and the 2-twirl $\Xi^{(2)}$ as

$$\Xi^{(1)}[(\cdot)] = \mathsf{Wg}([1], d)\operatorname{tr}[(\cdot)], \tag{4.8}$$

$$\Xi^{(2)}[(\cdot)] = \mathsf{Wg}([\mathbf{1}^2], d) Z + \mathsf{Wg}([2], d)\,\textsc{swap}\, Z, \quad \text{where} \quad Z = \operatorname{tr}[(\cdot)]\mathbb{1} + \operatorname{tr}[\textsc{swap}(\cdot)]\,\textsc{swap}, \tag{4.9}$$

and so similarly, the n-twirl can be written in terms of the corresponding n Weingarten functions.

As mentioned before, this is generally not an easy thing to do and is usually done only for moments of small n; however, the asymptotic behaviour is usually one of interest in random matrix theory (or more generally in any theory of non-commutative random variables) and in our case it will help us establish our results for Haar distributed quantum processes. Ultimately, the asymptotic behavior of the moments of the unitary group boils down to that of the Wg function.

Remark 4.2.4. In Ref. [187] it is shown that

$$\mathsf{Wg}(\sigma \in \mathfrak{G}_n, d) \sim \frac{1}{d^{2n-\#\sigma}}, \text{ as } d \to \infty, \tag{4.10}$$

as a refinement of a result in Ref. [188], where $\#\sigma$ is the number of cycles of the permutation σ counting also fixed points (assignments from an element to itself, $\sigma(x) = x$).

Let us then consider the average process in the constant interaction case, $\mathcal{U}_i = \mathcal{U}_j = \mathcal{U}$, $\forall i, j \leq k$. As it is clear that for us $d = d_{\mathsf{SE}}$, we will omit the dimension dependence in Wg. As we originally derived[3] in Ref. [2] and we reproduce in Appendix B.1, the average k-step process

[3] In Ref. [2] we normalized each maximally entangled state entering the Choi state of the process tensor; the reason for this will be made clear in the following sections.

tensor in the constant interaction case can be written by means of a set $\{|s_i^{(\prime)}\rangle\}_{s_i=1}^{d_\mathsf{S}}$ of S system bases for $i = 0, 1, \ldots, k$ as

$$\mathbb{E}_\mathsf{h}[\Upsilon] = \sum_{\sigma,\tau\in\mathfrak{G}_{k+1}} \rho_{\tau(0);0}\mathsf{Wg}(\tau\sigma^{-1})\Delta_{k,\sigma,\tau}^{(d_\mathsf{E})}|s_{\sigma(k)}\rangle\langle s_k|\bigotimes_{j=1}^{k}|s_{\sigma(j-1)}s'_{\tau(j)}\rangle\langle s_{j-1}s'_j|, \tag{4.11}$$

with implicit sum over all repeated basis $(s_i^{(\prime)})$ indices, where here $\mathfrak{G}_{k+1}$ is the symmetric group on $\{0, 1, \ldots, k\}$, and with the definitions

$$\rho_{\tau(0);0} = \langle e'_{\tau(0)}s'_{\tau(0)}|\rho|e'_0s'_0\rangle, \tag{4.12}$$

$$\Delta_{k,\sigma,\tau}^{(d_\mathsf{E})} = \delta_{e_{\sigma(k)}e_k}\prod_{\ell=1}^{k}\delta_{e_{\sigma(\ell-1)}e'_{\tau(\ell)}}\delta_{e_{\ell-1}e'_\ell}, \tag{4.13}$$

where $\{|e_i\rangle\}_{e_i=1}^{d_\mathsf{E}}$, $\{|e'_i\rangle\}_{e'_i=1}^{d_\mathsf{E}}$, with $i = 0, 1, \ldots, k$, also implicitly summed over all elements e_i and e'_i, is a set of E bases and the Δ term is simply a monomial in d_E, with degree determined by σ and τ.

The case $k = 0$ recovers $\mathbb{E}_\mathsf{h}[\Upsilon_{0:0}] = \mathbb{1}_S/d_\mathsf{S}$ as expected, as no process occurs; the $1/d_\mathsf{S}$ factor arises from the $\mathsf{Wg}([1], d_\mathsf{SE})$ function, with the unifying d_E factor coming from $\Delta_{0,\sigma,\tau}^{d_\mathsf{E}}$. In Appendix B.1.1 we also write the case $k = 1$ for a superchannel and its purity, which can be seen to be close to the maximally mixed one and coincide with it in the large E limit, i.e. it coincides with the random interaction case.

As we saw in the whole Chapter 1, the small subsystem limit, $d_\mathsf{E} \gg d_\mathsf{S}$, is of particular interest. For the average process tensor, we saw that the random interaction case in Eq. (4.3) is independent of d_E, however for the constant interaction case we will get terms in inverse powers of d_E arising from the Wg functions. In particular, when looking at the limit $d_\mathsf{E} \to \infty$ of Eq. (4.11), the only term that does not vanish is the one with $\sigma, \tau = \mathbf{1}^{k+1}$, i.e. with both permutations being identities, as these generate the most numerator powers in d_E in the $\Delta_{k,\sigma,\tau}^{(d_\mathsf{E})}$ term in Eq.(4.13) when summed over all e_i's. In other words, we get the contribution from $\Delta_{k,\sigma,\tau}^{(d_\mathsf{E})}$,

$$\sum_{\substack{e_0,e_1,\ldots,e_k=1\\ e'_1,e'_2,\ldots,e'_k=1}}^{d_\mathsf{E}}\left[\delta_{e_ke_k}\prod_{\ell=1}^{k}\delta_{e_{\ell-1}e'_\ell}\delta_{e_{\ell-1}e'_\ell}\right] = d_\mathsf{E}^{k+1}, \tag{4.14}$$

and all other terms will vanish because of the d_E powers in the denominator generated by the Wg functions will dominate those from $\Delta_{k,\sigma,\tau}^{(d_\mathsf{E})}$. Given the asymptotic limit of Wg in Eq. (4.10) we can see that indeed the least powers in d produced by it are those when $\sigma\tau^{-1} = \mathbf{1}^{k+1}$ because $\#\mathbf{1}^n = n$, i.e. identity produces the greatest number of cycles, being the number of all possible fixed points. Finally, as $\sum_{\epsilon'_0\varsigma'_0}\langle\epsilon'_0\varsigma'_0|\rho|\epsilon'_0\varsigma'_0\rangle = \operatorname{tr}\rho = 1$, we get

$$\begin{aligned}\mathbb{E}_\mathsf{h}[\Upsilon] &\sim d_\mathsf{E}^{k+1}\mathrm{Wg}(\mathbf{1}^{k+1})\,\mathbb{1}_{\mathsf{SA}_1\mathsf{B}_1\ldots\mathsf{A}_k\mathsf{B}_k}\\ &\sim \frac{\mathbb{1}_{\mathsf{SA}_1\mathsf{B}_1\ldots\mathsf{A}_k\mathsf{B}_k}}{d_\mathsf{S}^{k+1}}, \quad \text{when} \quad d_\mathsf{E}\to\infty,\end{aligned} \tag{4.15}$$

coinciding with the average over a random interaction process.

4.3 A bound on non-Markovianity & average noisiness

As we saw previously in Section 2.4.5, the process tensor leads to a well-defined Markov criterion from which it is possible to construct a family of operationally meaningful measures of non-Markovianity, many of which can be stated simply as *distances* between a process tensor's Choi state Υ and the closest Markovian one $\Upsilon^{(\mathrm{M})}$. We saw as well that $\Upsilon^{(\mathrm{M})}$ must take the form of a tensor product of quantum maps $\mathcal{Z}_{i:i-1}$ connecting adjacent pairs of time steps, $\Upsilon^{(\mathrm{M})} = \bigotimes_{i=1}^{k} \mathcal{Z}_{i:i-1} \otimes \rho_{\mathsf{S}}^{(0)}$.

We are now interested in studying the non-Markovianity of a generic quantum process. In Ref. [2], in analogy with the seminal studies on equilibration that we introduced in Chapter 1, we chose the measure of non-Markovianity defined in terms of the trace distance D as

$$\mathcal{N}_1 := \min_{\Upsilon^{(\mathrm{M})}} D\left(\Upsilon, \Upsilon^{(\mathrm{M})}\right) \propto \min_{\Upsilon^{(\mathrm{M})}} \|\Upsilon - \Upsilon^{(\mathrm{M})}\|_1 \tag{4.16}$$

where $\|X\|_1 := \mathrm{tr}\sqrt{XX^\dagger}$ is the trace norm (or Schatten 1-norm as defined by Eq. (19)). In particular, this trace-distance measure is related to relative entropy, briefly discussed in Section 2.4.5, through the so-called quantum Pinsker inequality, $\mathcal{R}(\Upsilon\|\Upsilon^{(\mathrm{M})}) \geq 2\, D^2(\Upsilon, \Upsilon^{(\mathrm{M})})$. Now, given that $\mathcal{N}_1$ is a trace-distance measure, we demand that it satisfies $0 \leq \mathcal{N}_1 \leq 1$. This can be done imposing a normalization factor or directly normalizing the Choi states. We chose the latter,[4] with the caveat that we employ it consistently only for the purposes of employing this measure of non-Markovianity.

Notation 4.3.1. From this point onward we normalize the Choi state of the process tensor to unity, which is equivalent to defining these via properly normalized maximally entangled states, i.e. from here on we redefine Υ as

$$\Upsilon = \mathrm{tr}_{\mathsf{E}}[\mathcal{U}_k\mathcal{S}_k\mathcal{U}_{k-1}\mathcal{S}_{k-1}\cdots\mathcal{U}_1\mathcal{S}_1\mathcal{U}_0(\rho\otimes\Psi^{\otimes k})], \quad \text{where} \quad \Psi = \frac{1}{d_{\mathsf{S}}}\tilde{\Psi}, \tag{4.17}$$

where Ψ are now rightful maximally entangled states, so that now Υ satisfies $\mathrm{tr}[\Upsilon] = 1$. This implies that now we explicitly have

$$\mathcal{N}_1 := \frac{1}{2}\min_{\Upsilon^{(\mathrm{M})}} \|\Upsilon - \Upsilon^{(\mathrm{M})}\|_1. \tag{4.18}$$

Now, we should point out that the choice of this measure is historical as well, as the trace distance is rather a distinguishability measure for quantum states rather than for quantum maps, and a distinguishability measure on Choi states would be directly a measure on these states rather than an operationally meaningful one on the corresponding processes. As we will see throughout the next chapter, this is not an issue in itself as we can tightly bound the measure $\mathcal{N}_1$ with an operationally relevant measure of non-Markovianity known as the diamond distance, and the result we obtained in Ref. [2] is only changed by a multiplicative constant in a minor

[4] The reasons might be said to be somewhat historical; one reason that motivated this choice is that it is more intuitive to think of properly normalized maximally mixed states as maximally noisy processes, as opposed to having an identity with an incorrect normalization factor.

way, ultimately not changing its consequences. In fact, we will see that we can very often relate families of non-Markovianity measures, such as those given by Schatten norms, by at most a multiplicative factor. However, the results in this chapter and the numerical calculations were originally obtained with $\mathcal{N}_1$. We will thus, for consistency and for simplicity's sake, present the results as originally derived with this measure in this thesis.[5]

To begin with, given the difficulty in computing and minimizing the Markovian Choi state, we may upper bound the distance $\mathcal{N}_1$ by a trace distance with respect to the maximally mixed state, which as we have seen would correspond to the noisiest Markovian process possible,

$$\mathcal{N}_1 \leq D\left(\Upsilon, \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\right), \tag{4.19}$$

where the identity acting on S together with the $k+1$ ancillas is implied. We may further bound this by considering the following cases separately.

1. Case $d_{\mathsf{E}} < d_{\mathsf{S}}^{2k+1}$: We notice that $\operatorname{rank}(\Upsilon) \leq d_{\mathsf{E}}$.[6] Letting γ be the diagonal matrix of up to d_{E} non-vanishing eigenvalues λ_{γ_i} of the Choi state, we may write

$$\begin{aligned}\|\Upsilon - \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\|_1 &= \sum_{i=1}^{d_{\mathsf{E}}} \left|\lambda_{\gamma_i} - \frac{1}{d_{\mathsf{S}}^{2k+1}}\right| + \sum_{j=d_{\mathsf{E}}+1}^{d_{\mathsf{S}}^{2k+1}} \left|-\frac{1}{d_{\mathsf{S}}^{2k+1}}\right| \\ &= \|\gamma - \frac{\mathbb{1}_{\mathsf{E}}}{d_{\mathsf{S}}^{2k+1}}\|_1 + 1 - \frac{d_{\mathsf{E}}}{d_{\mathsf{S}}^{2k+1}},\end{aligned} \tag{4.20}$$

where $|\cdot|$ denotes the standard absolute value, so using the inequality $\|X\|_1 \leq \sqrt{\dim(X)}\|X\|_2$ for a square matrix X, where $\|X\|_2 = \sqrt{\operatorname{tr}(XX^\dagger)}$ is the Schatten 2-norm,

$$\begin{aligned}\|\Upsilon - \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\|_1 &\leq \sqrt{d_{\mathsf{E}}}\|\gamma - \frac{\mathbb{1}_{\mathsf{E}}}{d_{\mathsf{S}}^{2k+1}}\|_2 + 1 - \frac{d_{\mathsf{E}}}{d_{\mathsf{S}}^{2k+1}} \\ &= \sqrt{d_{\mathsf{E}} \operatorname{tr}[\Upsilon^2] + \frac{d_{\mathsf{E}}^2}{d_{\mathsf{S}}^{4k+2}} - \frac{2d_{\mathsf{E}}}{d_{\mathsf{S}}^{2k+1}}} + 1 - \frac{d_{\mathsf{E}}}{d_{\mathsf{S}}^{2k+1}},\end{aligned} \tag{4.21}$$

Furthermore, applying Jensen's inequality for the square-root, $\mathbb{E}[\sqrt{X}] \leq \sqrt{\mathbb{E}[X]}$, for any $\mathbb{E}_{\mathsf{h}}$ or $\mathbb{E}_{\mathfrak{h}}$, we have

$$\mathbb{E}\left[\mathcal{N}_1\right] \leq \frac{1}{2}\left(\sqrt{d_{\mathsf{E}}\mathbb{E}[\operatorname{tr}(\Upsilon^2)] + \frac{d_{\mathsf{E}}^2}{d_{\mathsf{S}}^{4k+2}} - \frac{2d_{\mathsf{E}}}{d_{\mathsf{S}}^{2k+1}}} + 1 - \frac{d_{\mathsf{E}}}{d_{\mathsf{S}}^{2k+1}}\right). \tag{4.22}$$

2. Case $d_{\mathsf{E}} \geq d_{\mathsf{S}}^{2k+1}$: This case is a small subsystem limit for most k. Directly applying $\|X\|_1 \leq \sqrt{\dim(X)}\|X\|_2$ as before,

$$\|\Upsilon - \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\|_1 \leq \sqrt{d_{\mathsf{S}}^{2k+1}}\|\Upsilon - \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\|_2 = \sqrt{d_{\mathsf{S}}^{2k+1}\operatorname{tr}[\Upsilon^2] - 1}, \tag{4.23}$$

[5] See Eq. (5.14) for the main result in this Chapter in terms of diamond norm.

[6] One may see this by looking at the pure state $|\Phi\rangle\langle\Phi| \equiv \mathcal{U}(\Theta \otimes \Psi^{\otimes k})\mathcal{U}^\dagger$ with Schmidt decomposition $|\Phi\rangle = \sum_{i=1}^{n}\sqrt{\lambda_i}|e_i s_i\rangle$ where $n = \min\left(d_{\mathsf{E}}, d_{\mathsf{S}}^{2k+1}\right)$.

Similarly, taking the average over evolution, by means of Jensen's inequality,

$$\mathbb{E}\left[\mathcal{N}_1\right] \leq \frac{1}{2}\sqrt{d_{\mathsf{S}}^{2k+1}\mathbb{E}\left[\mathrm{tr}(\Upsilon^2)\right] - 1}. \tag{4.24}$$

This means that we can construct a piecewise function

$$\mathcal{B}_k(d_{\mathsf{E}}, d_{\mathsf{S}}) \equiv \begin{cases} \dfrac{\sqrt{d_{\mathsf{E}}\,\mathbb{E}[\mathrm{tr}(\Upsilon^2)] - x} + y}{2} & \text{if} \quad d_{\mathsf{E}} < d_{\mathsf{S}}^{2k+1} \\ \dfrac{\sqrt{d_{\mathsf{S}}^{2k+1}\mathbb{E}[\mathrm{tr}(\Upsilon^2)] - 1}}{2} & \text{if} \quad d_{\mathsf{E}} \geq d_{\mathsf{S}}^{2k+1} \end{cases}, \tag{4.25}$$

with $x = d_{\mathsf{E}} d_{\mathsf{S}}^{-(2k+1)}\,(1+y)$ and $y = 1 - d_{\mathsf{E}} d_{\mathsf{S}}^{-(2k+1)}$, which provides an upper bound on the average non-Markovianity, i.e.

$$\mathbb{E}[\mathcal{N}_1] \leq \mathcal{B}_k. \tag{4.26}$$

So similar to Section 1.3, we now need to compute the average purity of Υ.

The purity is a quantifier of the mixedness, or uniformity of eigenvalues of a positive operator, and as we saw as well in Section 1.3, when computed on reduced states of bipartite systems it can serve as a quantifier of entanglement. In Appendix B.2 and in Appendix B.3, we reproduce the computation of the expected purity of the Choi state of a process tensor $\mathbb{E}[\Upsilon^2]$ in the random and the constant interaction pictures, respectively, which we derived originally in Ref. [2]. This average purity can be directly translated as a quantifier for *noisiness* of the quantum process itself and can serve as well to measure the entanglement between system and environment. Bear in mind again that here the process tensors are normalized to unity. The average purities take the form

$$\mathbb{E}_{\mathfrak{h}}[\mathrm{tr}\left(\Upsilon^2\right)] = \frac{d_{\mathsf{E}}^2 - 1}{d_{\mathsf{E}}(d_{\mathsf{SE}} + 1)}\left(\frac{d_{\mathsf{E}}^2 - 1}{d_{\mathsf{SE}}^2 - 1}\right)^k + \frac{1}{d_{\mathsf{E}}}, \tag{4.27}$$

for the random interaction picture, where we have assumed that the fiducial state ρ is pure, and

$$\mathbb{E}_{\mathsf{h}}[\mathrm{tr}\left(\Upsilon^2\right)] = d_{\mathsf{S}}^{-2k} \sum_{\sigma,\tau\in\mathfrak{S}_{2k+2}} \mathsf{Wg}(\tau\sigma^{-1})\,\rho_{\tau(0);k+1}\rho_{\tau(k+1);0}\,\Delta_{k,\sigma,\tau}^{(d_{\mathsf{E}},d_{\mathsf{S}})}, \tag{4.28}$$

in the constant interaction case, where $\rho_{\tau(\cdot);\ell}$ uses the same notation as in Eq. (4.12) and is shown explicitly in Eq. (B.29) on Appendix B.3, and Δ is a product, scaling with k, of monomials in d_{E} and d_{S} depending on permutations σ and τ, shown in full in Eq. (B.30) also within Appendix B.3.

4.3.1 Limiting cases

In both the constant and random interaction cases, the bound $\mathcal{B}_k$ on the non-Markovianity $\mathcal{N}_1$ is a well-behaved rational function of d_{E}, d_{S} and k. In the constant interaction case, Eq. (4.28) takes a non-trivial form mainly because of the Wg function (which is intrinsic to the Haar-unitary averaging in the constant interaction picture). However, due to the results in Ref. [187, 188], we can still study analytically the behaviour of the bound $\mathcal{B}_k$ for both cases in the following limits.

Small subsystem limit

For both interaction pictures,

$$\lim_{d_{\mathsf{E}}\to\infty} \mathbb{E}[\mathrm{tr}(\Upsilon^2)] = \frac{1}{d_{\mathsf{S}}^{2k+1}}, \tag{4.29}$$

which corresponds to the purity of the maximally mixed state. Note that the averaging occurs after computing the purity independently of which case is considered, i.e. this does not correspond to the purity of the average process but the average purity of a process.

This implies that on average, in the small subsystem limit, a process will be indistinguishable from the maximally noisy (and hence Markovian) one,

$$\lim_{d_{\mathsf{E}}\to\infty} \mathbb{E}[\mathcal{N}_1] = \lim_{d_{\mathsf{E}}\to\infty} \mathcal{B}_k(d_{\mathsf{E}}, d_{\mathsf{S}}) = 0, \tag{4.30}$$

and from Eq. (4.27) we know it does so at a rate $\mathcal{O}(1/d_{\mathsf{E}})$ in the random interaction case.

Long time limit

The other interesting limiting case is the one where the SE dimension is fixed, but the number of time steps is taken to be very large. The resulting process encodes all high order correlation functions between observables over a long period of time. Again for both cases, the expected purity in this limit goes as

$$\lim_{k\to\infty} \mathbb{E}[\mathrm{tr}(\Upsilon^2)] = \frac{1}{d_{\mathsf{E}}}, \tag{4.31}$$

which corresponds to the maximally mixed purity of the environment. This implies that the Choi state of the full SE unitary process is maximally entangled between S and E. In this limit, we get correspondingly

$$\lim_{k\to\infty} \mathbb{E}[\mathcal{N}_1] \leq \lim_{k\to\infty} \mathcal{B}_k(d_{\mathsf{E}}, d_{\mathsf{S}}) = 1, \tag{4.32}$$

meaning only that we cannot say much about non-Markovianity in this limit, i.e. a typical process in this limit could be highly non-Markovian. Indeed, we expect this to be the case, since the finite-dimensional SE space will have a finite recurrence time.

Average state purity limit

As expected as well, our result generalizes the well-known result for quantum states, i.e., when $k = 0$. In this case we recover the average purity

$$\mathbb{E}[\mathrm{tr}(\rho_S^2)] = \frac{d_{\mathsf{E}} + d_{\mathsf{S}}}{d_{\mathsf{SE}} + 1}, \tag{4.33}$$

where $\rho_S \equiv \Upsilon_{0:0} = \mathrm{tr}_E(U\rho U^\dagger)$, as we re-derived in Eq. (1.57).

4.4 Concentration around Markovian processes

From the discussion in Section 1.3.1 it seems clear that we have almost all the ingredients to turn the statements above into a concentration of measure result. We proved the following.

Theorem 4.4.1 (Concentration of measure around Markovian processes [2]). Let $\Upsilon \sim \mu_{\mathsf{h}}$ be a k-step quantum process undergone by a d_{S}-dimensional subsystem of a larger d_{SE}-dimensional composite, sampled at random according to the Haar measure, then for any $\delta > 0$,

$$\mathbb{P}_{\mathsf{h}}[\mathcal{N}_1 \geq \mathcal{B}_k(d_{\mathsf{E}}, d_{\mathsf{S}}) + \delta] \leq \mathrm{e}^{-\mathscr{C}(d_{\mathsf{E}}, d_{\mathsf{S}})\delta^2}, \tag{4.34}$$

where $\mathscr{C}(d_{\mathsf{E}}, d_{\mathsf{S}}) = c\, d_{\mathsf{SE}} \left(\frac{d_{\mathsf{S}}-1}{d_{\mathsf{S}}^{k+1}-1}\right)^2$ with $c = 1/4$ for a constant interaction process and $c = (k+1)/4$ for a random interaction process. The function $\mathcal{B}_k$ is an upper bound on the expected non-Markovianity $\mathbb{E}[\mathcal{N}_1]$, given in Eq. (4.25), whose details depend on the way in which processes are sampled.

Proof. We have previously obtained the function $\mathcal{B}_k$ completely, so we now derive the Lipschitz constants $\mathscr{L}$ for both the constant and the random interaction cases, as well as the concentration rate.

Lipschitz constant (constant interaction)

Let $\tau : \mathbb{U}(d_{\mathsf{SE}}) \to \mathbb{R}$ defined by $\tau(U) = D\left(\Upsilon(U), \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\right) = \frac{1}{2}\|\Upsilon(U) - \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\|_1$, where we explicitly mean $\Upsilon(U) = \mathrm{tr}_E[\mathsf{U}_{k:0}\Theta\mathsf{U}_{k:0}^\dagger]$ where $\mathsf{U}_{k:0} = U_k\mathcal{S}_k\cdots U_1\mathcal{S}_1U_0$ and with $\Theta = \rho \otimes \Psi^{\otimes k}$ a pure state, then for all $V \in \mathbb{U}(d_{\mathsf{SE}})$,

$$\begin{aligned}
|\tau(U) - \tau(V)| &= \left| D\left(\Upsilon(U), \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\right) - D\left(\Upsilon(V), \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\right)\right| \\
&\leq D(\Upsilon(U), \Upsilon(V)) \\
&\leq D(\mathsf{U}_{k:0}\Theta\mathsf{U}_{k:0}^\dagger, \mathsf{V}_{k:0}\Theta\mathsf{V}_{k:0}^\dagger),
\end{aligned} \tag{4.35}$$

where here similarly $\mathsf{V}_{k:0} = V_k\mathcal{S}_k\cdots V_1\mathcal{S}_1V_0$.

Consider now different labelings for the unitary at each time step so that we can easily track each one. By the triangle inequality,

$$\begin{aligned}
\|\mathsf{U}_{k:0}\Theta\mathsf{U}_{k:0}^\dagger - \mathsf{V}_{k:0}\Theta\mathsf{V}_{k:0}^\dagger\|_1 &\leq \|\mathsf{U}_{k:0}\Theta\mathsf{U}_{k:0}^\dagger - U_k\mathcal{S}_k\mathsf{V}_{k-1:0}\Theta\mathsf{V}_{k-1:0}^\dagger\mathcal{S}_k^\dagger U_k^\dagger\|_1 \\
&\quad + \|U_k\mathcal{S}_k\mathsf{V}_{k-1:0}\Theta\mathsf{V}_{k-1:0}^\dagger\mathcal{S}_k^\dagger U_k^\dagger - \mathsf{V}_{k:0}\Theta\mathsf{V}_{k:0}^\dagger\|_1 \\
&= \|\mathsf{U}_{k-1:0}\Theta\mathsf{U}_{k-1:0}^\dagger - \mathsf{V}_{k-1:0}\Theta\mathsf{V}_{k-1:0}^\dagger\|_1 \\
&\quad + \|U_k\mathcal{S}_k\mathsf{V}_{k-1:0}\Theta\mathsf{V}_{k-1:0}^\dagger\mathcal{S}_k^\dagger U_k^\dagger - \mathsf{V}_{k:0}\Theta\mathsf{V}_{k:0}^\dagger\|_1,
\end{aligned} \tag{4.36}$$

where it is clear by context that U stands for $U \otimes \mathbb{1}$.

Now we use Lemma 1 of Ref. [196], which states that

$$\|A\sigma A^\dagger - B\sigma B^\dagger\|_1 \leq 2\|A-B\|_2 \tag{4.37}$$

for two unitaries A, B and any σ. For the simplest case with $k=1$, this gives

$$\begin{aligned}
&\|U_1\mathcal{S}_1V_0\Theta V_0^\dagger\mathcal{S}_1^\dagger U_1^\dagger - V_1\mathcal{S}_1V_0\Theta V_0^\dagger\mathcal{S}_1^\dagger V_1^\dagger\|_1 \\
&\qquad = \|\frac{\sum}{d_\mathsf{S}}(U_1\,\mathfrak{S}_{\alpha\beta}\,V_0\rho V_0^\dagger\,\mathfrak{S}_{\gamma\delta}\,U_1^\dagger - V_1\,\mathfrak{S}_{\alpha\beta}\,V_0\rho V_0^\dagger\,\mathfrak{S}_{\gamma\delta}\,V_1^\dagger)\otimes|\beta\alpha\rangle\langle\gamma\delta|\|_1 \\
&\leq \frac{\sum_{\alpha\beta}}{d_\mathsf{S}}\,\mathrm{tr}\Bigg[\sum_{\gamma\delta}(U_1\,\mathfrak{S}_{\alpha\beta}\,V_0\rho V_0^\dagger\,\mathfrak{S}_{\gamma\delta}\,U_1^\dagger - V_1\,\mathfrak{S}_{\alpha\beta}\,V_0\rho V_0^\dagger\,\mathfrak{S}_{\gamma\delta}\,V_1^\dagger) \\
&\qquad\qquad (U_1\,\mathfrak{S}_{\beta\alpha}\,V_0\rho V_0^\dagger\,\mathfrak{S}_{\delta\gamma}\,U_1^\dagger - V_1\,\mathfrak{S}_{\beta\alpha}\,V_0\rho V_0^\dagger\,\mathfrak{S}_{\delta\gamma}\,V_1^\dagger)\Bigg]^{1/2} \\
&\leq \frac{\sum_{\alpha\beta}}{d_\mathsf{S}}\|U_1\,\mathfrak{S}_{\alpha\beta}\,V_0\rho V_0^\dagger\left(\sum\mathfrak{S}_{\gamma\delta}\right)U_1^\dagger - V_1\,\mathfrak{S}_{\alpha\beta}\,V_0\rho V_0^\dagger\left(\sum\mathfrak{S}_{\gamma'\delta'}\right)V_1^\dagger\|_1 \\
&\leq \frac{2}{d_\mathsf{S}}\sum_{\alpha,\beta=1}^{d_\mathsf{S}}\|U_1-V_1\|_2 = 2d_\mathsf{S}\|U_1-V_1\|_2,
\end{aligned} \tag{4.38}$$

and doing similarly, for the i-th step,

$$\|U_i\mathcal{S}_i\mathsf{V}_{i-1:0}\Theta\mathsf{V}_{i-1:0}^\dagger\mathcal{S}_i^\dagger U_i^\dagger - \mathsf{V}_{i:0}\Theta\mathsf{V}_{i:0}^\dagger\|_1 \leq 2d_\mathsf{S}^i\|U_i-V_i\|. \tag{4.39}$$

Thus bounding iteratively expression (4.36), it follows that

$$\|\mathsf{U}_{k:0}\Theta\mathsf{U}_{k:0}^\dagger - \mathsf{V}_{k:0}\Theta\mathsf{V}_{k:0}^\dagger\|_1 \leq 2\sum_{\ell=0}^{k} d_\mathsf{S}^\ell\|U_\ell - V_\ell\|_2, \tag{4.40}$$

finally giving

$$|\tau(U)-\tau(V)| \leq \left(\frac{d_\mathsf{S}^{k+1}-1}{d_\mathsf{S}-1}\right)\|U-V\|_2. \tag{4.41}$$

Lipschitz constant (random interaction)

On the other hand, for the ergodic case, let $\mathbb{U}^{\times(k+1)}(d) = \underbrace{\mathbb{U}(d)\times\cdots\times\mathbb{U}(d)}_{k+1\text{ times}}$ be the $k+1$ Cartesian product space of d-dimensional unitary groups, then we define $\zeta : \mathbb{U}^{\times(k+1)}(d_\mathsf{SE}) \to \mathbb{R}$ by $\zeta(\vec{U}) = D\left(\Upsilon(\vec{U}), \frac{\mathbb{1}}{d_\mathsf{S}^{2k+1}}\right)$ where now $\vec{U} = (U_0,\cdots,U_k)$. Similarly as before, we now have

$$\left|\zeta(\vec{U})-\zeta(\vec{V})\right| \leq \sum_{\ell=0}^{k} d_\mathsf{S}^\ell\|U_\ell - V_\ell\|_2, \tag{4.42}$$

and we may let the metric on $\mathbb{U}^{\times(k+1)}(d_\mathsf{SE})$ be the 2-product metric $\delta_\mathbb{U}$ defined [197] by

$$\delta_\mathbb{U}(\vec{x},\vec{y}) = \sqrt{\sum_{\ell=0}^{k}\|x_\ell - y_\ell\|_2^2}, \tag{4.43}$$

which then satisfies

$$\sum_{\ell=0}^{k} d_{\mathsf{S}}^{\ell}\|U_\ell - V_\ell\|_2 \leq \sum_{\ell=0}^{k} d_{\mathsf{S}}^{\ell}\sqrt{\sum_{\ell'=0}^{k}\|U_{\ell'} - V_{\ell'}\|_2^2} = \left(\frac{d_{\mathsf{S}}^{k+1}-1}{d_{\mathsf{S}}-1}\right)\delta_{\mathbb{U}}(\vec{U},\vec{V}), \tag{4.44}$$

and thus we conclude that

$$\left|\zeta(\vec{U}) - \zeta(\vec{V})\right| \leq \left(\frac{d_{\mathsf{S}}^{k+1}-1}{d_{\mathsf{S}}-1}\right)\delta_{\mathbb{U}}(\vec{U},\vec{V}), \tag{4.45}$$

so the (bound on) Lipschitz constants coincide with

$$\mathscr{L} \leq \frac{d_{\mathsf{S}}^{k+1}-1}{d_{\mathsf{S}}-1}, \tag{4.46}$$

which essentially behaves as $\mathcal{O}\left(d_{\mathsf{S}}^{k}\right)$.

The concentration function (constant interaction)

We now make use of a result related to the Gromov-Bishop inequality (see e.g. Theorem 7 in Ref. [198]) stating that if $\mathrm{Ric}(M) \geq \mathrm{Ric}(\Sigma^n(R)) = \frac{n-1}{R^2}$ for an n-dimensional manifold M, where $\Sigma^n(R)$ is the n-dimensional sphere of radius R and $\mathrm{Ric}(X)$ is the infimum of diagonal elements of the Ricci curvature tensor on X, then the respective concentration functions satisfy $\alpha_M(x) \leq \alpha_{\Sigma^n(R)}(x) \leq \exp\left[-\frac{x^2(n-1)}{2R^2}\right]$ [70].

For the constant interaction case, the corresponding manifold is the group manifold $\mathscr{U}$ of $\mathbb{U}(d)$ (where here $d = d_{\mathsf{SE}}$), which is diffeomorphic to $\mathbb{SU}(d) \times \Sigma^1(1)$ [70], where $\mathbb{SU}(d)$ denotes the special unitary group (i.e. with added $\det(U) = 1$ for any element) and thus has $\mathrm{Ric}(\mathscr{U}) = d/2$ and $\dim(\mathscr{U}) = d^2$. Then it follows that $\mathrm{Ric}(\mathscr{U}) \geq \mathrm{Ric}(\Sigma^{d^2}(r))$ if $r^2 \geq 2(d^2-1)/d$ and hence, taking the minimal case,

$$\alpha_{\mathscr{U}}(\delta/\mathscr{L}) \leq \alpha_{\Sigma^{d^2}(r)}(\delta/\mathscr{L}) \leq \exp\left(-\frac{\delta^2 d_{\mathsf{SE}}}{4\mathscr{L}^2}\right). \tag{4.47}$$

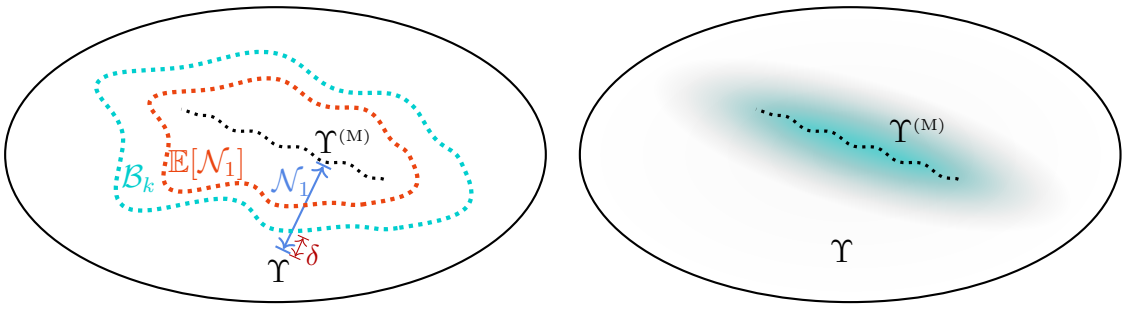
The concentration function (random interaction)

For the random interaction case the corresponding manifold is the group manifold $\mathbf{U}$ of the $k+1$ Cartesian product space $\mathbb{U}(d)\times\cdots\times\mathbb{U}(d)$. In this case $\mathrm{Ric}(\mathbf{U}) = (k+1)d/2$ and $\dim(\mathbf{U}) = (k+1)d^2$. Then it follows that $\mathrm{Ric}(\mathbf{U}) \geq \mathrm{Ric}(\Sigma^{(k+1)d^2}(\mathcal{R}))$ if $\mathcal{R}^2 \geq 2[(k+1)d^2-1]/[d(k+1)]$ and hence, taking the minimal case,

$$\alpha_{\mathbf{U}}(\delta/\mathscr{L}) \leq \alpha_{\Sigma^{(k+1)d_{\mathsf{SE}}^2}(\mathcal{R})}(\delta/\mathscr{L}) \leq \exp\left(-\frac{\delta^2(k+1)d_{\mathsf{SE}}}{4\mathscr{L}^2}\right). \tag{4.48}$$

The factor $d_{\mathsf{SE}}\mathscr{L}^{-2}$ behaves as $\mathcal{O}\left(d_{\mathsf{E}}d_{\mathsf{S}}^{-2k+1}\right)$, so both concentration functions will be small whenever $d_{\mathsf{E}} \gg d_{\mathsf{S}}^{2k}$. ■

Theorem 4.4.1 assures that the probability for the non-Markovianity $\mathcal{N}_1$ to exceed a function of k, d_{S} and d_{E}, that becomes very small in the large d_{E} limit, itself becomes small in that limit.



(a) (b)

Figure 4.2: **Concentration around Markovian processes in large dimensional environments:** (a) A geometric cartoon of our main result in a space of process tensors: the probability of the non-Markovianity $\mathcal{N}_1$ of deviating from $\mathcal{B}_k$ by some $\delta > 0$ decreases exponentially in δ^2. In (b), quantum processes Υ on large dimensional environments (such that $d_E \gg d_S^{2k+1}$) concentrate around the Markovian ones $\Upsilon^{(\mathrm{M})}$.

Our result is meaningful when both $\mathcal{B}_k + \delta$ and $\exp\left(-\mathscr{C}\delta^2\right)$ are small; the latter is fulfilled in the *small subsystem* or *large environment* limit, which in our setting means $d_{\mathsf{E}} \gg d_{\mathsf{S}}^{2k+1}$. We may also state the minimal value for δ that, assuming the large environment limit, renders both sides small,[7] i.e., such that $\delta^2 d_{\mathsf{E}} \gg 1 \gg \delta$; this is fulfilled for $\delta = d_{\mathsf{E}}^{-1/3}$. A geometrical cartoon to illustrate the result is presented in Fig. 4.2.

Theorem 4.4.1 can be said to state that *almost all quantum processes are almost Markovian.*

Our results imply that it is fundamentally hard to observe non-Markovianity in a typical process and thus go some way to explaining the overwhelming success of Markovian theories.

Specifically, Theorem 4.4.1 shows that even when interacting strongly with the wider composite system, a subsystem will typically undergo highly Markovian dynamics when the rest of the system has a sufficiently large dimension, and that the probability to be significantly non-Markovian vanishes with the latter. Our main result formalizes the notion that in the large environment limit a quantum process, taken uniformly at random, will be almost Markovian with very high probability. This corroborates the common understanding of the Born-Markov approximation, discusses in Section 1.3, but, crucially, we make no assumptions about weak coupling between E and S. Instead, in the Haar random interactions we consider, every part of the system typically interacts significantly with every part of E. This is in contrast to many open systems models, even those with superficially infinite dimensional baths, where the effective dimension of the environment is relatively small [199]; it can always be bounded by a function of time scales in the system-environment Hamiltonian [200], which could be encoded in a bath spectral density. Our result is also more general than the scenario usually considered, since it accounts for interventions and thus the flow of information between S and E across multiple times.

[7] Similar to Ref. [65] detailed in Section 1.3, here we look for an $x > 0$ such that $\delta = d_{\mathsf{E}}^{-x}$ and $\delta^2 d_{\mathsf{E}} = d_{\mathsf{E}}^{x}$.

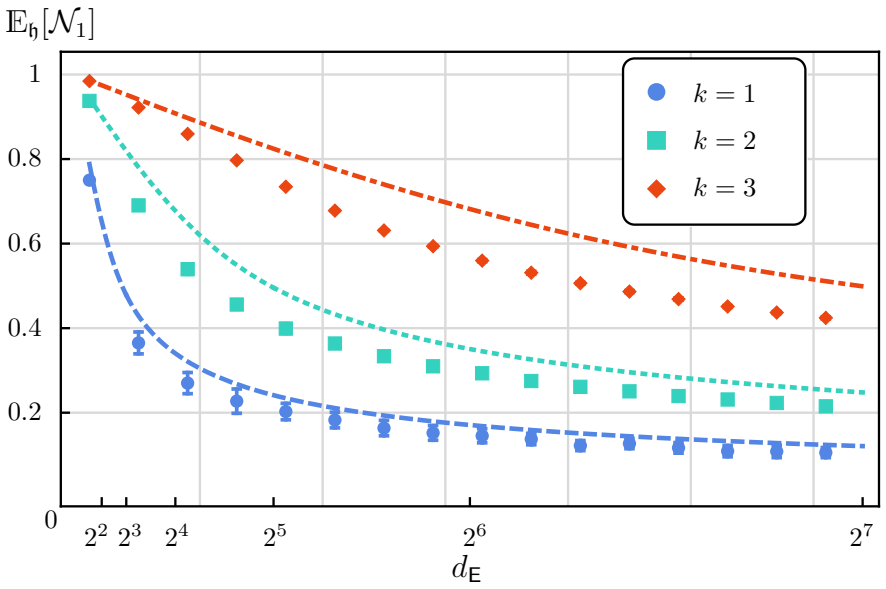


Figure 4.3: **Average non-Markovianity $\mathbb{E}_\mathfrak{h}[\mathcal{N}_1]$ of a random interaction process for a qubit in the environment dimension d_E at fixed time steps k:** Discrete values are shown for numerical averages over $\lfloor 40/k \rfloor$ randomly generated process tensors Υ at time steps $k = 1, 2, 3$ and with fixed $d_\mathsf{S} = 2$; error bars denote the standard deviation due to sampling error. The lines above each set of points denote the upper bound $\mathcal{B}_k(d_E, 2)$. Process tensors were generated by sampling Haar random unitaries according to Ref. [66] and described in Appendix C

4.4.1 Numerical sampling

To support our results, we sampled process tensors Υ numerically in the random interaction case and computed their corresponding average non-Markovianity $\mathbb{E}[\mathcal{N}_1]$ as a function of environment dimension d_E for a fixed system dimension $d_\mathsf{S} = 2$, obtaining the behaviour shown in Fig. 4.3. The details on how this was done can be seen in Appendix C.

For constant interaction the numerical results are practically indistinguishable from those in the random case, but as mentioned, the analytical bound $\mathcal{B}_k$ is much harder to compute exactly. This suggests that either a simpler bound exists or that it might be possible to simplify the one we have obtained.[8] As expected, our numerical results fall within the bound $\mathcal{B}_k(d_\mathsf{E}, d_\mathsf{S} = 2)$ and they behave similarly; we notice that the bound in general seems to be somewhat loose, and become loosest when $d_\mathsf{E} \simeq d_\mathsf{S}^{2k+1}$, implying that non-Markovianity might be hard to detect even when not strictly in the large environment limit. However, it does saturate rapidly as d_E increases.

So far, our results are valid for process tensors constructed with Haar random unitaries at k evenly spaced steps; we are effectively considering a strong interaction between system and environment which rapidly scrambles quantum information in both [69]. As we saw in Section 1.2,

[8] Notice that despite expressions in Eq. (4.11) and Eq. (4.28) being seemingly complicated, this complexity arises only because of the permutations and the Weingarten function.

the mechanism for equilibration is precisely that of dephasing, or effectively, the scrambling of information on the initial state of the system. This suggests that, even when timescales will differ with the type of evolution considered, most physical evolutions fall within our result, with e.g. a weaker behavior in k.

This is perhaps the most contentious part of our results from Ref. [2] presented in this chapter: as opposed to the typicality of quantum states where the Haar measure only plays a role in the sampling, in the case of quantum processes it has an implication on the class of dynamics that are being considered, and as we are well aware nature is far from random. We will discuss this point in more precise terms in the next chapter, where we directly approached this issue. Despite these features, we will now show that our results still hold at a coarse-grained level, where the intermediate dynamics corresponds to products of Haar random unitaries, which are not themselves Haar random.

4.5 Observing non-Markovianity

The choice of unitaries going into the process tensor in the previous sections (in our case drawn from the Haar measure), dictates a time scale for the system, up to a freely chosen energy scale. However, as we saw in Section 2.4.3, illustrated in Fig. 2.10, the process tensor satisfies a containment property, so that we can straightforwardly construct process tensors on a longer, coarse-grained time scale by simply allowing the system to evolve, or equivalently be acted with an identity operation, between some subset of time steps. And in fact, as we also mentioned before, process tensors at all time scales should be related in this way to an underlying process tensor with an infinite number of steps [154]. We thus now refer to this construction through the containment property as *coarse graining*.

To see that our main result directly applies to any coarse-grained process tensor, we again consider the definition of our non-Markovianity measure in Eq. (4.16). Consider the coarser grained process tensor

$$\Upsilon_{\text{coarse}} = \Upsilon_{k:0\setminus\{i\in[0,k-1]\}}, \tag{4.49}$$

where a subset of operations $\{\mathcal{A}_i\}_{i\in[0,k-1]}$ are replaced by identity operations, i.e. the system is simply left to evolve. Letting $\mathcal{N}_{\text{coarse}} \equiv \min_{\Upsilon^{(\text{M})}_{\text{coarse}}} \mathcal{D}(\Upsilon_{\text{coarse}}, \Upsilon^{(\text{M})}_{\text{coarse}})$, we have

$$\mathcal{N}_{\text{coarse}} \leq \mathcal{N}_1, \tag{4.50}$$

since the set of allowed $\Upsilon^{(\text{M})}_{\text{coarse}}$ strictly contains the allowed $\Upsilon^{(\text{M})}$ at the finer-grained level. This renders the new process less distinguishable from a Markovian one, i.e., coarse-graining can only make processes more Markovian.

The physical intuition behind this result is that the amount of information which can be encoded in a coarse grained process tensor is strictly less than that in its parent process tensor. In fact, this is a key feature of non-Markovian memory: the memory should decrease under coarse graining. On the other hand, due to the same reasoning, we cannot say anything about

finer-grained dynamics once given a process up to a finite number of time-steps. One approach to tackling this issue would be to choose a different sampling procedure which explicitly takes scales into account and can deal with this problem directly incorporating this notion of graining.

Moreover, there is another important limitation for observing non-Markovianity. The operational interpretation of the trace distance, discussed in the previous section, implies that observing non-Markovianity requires applying a measurement that is an eigenprojector operator of $\Upsilon - \Upsilon^{(\mathrm{M})}$. The optimal measurement will, in general, be entangled across all time steps. In practice, this is hard to achieve and typically one considers a sequence of local measurements $m \in \mathbb{M}$. In general, for an any set of measurements $\mathbb{M}$ we can define a restricted measure of non-Markovianity detectable with that set: $\mathcal{D}_{\mathbb{M}}(\Upsilon, \Upsilon^{(\mathrm{M})}) \equiv \max_{m\in\mathbb{M}} \frac{1}{2}|\operatorname{tr}[m(\Upsilon - \Upsilon^{(\mathrm{M})})]| \leq \mathcal{D}(\Upsilon, \Upsilon^{(\mathrm{M})})$, which means that the detectable non-Markovianity will be smaller. This is akin to the *eigenstate thermalization hypothesis* [201], where all eigenstates of a physical Hamiltonian look uniformly distributed with respect to most *physically reasonable* observables. In our setting, this means that looking for non-Markovianity with observables that are local in time –i.e., physically reasonable– we find almost no temporal correlations.

The locality constraint, along with monotonicity of non-Markovianity under coarse graining, have further important consequences for a broad class of open systems studies where master equations are employed [202]. Since master equations usually only account for two-point correlations with local measurements, they will be insensitive to most of the temporal correlations being accounted for by our measure, leading to an even greater likelihood for their descriptions to be Markovian. We will also discuss this point in the next chapter when going beyond the Haar measure.

4.6 Conclusions

The generic form of open quantum dynamics is non-Markovian, but, despite this, it is often very well approximated by simpler Markovian dynamics. How this *memorylessness* emerges is not dissimilar to questions, regarding the emergence of thermodynamic behaviour, which have pervaded quantum mechanics since its conception. We now know that the fundamental postulate of equal a-priori probabilities of statistical mechanics can be traced back to the entanglement between subsystems and their environment [65]. With Theorem 4.4.1, obtained originally in Ref. [2], we have shown that very similarly, if we sample a generic quantum process occurring in a large finite environment at random, it will be almost Markovian with very high probability.

Specifically, we have showed that, even when interacting strongly with the wider composite system, a subsystem will typically undergo highly Markovian dynamics when the rest of the system has a sufficiently large dimension, and that the probability to be significantly non-Markovian vanishes with the latter. Theorem 4.4.1 formalizes the notion that in the large environment limit a quantum process, taken uniformly at random, will be almost Markovian with very high probability, in turn also corroborating the common understanding of the Born-Markov

approximation described in Section 2.3, where crucially, however, we make no assumptions about weak coupling between E and S. It is important that this is different to the case of many open systems models which consider superficially infinite dimensional baths, i.e. where the effective dimension of the environment is relatively small [199], as these can always be bounded by a function of time scales in the system-environment Hamiltonian [200] and which could be encoded in a bath spectral density. Most importantly, as have been stressed throughout Chapter 2, our result naturally contains the scenario usually considered, since it accounts for multiple interventions and thus the flow of information between S and E across multiple times.

It is also important to highlight that while it may still be possible to observe non-Markovian behaviour at a time scale that is smaller than the fundamental time scale set by the chosen unitaries, Eq. (4.50) tells us that any coarse grained process will remain concentrated around the Markovian ones in the large environment limit. Otherwise, for larger and larger systems, one needs an ever increasing number of time steps, corresponding to higher order correlations, in order to increase the probability of witnessing non-Markovianity. However, even in this case, from the discussion in the previous section, we know that the measurement on this large number of times steps will be temporally entangled, which may also be difficult to achieve.

Finally, we highlight the interplay of the typicality of Markovian processes and dynamical equilibration on multiple time-steps. We have drawn an analogy before with the relation between equilibration, thermalization and the typicality of canonical states, and it now seems highly plausible that an analogous relationship holds between multitime equilibration, *Markovianization* and the typicality of Markovian processes. While a fully dynamical characterization of a notion of Markovianization is still to be achieved, in the following chapter we will describe a step towards this notion as an emergence of Markovian processes by addressing the bothersome aspect in Theorem 4.4.1 of being somewhat far from physical, that is, we will show that Markovian processes still satisfy a large deviation bound even if we step away from the Haar measure and consider more physically motivated evolutions.

CHAPTER 5

Markovianization by Design

Matter is matter, neither noble nor vile, infinitely transformable, and its proximate origin is of no importance whatsoever.

– Primo Levi (*The Periodic Table*)

An important limitation of the typicality of Markovian processes in large environments by means of Theorem 4.4.1, is that it encompasses too wide a class of SE interactions, many of which can be deemed highly unphysical. Implementing a Haar random unitary requires an exponential number two-body interactions and random bits [203], so Haar random dynamics cannot be obtained efficiently in a physical setting. This seems to be at odds with the applicability of the Born-Markov assumption on a wide variety of physical models [100–103]. Forgetfulness is indeed a common feature of the world around us, and one that is crucial for doing science: without forgetfulness, repeatability would be impossible. From a somewhat philosophical standpoint, consider that if any given atom remembered its own past, then it would be unique and there would be no sense in classifying atoms and molecules.

So can we say physically relevant models satisfy a concentration of measure with respect to Markovianity?

In addressing this problem, in Ref. [3] we identified a class of isolated physical processes which approximately Markovianize. We show that, similar to the way in which quantum states thermalize, quantum processes can *Markovianize* in the sense that they can converge to a class of typical processes, satisfying a meaningful large deviations principle whenever they are undergone within a large environment and under complex enough—but not necessarily fully random—dynamics. To accomplish this, we employ large deviation bounds[1] for so-called *approximate unitary designs* derived in Ref. [204], and apply them to the process tensor formalism.

The concept of a unitary design provides a finite approximation to uniform Haar randomness and it refers to a set of unitaries that reproduce a finite number of moments of the unitary group with the Haar measure. In a physical scenario, unitary designs reproducing an ever-increasing number of moments of the Haar measure can be seen to arise naturally from seemingly simple

[1] I.e. bounds on probabilities for rare events.

situations as the evolution time increases [69, 205, 206]. One can further allow some small error in such reproduction of the Haar moments, thus rendering the respective unitary designs approximate. To further establish these ideas together with the main claim in our result, we make use of an efficient construction of an approximate unitary design with an n-qubit quantum circuit using two-qubit interactions only, showing how seemingly simple, physically motivated systems, can speedily become forgetful.

Given the ever-increasing interest and relevance in determining the breakdown of the Markovian approximation in modern experiments [118, 130, 134, 135], we discuss potential applications and extensions to our results, as well as their limitations and possible ways to overcome them.

5.1 Approximate unitary designs

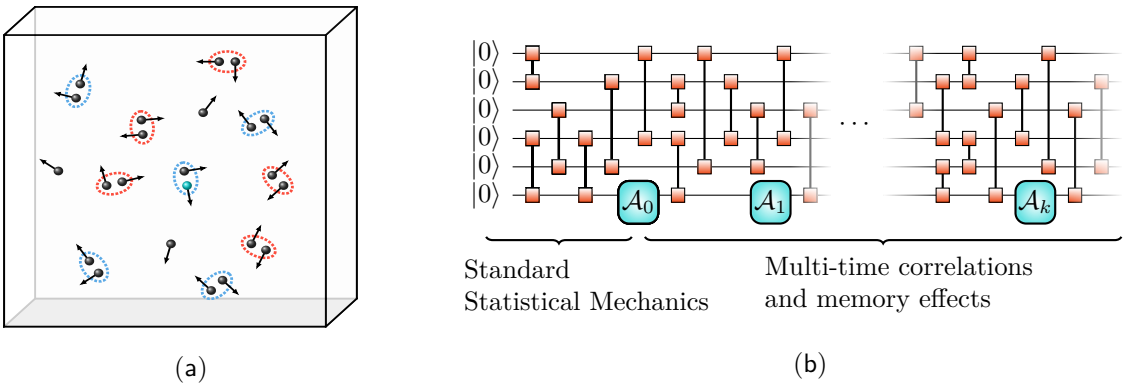
The core of the issue with our previous approach for the typicality of Markovian processes lies in the unitaries entering a process being uniformly distributed. Physical unitary evolution has a very specific structure determined by a Hamiltonian, and in this sense it can be said to be far from random. In some circumstances, however, there are physical processes which can approximate some of the statistical features of the Haar measure [18, 19, 190, 191]. For example, consider the toy model depicted in Fig. 5.1(a), comprising a dilute gas of n particles evolving autonomously in a closed box. The gas particles interact with each other in one of two ways as they randomly move inside the box. Following and intervening on an impurity particle, taken to be the system, this model can be well described by a circuit such as the one in Fig. 5.1(b). The simplicity of this system suggests that it can only *uniformly randomize* after a large number of random two-qubit interactions, progressively resembling genuine Haar random dynamics. One possible way to quantify this progressive resemblance of the Haar measure is given by the concept of unitary designs.

Definition 5.1.1 (Exact unitary t-design [204])**.** An exact unitary t-design is defined as a probability measure μ_{t} on $\mathbb{U}(d)$ such that for all positive $s \leq \mathsf{t}$, and all $d^s \times d^s$ complex matrices X,

$$\mathbb{E}_{\mathsf{t}}\left[\mathcal{V}^{\otimes s}(X)\right] = \mathbb{E}_{\mathsf{h}}\left[\mathcal{U}^{\otimes s}(X)\right], \quad \forall s \leq \mathsf{t}, \tag{5.1}$$

where $\mathcal{U}(\cdot) := U(\cdot)U^\dagger$ and $\mathcal{V}(\cdot) := V(\cdot)V^\dagger$ are unitary maps with $U, V \in \mathbb{U}(d)$.

Here, as above, the notation $\mathbb{E}_\Omega$ indicates the expectation value with respect to a given probability measure μ_Ω, i.e. we have $V \sim \mu_{\mathsf{t}}$ and $U \sim \mu_{\mathsf{h}}$ in Eq. (5.1). In the case we will be interested in, the unitary maps will correspond to SE unitaries distributed according to the either the Haar measure or a unitary design. As per the definition in Eq. (5.1), a unitary t-design reproduces up to the t-th moment over the uniform distribution given by the Haar measure. In particular, μ_{t} can consist of a finite ensemble $\{V_i, p_i\}_{i=1}^N$ of unitaries V_i and probabilities p_i, as is now common in applications such as so-called randomized benchmarking of error rates in quantum gates [207, 208].



(a) (b)

Figure 5.1: **Unitary designs, from simplicity to randomness:** (a) A toy model with dynamics given by a unitary design is that of an impurity particle (teal) immersed in a gas of n_{E} particles within a closed box, where all particles interact in pairs in one of two ways at random. (b) An SE-system with random two-qubit gate interactions only, and multiple interventions $\{\mathcal{A}_i\}$. While the standard approach towards typicality or equilibrium properties concerns the whole SE dynamics and/or a single measurement on system S, we show that complex dynamics within large environments will be highly Markovian with high probability. On the other hand, if probed enough times, information about past correlations will eventually become non-negligible.

Moreover, this definition of a unitary design can be relaxed by allowing a small error ε. In general an ε-approximate t-design, which we denote $\mu_{\mathsf{t}_\varepsilon}$, can be defined through

$$\left|\mathbb{E}_{\mathsf{t}_\varepsilon}\left[\mathcal{V}^{\otimes s}(X)\right] - \mathbb{E}_{\mathsf{h}}\left[\mathcal{U}^{\otimes s}(X)\right]\right| \leq \varepsilon, \quad \forall\, s \leq \mathsf{t} \tag{5.2}$$

for a suitable metric $|\cdot|$, with all quantities as in Eq. (5.1), now with $V \sim \mu_{\mathsf{t}_\varepsilon}$.

We specifically employed the definition of Ref. [204] for unitary designs, which uses the fact that the definition of an exact t-design can be written in terms of a balanced monomial Θ of degree less or equal to t in the components of the unitaries U.

Definition 5.1.2 (Balanced monomial [204])**.** A (unitary) balanced monomial $\Theta_s : \mathbb{U}(d) \to \mathbb{R}$ of degree s is a monomial in unitary elements with s conjugated and s unconjugated elements.

This means, for example, that $\Theta_2(U) = U_{ab}U_{cd}U^*_{ef}U^*_{hg}$ is a balanced monomial of degree 2 for given components $a, b, \ldots, g$. In this language, we have seen as well that the n-moments of the unitary group are non-vanishing only for balanced monomials. Thus, writing Eq. (5.1) in terms of matrix elements, this can be seen to be equivalent to requiring $\mathbb{E}_{\mathsf{t}}[\Theta_s(V)] = \mathbb{E}_{\mathsf{h}}[\Theta_s(U)]$ for all monomials Θ_s of degree $s \leq \mathsf{t}$. Similarly then, for an ε-approximate t-design, we adopt the definition of Ref. [204] with Eq. (5.2), which implies the following.

Definition 5.1.3 (ε-approximate unitary t-design)**.** The distribution $\mu_{\mathsf{t}_\varepsilon}$ is a ε-approximate unitary t-design if for all balanced monomials Θ_s of degree $s \leq \mathsf{t}$,

$$|\mathbb{E}_{\mathsf{t}_\varepsilon}\Theta_s(V) - \mathbb{E}_{\mathsf{h}}\Theta_s(U)| \leq \frac{\varepsilon}{d^{\mathsf{t}}}, \tag{5.3}$$

with $U, V \in \mathbb{U}(d)$.

From now on we focus on the more general approximate designs. We will see below that the degree ε to which the distribution of the unitary dynamics on $\mu_{\mathsf{t}_\varepsilon}$ differs from an exact design for given t depends on the complexity of the model.

Notice what this means for the toy model of Fig. 5.1(a): as individual random two-body interactions of each kind accumulate, what we expect is for the dynamics to start *scrambling* their information across the whole gas in the box, progressively becoming more complex and uniformly random [69]. Unitary designs give us this finite quantification of the resemblance towards uniform Haar randomness and, in this case, it can give us a precise way to account for the progressive emergence of complexity from seemingly simple individual two-body interactions.

Unitary designs for $\mathsf{t} = 2, 3$ have been widely studied [206–214] and efficient constructions are known for larger values of t [205, 211, 215]. The latter are of particular relevance, precisely as designs for large t, i.e., those with a higher complexity [69], are expected to satisfy tighter large deviation bounds. Indeed, a statement of the form $\mathbb{P}_{\mathsf{t}_\varepsilon}[\mathcal{N} \geq \delta]$ for a non-Markovianity measure $\mathcal{N}$ is expected to satisfy a bound similar to that in Theorem 4.4.1, approaching genuine concentration of measure as the level and quality of the design increases. Such large deviation bounds –which here simply refers to bounds on probabilities for *rare* events– over approximate unitary designs were derived in a general form in Ref. [204] for a polynomial function satisfying a concentration of measure bound.

Theorem 5.1.4 (Large deviation bounds for t-designs [204]). Let $\mathcal{X}$ be a polynomial of degree T. Let $f(U) = \sum_i \alpha_i \Theta_{s_i}(U)$ where $\Theta_{s_i}(U)$ are monomials and let $\alpha(f) = \sum_i |\alpha_i|$. Suppose that f has probability concentration

$$\mathbb{P}_{\mathsf{h}}[|f - \zeta| \geq \delta] \leq C \exp\left(-\mathscr{C}\delta^2\right), \tag{5.4}$$

and let $\mu_{\mathsf{t}_\varepsilon}$, be an ε-approximate unitary t-design, then

$$\mathbb{P}_{\mathsf{t}_\varepsilon}[|f - \zeta| \geq \delta] \leq \frac{1}{\delta^{2m}} \left(C \left(\frac{m}{\mathscr{C}}\right)^m + \frac{\varepsilon}{d^{\mathsf{t}}} (\alpha + |\zeta|)^{2m} \right), \tag{5.5}$$

for any integer m with $2m\mathsf{T} \leq \mathsf{t}$.

This is the most general expression, where ζ can be any quantity, in particular the expectation of f. The main idea from this result in Ref. [204] (similarly applied before in Ref. [215]) is that given a $\mu_{\mathsf{t}_\varepsilon}$ distribution as an ε-approximate unitary t-design and a concentration result for a polynomial f of degree T, then one can compute

$$\mathbb{E}_{\mathsf{t}_\varepsilon}[f^m] = \mathbb{E}_{\mathsf{h}}[f^m] + g(\varepsilon, \mathsf{t}, f), \tag{5.6}$$

where $m \leq \mathsf{t}/2\mathsf{T}$. Using Markov's inequality we have

$$\mathbb{P}_{\mathsf{t}_\varepsilon}(f \geq \delta) = \mathbb{P}_{\mathsf{t}_\varepsilon}(f^m \geq \delta^m) \leq \frac{\mathbb{E}_{\mathsf{t}_\varepsilon}[f^m]}{\delta^m} = \frac{1}{\delta^m}\left[\mathbb{E}_{\mathsf{h}}[f^m] + g(\varepsilon, \mathsf{t}, f)\right], \tag{5.7}$$

which is the form of the main large deviations bound in Eq. (5.5). More precisely, the other two main results that come along with the proof of Theorem 5.1.4 given in Ref. [204] and allowing to compute the right hand-side of Eq. (5.7) are the following.

Lemma 5.1.5 (3.4 of Ref. [204]). Let $\mathcal{X}$ be a polynomial of degree T and ζ any constant. Let $f(U) = \sum_i \alpha_i \Theta_{s_i}(U)$ where $\Theta_{s_i}(U)$ are monomials and let $\alpha(f) = \sum_i |\alpha_i|$. Then for an integer m such that $2m\mathsf{T} \leq \mathsf{t}$ and $\mu_{\mathsf{t}_\varepsilon}$ an ε-approximate unitary t-design,

$$\mathbb{E}_{\mathsf{t}_\varepsilon}\left[|f-\zeta|^{2m}\right] \leq \mathbb{E}_{\mathsf{h}}\left[|f-\zeta|^{2m}\right] + \frac{\varepsilon}{d^{\mathsf{t}}}\left(\alpha + |\zeta|\right)^{2m}. \tag{5.8}$$

Lemma 5.1.6 (5.2 of Ref. [204]). Let X be any non-negative random variable with probability concentration

$$\mathbb{P}(X \geq \delta + \gamma) \leq C\exp(-\mathscr{I}\,\delta^2), \tag{5.9}$$

where $\gamma \geq 0$, then

$$\mathbb{E}[X^m] \leq C\left(\frac{2m}{\mathscr{I}}\right)^{m/2} + (2\gamma)^m, \tag{5.10}$$

for any $m > 0$.

So, in essence, to solve our problem of finding a large deviations bound for quantum processes, what we can do given the results from Ref. [204], is to determine the right-hand sides of Eq. (5.8) and Eq. (5.10) by phrasing our measure of non-Markovianity and all the other relevant quantities in such terms.

5.2 Large deviations on unitary designs around Markovian processes

Let us now then revisit Theorem 4.4.1. For concreteness, in Ref. [3] we focused only in the random interaction picture ($\mathcal{U}_i \neq \mathcal{U}_j$). As we mention in the previous chapter, the choice of the non-Markovianity measure as the trace distance between a quantum process and the closest Markovian one in Ref. [2] was motivated mainly due to its relation with the equilibration and state typicality results; strictly speaking this measure gives the distinguishability between explicitly constructed Choi states of corresponding process tensors and has no operational meaning. In this case a more suitable choice is the so-called *diamond norm*.

While trace distance is a natural metric for differentiating two quantum states, the natural distance for differentiating two quantum channels is the diamond norm, which allows for the use of additional ancillas, as in discussed in Ref. [216]. We are interested in optimally distinguishing between a non-Markovian process from a Markovian one, which leads to the multitime diamond distance:

$$\mathcal{N}_\blacklozenge := \frac{1}{2}\min_{\Upsilon^{(\mathrm{M})}} \|\Upsilon - \Upsilon^{(\mathrm{M})}\|_\blacklozenge, \tag{5.11}$$

where

$$\|\Upsilon\|_\blacklozenge := \sup_{\mathcal{J}=\{\mathcal{O}_i\}} \left\| \sum_i \mathrm{tr}[\mathcal{O}_i \Upsilon \otimes \mathbb{1}]|i\rangle\langle i| \right\|_1 \tag{5.12}$$

is a generalized diamond norm [158, 217–220], with supremum over the instrument $\mathcal{J} = \{\mathcal{O}_i\}$. This definition generalizes the diamond norm for quantum channel distinguishability [221] (also called cb-norm in Ref. [222] or completely bounded trace norm in Ref. [11]), reducing to it

for a single step process tensor, and similarly being interpreted as the optimal probability to discriminate a process from the closest Markovian one in a single shot, given any set of measurements together with an ancilla.

As we saw in the previous two chapters, this choice of non-Markovianity measure is not unique; more generally for any Schatten p-norm $\|X\|_p := \mathrm{tr}(|X|^p)^{\frac{1}{p}}$ we can define a family of non-Markovianity measures as

$$\mathcal{N}_p := \frac{1}{2} \min_{\Upsilon^{(\mathrm{M})}} \|\Upsilon - \Upsilon^{(\mathrm{M})}\|_p, \tag{5.13}$$

as done with $p = 1$ in Theorem 4.4.1, whenever Υ is normalized such that $\mathrm{tr}[\Upsilon] = \mathrm{tr}[\Upsilon^{(\mathrm{M})}] = 1$. Then, we have the hierarchy $\mathcal{N}_1 \geq \mathcal{N}_2 \geq \ldots$, induced by that of the Schatten norms in Eq. (20). In particular, we have the relation $d_{\mathsf{S}}^{-2k-1}\mathcal{N}_\blacklozenge \leq \mathcal{N}_1 \leq \mathcal{N}_\blacklozenge$. This implies that the result in Theorem 4.4.1, can be written equivalently as

$$\mathbb{P}_{\mathsf{h}}\left[\mathcal{N}_\blacklozenge \geq d_{\mathsf{S}}^{2k+1}\mathcal{B} + \delta\right] \leq \exp\left\{-4\,\mathscr{C}\,\delta^2 d_{\mathsf{S}}^{-2(2k+1)}\right\}, \tag{5.14}$$

where

$$\mathscr{C} = \frac{d_{\mathsf{SE}}(k+1)}{16}\left(\frac{d_{\mathsf{S}}-1}{d_{\mathsf{S}}^{k+1}-1}\right)^2, \tag{5.15}$$

is the constant related to the Lipschitz constant of $\mathcal{N}_1$ in Theorem 4.4.1 up to a dimensional multiplicative factor of 4, and where we now simply denote $\mathcal{B} \geq \mathbb{E}[\mathcal{N}_1]$ the upper bound on the non-Markovianity $\mathcal{N}_1$ given in Eq. (4.25) with average purity in the random interaction case given in Eq. (4.27). In Ref. [3] we refer to $\mathscr{C}$ as *the* Lipschitz constant of $\mathcal{N}_1$, but strictly speaking this is an inverse and rescaled version of $\mathscr{L}$ in Eq. (4.46). We recall that, holding everything else constant, we have the limiting cases $\mathcal{B} = 0$ when $d_{\mathsf{E}} \to \infty$ and $\mathcal{B} = 1$ when $k \to \infty$, so that the expected non-Markovianity vanishes in the small subsystem limit and becomes loosest in the long time limit case.

Theorem 5.2.1 (Markovianization with approximate unitary designs [3]). Given a k-step process Υ on a d_{S} dimensional subsystem, generated from global unitary d_{SE} dimensional SE dynamics distributed according to an ε-approximate unitary t-design $\mu_{\mathsf{t}_\varepsilon}$, the likelihood that its non-Markovianity exceeds any $\delta > 0$ is bounded as

$$\mathbb{P}_{\mathsf{t}_\varepsilon}\left[\mathcal{N}_\blacklozenge \geq \delta\right] \leq \mathsf{B}, \tag{5.16}$$

where B is defined as

$$\mathsf{B} := \frac{d_{\mathsf{S}}^{3m(2k+1)}}{\delta^{2m}}\left[\left(\frac{m}{\mathscr{C}}\right)^m + (2\mathcal{B})^{2m} + \frac{\varepsilon}{d_{\mathsf{SE}}^{\mathsf{t}}}\eta^{2m}\right], \tag{5.17}$$

for any $m \in (0, \mathsf{t}/4]$ and

$$\eta := (d_{\mathsf{SE}}^4 d_{\mathsf{S}}^{2k} + d_{\mathsf{S}}^{-(2k+1)})/4, \tag{5.18}$$

where $\mathscr{C}$ is defined in Eq. (5.15) and $\mathcal{B}$ an upper bound on the expected norm-1 non-Markovianity $\mathbb{E}_{\mathsf{h}}[\mathcal{N}_1]$, defined in Eq. (4.25) with average noisiness given by Eq. (4.27).

Proof. The overall strategy is as in Ref. [204]: a bound on the moments $\mathbb{E}_{\mathsf{t}_\varepsilon}[\mathcal{N}_\blacklozenge^m]$ is given in terms of $\mathcal{B}$, $\mathscr{C}$ and η, followed by Markov's inequality. As we mention by the end of the previous section, this amounts to determining the right-hand sides of Eq. (5.8) and Eq. (5.10). For us the relevant quantity is the non-Markovianity $\mathcal{N}_\blacklozenge$, which is a fairly hard quantity to work with, however, we can use the relationship of this norm with the family of Schatten-norm non-Markovianity measures to compute relevant bounds.

Unsurprisingly, the hardest quantity to compute is the sum of moduli of coefficients of the non-Markovianity expressed as a polynomial in the unitaries, so the most accessible way to do this is to turn to the non-Markovianity $\mathcal{N}_2$.

In general, $\|X\|_1 \geq \|X\|_2$ for Schatten norms, so given the concentration result for $\mathcal{N}_1$ in Theorem 4.4.1 and the upper bound $\mathcal{B} \geq \mathbb{E}[\mathcal{N}_1]$ given in Eq. (4.25), this also implies

$$\mathbb{P}_{\mathsf{h}}\left[\mathcal{N}_2 \geq \mathcal{B} + \delta\right] \leq \mathrm{e}^{-\mathscr{C}\delta^2}, \tag{5.19}$$

so that in turn Eq. (5.10) implies that

$$\mathbb{E}_{\mathsf{h}}\left[\mathcal{N}_2^{2m}\right] \leq \left(\frac{m}{\mathscr{C}}\right)^m + (2\mathcal{B})^{2m} = \left[\frac{16m}{(k+1)d_{\mathsf{SE}}}\left(\frac{d_{\mathsf{S}}^{k+1}-1}{d_{\mathsf{S}}-1}\right)^2\right]^m + (2\mathcal{B})^{2m}, \tag{5.20}$$

for any $m > 0$, and where we have absorbed a multiplicative factor of 4 in the definition of $\mathscr{C}$, which is now as in Eq. (5.15).

For the case of all unitaries at each step being independently sampled, $\mathcal{N}_2^2$ is a polynomial of degree $p = 2$ when the unitaries are all distinct (random interaction type). We can thus take $\mathcal{N}_2^2$ and apply Lemma 5.1.5 of Ref. [204] for a unitary t-design $\mu_{\mathsf{t}_\varepsilon}$ with $\mathsf{t} \geq 4m$, which holds for real $m > 0$,[2] as

$$\mathbb{E}_{\mathsf{t}_\varepsilon}\left[\mathcal{N}_2^{2m}\right] \leq \mathbb{E}_{\mathsf{h}}\left[\mathcal{N}_2^{2m}\right] + \frac{\varepsilon}{d_{\mathsf{SE}}^{\mathsf{t}}}\,\eta^{2m}, \tag{5.21}$$

where η is the sum of the moduli of the coefficients of

$$\mathcal{N}_2^2 = \left(\frac{1}{2}\min_{\Upsilon^{(\mathrm{M})}}\|\Upsilon - \Upsilon^{(\mathrm{M})}\|_2\right)^2 \leq \frac{1}{4}\|\Upsilon - \frac{\mathbb{1}}{d_{\mathsf{S}}^{2k+1}}\|_2^2 = \frac{1}{4}\left[\mathrm{tr}\left(\Upsilon^2\right) - d_{\mathsf{S}}^{-(2k+1)}\right], \tag{5.22}$$

so our problem now boils down to computing the sum of the moduli of the coefficients of the purity, or noisiness, of the process.

Let us explicitly write the process Υ as a function of the set of unitary maps $\mathfrak{U} := \{\mathcal{U}_i\}_{i=0}^k$, i.e.

$$\begin{aligned}\Upsilon[\mathfrak{U}] &= \mathrm{tr}_{\mathsf{E}}\left[\mathcal{U}_k\,\mathcal{S}_k\,\mathcal{U}_{k-1}\mathcal{S}_{k-1}\cdots\mathcal{U}_1\mathcal{S}_1\left(\rho\otimes\Psi^{\otimes k}\right)\right]\\ &= \sum_{\alpha,\ldots,\delta}\mathrm{tr}_{\mathsf{E}}\left[U_k\,\mathfrak{S}_{\alpha_k\beta_k}\cdots U_1\,\mathfrak{S}_{\alpha_1\beta_1}\,\rho\,\mathfrak{S}_{\delta_1\gamma_1}\,U_1^\dagger\cdots\mathfrak{S}_{\delta_k\gamma_k}\,U_k^\dagger\right]\otimes|\beta_1\alpha_1\cdots\beta_k\alpha_k\rangle\langle\delta_1\gamma_1\cdots\delta_k\gamma_k|,\end{aligned} \tag{5.23}$$

[2] The proof of Lemma 5.1.5 in Ref. [204] requires m to be an integer through the multinomial theorem; in the notation of the cited paper, this can be relaxed to be a real number when $\mu = 0$ and applying the multinomial theorem for a real power: convergence will require (an ordering such that) $|\alpha_t\mathbb{E}M_t| > 2^{1-n}|\alpha_{t-n}\mathbb{E}M_{t-n}|$ for each $n = 1,\ldots,t-1$ for both the approximate design and Haar expectations.

where the sum runs over all Greek indices from 1 to d_{S}, with Ψ being a maximally entangled state acting in the respective d_{S}-dimensional ancillary spaces $\mathsf{A}_i\mathsf{B}_i$, where $\mathcal{S}_i$ are swaps between S and ancillary system A_i at time-step i, and where $\mathfrak{S}_{\alpha\beta} = \mathbb{1}_{\mathsf{E}} \otimes |\alpha\rangle\langle\beta|$. Here as well, full detail can be revisited around the definition in Eq. (2.52).

Now, the standard approach to compute the sum of the moduli of the coefficients of a given polynomial is to evaluate on an argument (here a $d_{\mathsf{SE}} \times d_{\mathsf{SE}}$ matrix) full of ones (so that all single monomials equal to one) and take each summand to the corresponding modulus. We follow this approach, however, we first notice that the environment part in Eq. (5.23) is just a product of the environment parts of all unitaries and initial state.[3] This implies that at most d_{E} terms need to be set to one and we can evaluate Υ in a set of matrices $\mathscr{J} = \{\mathbb{1}_{\mathsf{E}} \otimes J_{\mathsf{S}}, \cdots, \mathbb{1}_{\mathsf{E}} \otimes J_{\mathsf{S}}, J_{\mathsf{E}} \otimes J_{\mathsf{S}}\}$ with J a matrix with each element equal to one in the respective E or S systems: let $\rho = \sum \rho_{ese's'}|es\rangle\langle e's'|$, then

$$\begin{aligned}
\Upsilon[\mathscr{J}] &= d_{\mathsf{S}}^{-k} \sum \rho_{ese's'} \operatorname{tr}\left[d_{\mathsf{E}} J_{\mathsf{E}}|e\rangle\langle e'|\right] J_{\mathsf{S}}|\alpha_k\rangle\langle\beta_k| \cdots |\alpha_1\rangle\langle\beta_1|J_{\mathsf{S}}|s\rangle\langle s'|J_{\mathsf{S}}|\delta_1\rangle\langle\gamma_1| \cdots |\delta_k\rangle\langle\gamma_k|J_{\mathsf{S}} \\
&\qquad \otimes |\beta_1\alpha_1 \cdots \beta_k\alpha_k\rangle\langle\delta_1\gamma_1 \cdots \delta_k\gamma_k| \\
&= \frac{d_{\mathsf{E}}}{d_{\mathsf{S}}^k} \sum \rho_{ese's'}\, J_{\mathsf{S}}|\alpha_k\rangle\langle\beta_k| \cdots |\alpha_1\rangle\langle\beta_1|J_{\mathsf{S}}|s\rangle\langle s'|J_{\mathsf{S}}|\delta_1\rangle\langle\gamma_1| \cdots |\delta_k\rangle\langle\gamma_k|J_{\mathsf{S}} \\
&\qquad \otimes |\beta_1\alpha_1 \cdots \beta_k\alpha_k\rangle\langle\delta_1\gamma_1 \cdots \delta_k\gamma_k|,
\end{aligned} \tag{5.24}$$

and hence (we now omit the subindex S on the J matrices for simplicity),

$$\begin{aligned}
\left(\frac{d_{\mathsf{S}}^k}{d_{\mathsf{E}}}\right)^2 \operatorname{tr}[\Upsilon^2(\mathscr{J})] &= \sum \rho_{ese's'}\rho_{\epsilon\sigma\epsilon'\sigma'} \operatorname{tr}[J|\alpha_k\rangle\langle\beta_k| \cdots |\alpha_1\rangle\langle\beta_1|J|s\rangle\langle s'|J|\delta_1\rangle\langle\gamma_1| \cdots \\
&\qquad J|\delta_k\rangle\langle\gamma_k|J^2|\gamma_k\rangle\langle\delta_k|J \cdots |\gamma_1\rangle\langle\delta_1|J|\sigma\rangle\langle\sigma'|J|\beta_1\rangle\langle\alpha_1|J \cdots |\beta_k\rangle\langle\alpha_k|J] \\
&= d_{\mathsf{S}}^2 \sum \rho_{ese's'}\rho_{\epsilon\sigma\epsilon'\sigma'} \operatorname{tr}[J|\alpha_k\rangle\langle\beta_k| \cdots |\alpha_1\rangle\langle\beta_1|J|s\rangle\langle s'|J|\delta_1\rangle\langle\gamma_1| \cdots \\
&\qquad \langle\gamma_{k-1}|J|\delta_k\rangle\langle\delta_k|J|\gamma_{k-1}\rangle \cdots |\gamma_1\rangle\langle\delta_1|J|\sigma\rangle\langle\sigma'|J|\beta_1\rangle\langle\alpha_1|J \cdots |\beta_k\rangle\langle\alpha_k|J] \\
&= d_{\mathsf{S}}^{2k+1} \sum \rho_{ese's'}\rho_{\epsilon\sigma\epsilon'\sigma'} \operatorname{tr}[J|\alpha_k\rangle\langle\beta_k| \cdots |\alpha_1\rangle\langle\beta_1|J|s\rangle\langle s'|J|\sigma\rangle\langle\sigma'|J|\beta_1\rangle\langle\alpha_1|J \cdots |\beta_k\rangle\langle\alpha_k|J] \\
&= d_{\mathsf{S}}^{2k+3} \sum \rho_{ese's'}\rho_{\epsilon\sigma\epsilon'\sigma'} \langle\beta_k|J|\alpha_{k-1}\rangle \cdots |\alpha_1\rangle\langle\beta_1|J|s\rangle\langle s'|J|\sigma\rangle\langle\sigma'|J|\beta_1\rangle\langle\alpha_1| \cdots \langle\alpha_{k-1}|J|\beta_k\rangle \\
&= d_{\mathsf{S}}^{2k+5} \sum \rho_{ese's'}\rho_{\epsilon\sigma\epsilon'\sigma'} \langle\beta_{k-1}|J|\alpha_{k-2}\rangle \cdots |\alpha_1\rangle\langle\beta_1|J|s\rangle\langle s'|J|\sigma\rangle\langle\sigma'|J|\beta_1\rangle\langle\alpha_1| \cdots \langle\alpha_{k-1}|J|\beta_{k-1}\rangle \\
&= d_{\mathsf{S}}^{2(2k+1)} \sum \rho_{ese's'}\rho_{\epsilon\sigma\epsilon'\sigma'},
\end{aligned} \tag{5.25}$$

where to obtain the second line we used the fact that $J^n = d_{\mathsf{S}}^{n-1}J$ for positive integers n, here applied for $n = 2$, together with the trace over system S given by $\sum\langle\gamma_k| \cdot |\gamma_k\rangle$. This is similarly done to get the third line by $\sum |\delta_k\rangle\langle\delta_k| = \mathbb{1}_{\mathsf{S}}$, and taking the trace summing over $|\gamma_{k-1}\rangle$, which can subsequently be done for all $|\gamma_i\rangle$ and $|\delta_i\rangle$. For the fourth line the cyclicity of the trace was used, followed by taken an identity summing up over $|\alpha_k\rangle$, using $J^2 = d_{\mathsf{S}}J$ and taking the trace.

[3] Let $U = \sum U^{es}_{e's'}|es\rangle\langle e's'|$ where $|e\rangle$ and $|s\rangle$ are E and S bases. Unitarity then implies $\sum \overline{U}^{ab}_{es}U^{ab}_{e\sigma} = \delta_{e\epsilon}\delta_{s\sigma}$, and so this means that $\operatorname{tr}_{\mathsf{E}}[V\mathfrak{S}_{\alpha\beta}U\rho U^\dagger\mathfrak{S}_{\gamma\delta}V^\dagger] = \sum V^{es}_{e's'}\overline{V}^{e\sigma}_{e'\sigma'}U^{e's_2}_{bs'_2}\overline{U}^{e'\sigma_2}_{b\sigma'_2}\rho^{br}_{bt}\,\phi(S)$ where $\phi(S)$ stands for the system S part; for each b index the rest of the terms are summed over e. This generalizes similarly for any number of unitaries.

This can be done through all remaining steps, giving the last equality. This, together with Eq. (5.22), implies that (now writing simply i, j for SE indices),

$$\begin{aligned} 4\eta &\leq d_{\mathsf{E}}^2 d_{\mathsf{S}}^{2(k+1)} \left(\sum |\rho_{ij}|\right)^2 + \frac{1}{d_{\mathsf{S}}^{2k+1}} \\ &\leq d_{\mathsf{E}}^4 d_{\mathsf{S}}^{2(k+2)} \sum |\rho_{ij}|^2 + \frac{1}{d_{\mathsf{S}}^{2k+1}} \\ &\leq d_{\mathsf{E}}^4 d_{\mathsf{S}}^{2(k+2)} + \frac{1}{d_{\mathsf{S}}^{2k+1}}, \end{aligned} \tag{5.26}$$

where in the second line we used $\|X\|_1^2 \leq d\|X\|_2^2$ for element-wise norms $\|X\|_p^p = (\sum |x_{ij}|^p)$ and in the third line we used $\|\rho\|_2^2 \leq 1$.

We can finally put everything together as follows. As $d_{\mathsf{S}}^{-2k1-1}\mathcal{N}_\blacklozenge \leq \mathcal{N}_1 \leq \sqrt{d_{\mathsf{S}}^{2k+1}}\mathcal{N}_2$, also for $0 < m \leq t/4$,

$$\begin{aligned} &\mathbb{P}_{\mathsf{t}_\varepsilon}[\mathcal{N}_\blacklozenge \geq \delta] \leq \mathbb{P}_{\mathsf{t}_\varepsilon}\left[\sqrt{d_{\mathsf{S}}^{3(2k+1)}}\,\mathcal{N}_2 \geq \delta\right] = \mathbb{P}_{\mathsf{t}_\varepsilon}\left[\mathcal{N}_2^{2m} \geq \frac{\delta^{2m}}{d_{\mathsf{S}}^{3m(2k+1)}}\right] \\ &\leq \frac{d_{\mathsf{S}}^{3m(2k+1)}\,\mathbb{E}_{\mathsf{t}_\varepsilon}\mathcal{N}_2^{2m}}{\delta^{2m}} \leq \left(\frac{d_{\mathsf{S}}^{3(2k+1)}}{\delta^2}\right)^m \left[\left(\frac{4m}{\mathscr{C}}\right)^m + (2\mathcal{B})^{2m} + \frac{\epsilon}{d_{\mathsf{SE}}^t}\eta^{2m}\right] \\ &= \left(\frac{d_{\mathsf{S}}^{3(2k+1)}}{\delta^2}\right)^m \left\{\left[\frac{16m}{(k+1)\,d_{\mathsf{SE}}}\left(\frac{d_{\mathsf{S}}^{k+1}-1}{d_{\mathsf{S}}-1}\right)^2\right]^m + (2\mathcal{B})^{2m} + \frac{\epsilon}{16^m d_{\mathsf{SE}}^t}\left(d_{\mathsf{E}}^4 d_{\mathsf{S}}^{2(k+2)} + \frac{1}{d_{\mathsf{S}}^{2k+1}}\right)^{2m}\right\}, \end{aligned} \tag{5.27}$$

where in the third line we used Markov's inequality. ■

The choice of $0 < m \leq \mathsf{t}/4$ can be made to optimize the right-hand-side of the inequality, which ideally should be small whenever δ is. The term $d_{\mathsf{S}}^{3(2k+1)}/\delta^2$ arises from bounding $\mathcal{N}_\blacklozenge$ and Markov's inequality, while the three summands within square brackets will be small provided *i)* $\mathscr{C}$ is large, *ii)* $\mathcal{B}$ is small and *iii)* the unitary design sufficiently small ε and large t is well-approximate and high enough. For conditions *i)* and *ii)*, as in Chapter 4, we require a fixed k such that $d_{\mathsf{E}} \gg d_{\mathsf{S}}^{2k+1}$: this implies $\mathcal{B} \approx 0$ so that

$$\begin{aligned} \mathbb{P}_{\mathsf{t}_\varepsilon}[\mathcal{N}_\blacklozenge \geq \delta] &\lesssim \left(\frac{d_{\mathsf{S}}^{3(2k+1)}}{\delta^2}\right)^m \left\{\left[\frac{16m}{(k+1)\,d_{\mathsf{SE}}}\left(\frac{d_{\mathsf{S}}^{k+1}-1}{d_{\mathsf{S}}-1}\right)^2\right]^m + \frac{\varepsilon}{16^m d_{\mathsf{SE}}^{\mathsf{t}}}\left(d_{\mathsf{E}}^4 d_{\mathsf{S}}^{2(k+2)} + \frac{1}{d_{\mathsf{S}}^{2k+1}}\right)^{2m}\right\} \\ &\approx \left\{\left[\frac{16m}{\delta^2(k+1)}\frac{d_{\mathsf{S}}^{2(4k+1)}}{d_{\mathsf{E}}}\right]^m + \varepsilon\frac{d_{\mathsf{E}}^{8m-\mathsf{t}} d_{\mathsf{S}}^{m(10k+11)-\mathsf{t}}}{\delta^{2m}16^m}\right\}. \end{aligned} \tag{5.28}$$

Now, supposing the t-design is exact, i.e. $\varepsilon = 0$, we require $m \leq \delta^2\frac{(k+1)d_{\mathsf{E}}}{16\,d_{\mathsf{S}}^{6k}}$, together with $m \leq \mathsf{t}/4$. On the other hand if ε is non-zero, we require

$$\varepsilon \ll \left[\delta^2\left(\frac{2}{d_{\mathsf{E}}^2 d_{\mathsf{S}}^{(10k+11)/4}}\right)^4\right]^m d_{\mathsf{SE}}^{\mathsf{t}}. \tag{5.29}$$

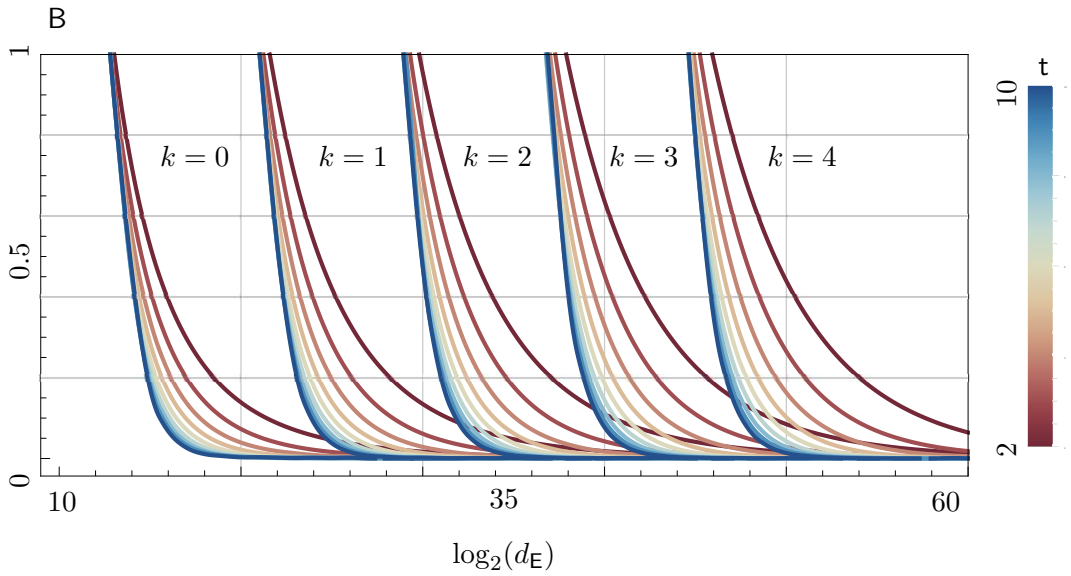


Figure 5.2: **Large deviations bound on non-Markovianity with unitary designs:** Upper bound B on $\mathbb{P}_{\mathsf{t}_\varepsilon}[\mathcal{N}_\blacklozenge \geq 0.1]$ defined by Eq. (5.16) against $\log_2(d_{\mathsf{E}})$ for a subsystem qubit undergoing a joint closed approximate unitary design interaction at each step. We fix an $\varepsilon = 10^{-12}$ approximate unitary t-design for different values $2 \leq \mathsf{t} \leq 10$ and fixed values of timesteps k, optimizing m for each case.

The choice of real m is only restricted by $0 < m \leq \mathsf{t}/4$, but is otherwise arbitrary. The right-hand side of Eq. (5.17) is not monotonic in m over all the remaining parameters, so it won't always be optimal for some fixed choice. One may thus optimize the choice of m numerically for each particular case.

Overall, the bound in Eq. (5.16) approaches concentration whenever d_{E} is large relative to d_{S} and k, together with large enough t, as shown in Fig. 5.2. Therefore, the vast majority of processes sampled from such a t-design are indistinguishable from Markovian ones in this limit. Despite the final bound in Theorem 5.2.1 being seemingly complicated, we can apply it to state-of-the-art efficient costructions of simple circuits that generate unitary designs. We now show how these processes can be modelled in terms of random circuits.

5.3 Markovianization by circuit design

While no explicit sets forming unitary t-designs for $\mathsf{t} \geq 4$ are known to date, several efficient constructions generating approximate unitary designs by quantum circuits are known. Using these constructions we can highlight the physical implications of the theorem above. We begin by discussing the details of one such construction. As suggested in Fig. 5.1(b), this construction only requires two-qubit interactions and, under certain conditions, yields an approximate unitary design, from which we can use Eq. (5.16) in Theorem 5.2.1 to verify that Markovianization emerges.

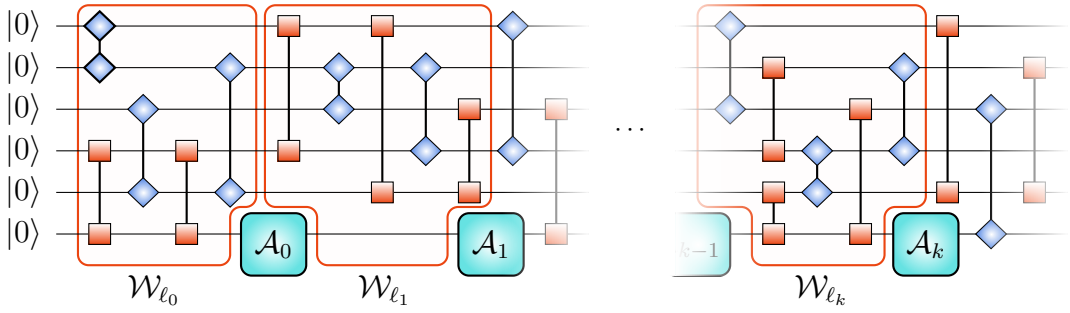


Figure 5.3: **Markovianization by circuit design:** Cartoon of a quantum process which can Markovianize under only two different types of 2-qubit interaction dynamics. For an n-qubit system, the unitaries $\mathcal{W}_\ell$ defined by Eq. (5.33) generate an ε-approximate unitary $\mathtt{t}$-design whenever $\ell \geq \mathtt{t} - \log_2(\varepsilon)/n$, as found in Ref. [205]. This can be thought as stemming from repeated alternate applications of random 2-qubit gates diagonal in only two Pauli bases (rectangles and diamonds). A qubit probed with a set of operations $\{\mathcal{A}_i\}$ on a system undergoing ε-approximate unitary $\mathtt{t}$-design dynamics $\mathcal{W}_\ell$ on a large environment will Markovianize for small design error ε and large complexity $\mathtt{t}$ as specified in the main text.

In our manuscript we focused specifically on Result 2 of Ref. [205], where a circuit with interactions mediated by two-qubit diagonal gates with three random parameters is introduced. To begin with, an efficient approximation for a unitary design on a system composed of n-qubits is shown in Ref. [205] for a circuit labeled RDC($\mathcal{I}_2$), where the name stands for *Random Diagonal Circuit*, and refers to a circuit where $\mathcal{I}_2 = \{I_i\}$ is a set of subsets of qubit labels $I_i \subset \{1, \ldots, n\}$, such that $|I_i| = 2$, i.e., at step i, I_i picks a pair of qubits, to which a Pauli-Z-diagonal gate with three random parameters is applied. The same idea follows in general for an arbitrary number of qubits $|I_i|$, but here we focus on the case of two-qubit interactions only. This construction can already be seen in Ref. [206] as arising from only two types of random diagonal interactions, which can be simplified into a product of Z-diagonal ones. The brilliance in this construction lies in the intuition that repeated alternate applications of these diagonal gates quickly randomizes the system.

This idea now fully captures the gas scenario depicted in Fig. 5.1(a), where we only have two types of random two-body interactions repeatedly occurring, and we focus on one of the particles of the gas. We can more concretely illustrate this idea in Fig. 5.3, where we depict an n-qubit SE composite with k interventions on one of the qubits, with the unitary interactions within the circuit being only between pairs of qubits and of only two kinds.

A particular case which further simplifies things by writing all two-qubit gates in a diagonal form is denoted by $\mathrm{RDC}^{(\mathtt{t})}_{\mathrm{disc}}(\mathcal{I}_2)$, where the subscript "disc" refers to *discrete* sets from which the parameters of the two-qubit gates will be sampled, and the superscript $\mathtt{t}$ is a natural number which determines this set. Specifically, all gates in $\mathrm{RDC}^{(\mathtt{t})}_{\mathrm{disc}}(\mathcal{I}_2)$ have the simplified form

$$(\mathrm{diag}\{1, e^{i\phi_1}\} \otimes \mathrm{diag}\{1, e^{i\phi_2}\})\ \mathrm{diag}\{1, 1, 1, e^{i\vartheta}\}, \tag{5.30}$$

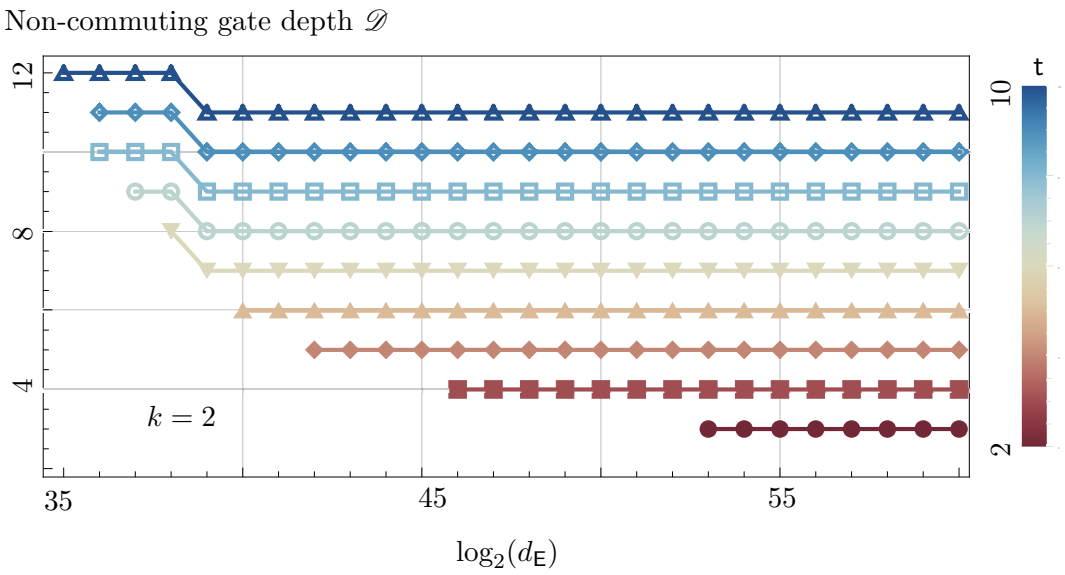


Figure 5.4: **Scaling of the non-commuting gate depth** $\mathscr{D}$, as in Eq. (5.35), equivalent to that of the minimum amount of repetitions ℓ in the $n = n_{\mathsf{E}} + 1$ qubit circuit $\mathcal{W}_\ell$ on SE, plotted against the environment qubits, $n_{\mathsf{E}} = \log_2(d_{\mathsf{E}})$, to generate an $\varepsilon = 10^{-12}$ approximate unitary t-design for $2 \leq \mathsf{t} \leq 10$, such that for a single-qubit system undergoing a process with $k = 2$ timesteps, $\mathbb{P}_{\mathsf{t}_\varepsilon}\left[\mathcal{N}_\blacklozenge \geq 0.1\right] \leq \mathsf{B} \leq 0.01$.

where diag denotes Pauli-Z diagonal, and with

$$\phi_1, \phi_2 \sim \{2\pi\, m/(\mathsf{t}+1) : m \in \{0, \ldots, \mathsf{t}\}\}, \tag{5.31}$$

chosen independently from such discrete set, and similarly

$$\vartheta \sim \{2\pi\, m/(\lfloor \mathsf{t}/2 \rfloor + 1) : m \in \{0, \ldots, \lfloor \mathsf{t}/2 \rfloor\}\}. \tag{5.32}$$

Notice that despite the apparent complexity of this construction, it is still just a circuit comprised only of 2-qubit diagonal gates with only three random parameters each, and therein lies its simplicity. Let us then state the main Result of Ref. [205] that we applied on our result for Markovianization.

Theorem 5.3.1 (Main Result 2 of Ref. [205]). Let $\mathsf{H}_n = \mathsf{H}^{\otimes n}$ be n copies of the Hadamard gate,[4] then for an n-qubit system, when t is of order $\sqrt{n}$, a circuit of the form

$$\mathcal{W}_\ell := \left(\mathrm{RDC}^{(\mathsf{t})}_{\mathrm{disc}}(\mathcal{I}_2)\ \mathsf{H}_n\right)^{2\ell}\ \mathrm{RDC}^{(\mathsf{t})}_{\mathrm{disc}}(\mathcal{I}_2), \tag{5.33}$$

yields an ε-approximate unitary t-design if

$$\ell \geq \mathsf{t} - \frac{\log_2(\varepsilon)}{n}, \tag{5.34}$$

up to leading order in n and t.

[4] In the computational basis the Hadamard gate is given by $\mathsf{H} = (|0\rangle\langle 0| + |1\rangle\langle 0| + |0\rangle\langle 1| - |1\rangle\langle 1|)/\sqrt{2}$.

Furthermore, of great relevance in this result is the fact that all the 2-qubit gates in each repetition of $\mathcal{W}_\ell$, except those in H_n, can be applied simultaneously because they commute [206, 223]. Therefore, if $\mathcal{W}_\ell$ yields an approximate unitary design as above, the order of the non-commuting gate depth $\mathscr{D}$, defined in Ref. [223] as the circuit depth when each commuting part of the circuit is counted as a single part, will coincide with the bound on the order of the number of repetitions ℓ. That is, the non-commuting gate depth asymptotes to

$$\mathscr{D} \sim \mathsf{t} - \frac{\log_2(\varepsilon)}{n}. \tag{5.35}$$

We can now take the system from the toy model of Fig. 5.1(a) as given by a spin locally interacting with a large, n_{E}-qubit environment via a random time-independent Hamiltonian, with Eq. (5.16) statistically predicting under which conditions memory effects can be neglected. In Fig. 5.4 we take such a system for a single qubit and demand a bound $\mathsf{B} \leq 0.01$ on the probability $\mathbb{P}_{\mathsf{t}_\varepsilon}[\mathcal{N}_\blacklozenge \geq 0.1]$ for a $k = 2$ timestep process; with this, we plot the scaling of the non-commuting gate depth $\mathscr{D}$ required to achieve an $\epsilon = 10^{-12}$ approximate unitary t-design using $\mathcal{W}_\ell$ circuits for different values of $2 \leq \mathsf{t} \leq 10$. While the number of 2-qubit gates is on the order of 10^4, the number of repetitions ℓ is at most 12 for an approximate 10-design and stays mostly constant as the number of environment qubits increases.

This construction naturally accommodates the cartoon example in Fig. 5.1: as long as the two interactions in the example together generate the necessary level of complexity, Markovianization will emerge. This shows, in principle, how simple dynamics described by approximate unitary designs can Markovianize under the right conditions, but, moreover, taking the physical interpretation of a qubit locally interacting through two-qubit diagonal unitaries with a large environment, it also hints at how macroscopic systems can display Markovianization of small subsystem dynamics in circuits requiring just a small gate depth. Furthermore, for macroscopic systems with coarse observables, the same Markovianization behaviour would remain resilient to a much larger number of interventions.

5.4 Conclusions

The results in this Chapter, based on Ref. [3] show that quantum processes with physically motivated interactions can Markovianize, in the sense of becoming Markovian with very high probability in suitable limits. Beyond foundational considerations, our results have direct consequences for the study of open systems using standard tools, such as master equations and dynamical maps. As we saw in Chapter 2, these can be seen as a family of one-step process tensors, where in the presence of initial SE correlations a minimum of two steps must be considered [130, 224]. Specifically, for the case of $k \leq 2$, our result can be used to estimate the time scale, using gate depth as a proxy, on which an approximate unitary design's open dynamics can be described (with high probability) with a truncated memory kernel [8, 202, 225], or even a Markovian master equation.

Conversely, for larger k, our results would have implications for approximations made in computing higher order correlation functions, such as the quantum regression theorem [226]. These higher order approximations are independent of those at the level of dynamical maps, which can, e.g., be divisible, even when the process is non-Markovian [79]. This is reflected in the loosening behaviour of the bound in Eq. (5.16) as the number of timesteps increases, which can be interpreted as a growing potential for temporal correlations to become relevant when more information about the process is accessible.

This breadth of applicability is in contrast with the results presented in Chapter 4 from Ref. [2], which has two main drawbacks: first, as stated above, Haar random interactions do not exist in nature and hence the relevance of the result is limited. Second, the rate of Markovianization is far too strong. Almost all processes, sampled according to the Haar measure, will look highly Markovian even for a large k. This, unlike our current result, misses almost all interesting physical dynamical processes.

While the behaviour of the large deviations bound in Theorem 5.2.1 is polynomial, rather than exponential, thus not exhibiting concentration per-se, we have nevertheless exemplified how, with modestly large environments and relatively simple interactions, almost Markovian processes can come about with high probability. Physical macroscopic environments will be far larger than the scale shown in Figs. 5.2 and 5.4.

Finally, despite the fundamental relevance of our result, it is well known that typicality arguments can have limited reach. For instance, the exotic Hamiltonians, introduced in Ref. [227], which lead to *strange* relaxation, may not Markovianize even though the SE process is highly complex with a large E. There is also still significant scope for further addressing physical aspects, such as the question of whether, and how, a time-independent Hamiltonian can give rise to an approximate unitary designs [205], the real-world time scales of Markovianization, or the potential role of different approaches to pseudo-randomness such as that in Ref. [228], where it is shown that driven quantum systems can converge rapidly to the uniform distribution.

Furthermore, a renewed wave of interest in thermalization has come along with the so-called eigenstate thermalization hypothesis, which is a stronger and seemingly more fundamental condition on thermalization [19, 23–25, 27, 229], and we would thus expect a deep connection in the sense of ETH between Markovianization, thermalization, and/or dynamical equilibration to be forthcoming. In any case, it is clear that many physical systems Markovianize at some scale, and it only remains to discover how.

PART III

Envoi

CHAPTER 6

Conclusions

Within this thesis, we have investigated the concepts of equilibration and typicality in the context of quantum stochastic processes, appealing to the motivation of advancing an understanding of the emergence of Markovianity purely from quantum mechanical laws. We have seen that questions regarding this emergence can be phrased in analogous terms to those related to the emergence of statistical mechanics from quantum mechanics.

Specifically, with the original results of Part II of this thesis we have:

Chapter 3 Extended the notion of equilibration on average on expectation values of observables to a multitime scenario, establishing sufficient conditions for the multitime statistics due to a sequence of observations on a quantum process to equilibrate.

Chapter 4 Formally shown—without resorting to the Born-Markov assumption or making any approximations—that quantum processes undergone within a finite large dimensional environment concentrate around Markovian ones.

Chapter 5 Obtained a large deviation bound for quantum processes described by approximate unitary designs, showing that quantum processes can Markovianize in the sense that they can converge to a class of Markovian processes whenever they are undergone within a large environment and under complex enough dynamics.

These results were made possible due to the developments in the topics of equilibration and typicality, which relate to the emergence of statistical mechanics solely from quantum mechanics, and the process tensor framework for quantum stochastic processes, both presented in Part I.

In **Chapter 1** we focused on the results on equilibration on average by Ref. [54] and typicality by Ref. [65], which we consider were pivotal regarding the emergence of statistical mechanics purely from quantum mechanics. Topics related to equilibration and typicality are currently highly active ones in fronts such as equilibration time-scales [18, 59] and the eigenstate thermalization hypothesis [19, 24–27].

In **Chapter 2** we presented the mathematical framework of quantum maps and some of their representations, which allowed us to approach the problem of temporal correlations in

open quantum systems. We saw that this is resolved by the process tensor framework for quantum stochastic processes, introduced in Ref. [151] (and equivalent frameworks with different perspectives in Ref. [156–166]), which establishes an unambiguous condition for Markovianity capturing all multitime phenomena and memory effects [152]. Not unexpectedly, the topic of quantum stochastic processes and the process tensor framework are highly active research areas, with investigations in machine learning [230], resource theories [231], open dynamics simulations [232] or the causal structure of quantum processes [167], to name but a few examples. Our work is rather a contribution to the area of quantum processes in the spirit of the approaches of equilibration and thermalization of Chapter 1.

We now discuss the outlook for future research related to this thesis and some possible ways to approach it.

6.1 Outlook

The results within this thesis make a significant step towards the understanding of the emergence of Markovianity in quantum processes. There are, nevertheless, several ways in which this understanding can be further advanced, as well as applied.

In the case of multitime equilibration, two main issues stand out. The first relates to the role that memory has, both in the dynamical process itself and within the external operations. While we obtained sufficient conditions under which quantum processes will equilibrate, the interplay with memory effects, both through the environment and the ancillary space in the interventions is as yet not entirely clear, e.g. under which circumstances finite-temporal resolution equilibration can occur without the dynamics being Markovian, or if the temporal correlations among interventions through the ancillary space can display a departure from equilibration within a finite-time.

The relation between multitime equilibration and Markovianity is somewhat akin to the link between equilibration and thermalization, in the sense that stronger conditions and a deeper characterization of the equilibrium process might be needed in order to determine whether it is generically indeed an almost Markovian process. This could be achieved e.g. by investigating the non-Markovianity of the equilibrium process Ω and the limits that make it exactly Markovian. In the case of the memory on the external operations, while it is expected that having access to a memory space that can keep track of the multitime statistics in a process would necessarily serve to witness a departure from equilibration, attempts to show this proved elusive during this PhD. Possible ways to move forward could be e.g. employing ideas topics such as state transfer [233] or algorithmic cooling [234], where it would find relevant technological applications as well.

A further issue in multitime equilibration will be setting time-scales in which equilibration occurs. This, however, is a longstanding problem even in the standard scenario, so tackling the problem in general quantum processes could be a very ambitious goal.

In the case of Markovianization, there are some open questions that might be possible to

address in a reasonable time-frame. In particular, further constraining Markovianization to general physical and experimentally realizable processes, is a goal that could be achieved in many ways with recent theoretical developments; one such way could be developing the notion of time-independent Hamiltonian unitary designs [205], or another could be looking at driven quantum systems that can quickly increase their complexity [228].

Finally, two topics in the border with the ones discussed in this thesis are the eigenstate thermalization hypothesis (ETH) and Randomized Benchmarking.

The ETH relates to the energy eigenstates of large and chaotic systems behaving as random variables and displaying the corresponding equilibrium statistical properties [23, 24]. The framework in which the ETH is set naturally overlaps with many of the technical concepts employed in this thesis and many questions regarding higher-order correlations could be approached with the process tensor and techniques from random matrix theory employed to study the typicality of Markovian processes.

On the other hand, Randomized Benchmarking refers to the estimation of average error rates in random circuits by implementing random quantum gates (elementary unitary operations) that would amount to identities if there was no error present. The mathematical formalism employed in the derivation for the results on the typicality of Markovian processes and Markovianization is mostly the same as the one in randomized benchmarking: the idea is to estimate the so-called fidelity, i.e. a measure of noisiness, of a class of sets of noisy random quantum gates entering a quantum process with respect to the corresponding noise-free ones by looking at the average over the noisy gates. Whenever the noise is Markovian, so that there is both no time-dependence and no dependence on the particular choice of gates, randomized benchmarking has been widely studied [208, 235–239]. However, randomized benchmarking for non-Markovian noise, i.e. due to memory effects, is still an open problem and the process tensor is the natural framework to approach it. This is a problem whose solution could represent a significant contribution with high impact in the field of quantum computing.

We remain confident that the work presented in this thesis will serve as a stepping stone to tackle these problems, as well as an inspiration to approach others, old or new, naturally connected to quantum stochastic processes and quantum information science.

Appendices

APPENDIX A

Notation

An effort has been made to maintain consistency in the notation throughout this thesis. Here we present the standard notation that is employed throughout the text.

Many symbols are only distinguished by their style, e.g. $\mathbb{P}$, $\mathcal{P}$, $\mathscr{P}$, $\wp$ and P all stand for different concepts, so we hope this stylistic distinction, together with the relevant context, is enough to easily tell them apart.

Keep in mind that we only work with finite dimensional quantum systems. Most abbreviated terminology, such as referring to density matrices as quantum states, Hilbert space as space, or stochastic process as process, is sufficiently clear by context and we do not reproduce it here. Similarly, some simplifications such as positivity of a quantum states meaning positive semi-definiteness, or removing the subindex for the number of time-steps in a process tensor is usually done when sufficiently clear by context.

We have also committed some abuse of notation when relevant for simplicity's sake, e.g. omitting identity operators when an operator acts only on a subpart of the whole system in question. When this is done it has been duly pointed out.

We have also been somewhat lax when it comes to the usual mathematical structure of introducing Definitions, Remarks or Notation in a bullet point style and opted for a more narrative structure; we have nevertheless used these when we have considered relevant, in particular with the original theorems from this PhD which are followed by their proof.

Finally, we made an effort to employ a consistent color scheme on the different figures, mainly using red to denote dynamical objects, blue/teal on experimental operations and/or maps and purple on process tensors.

General

$\mathbb{R}, \mathbb{R}^+, \mathbb{R}_0^+$	Set of real, positive real and non-negative real numbers
$\mathbb{C}$	Set of complex numbers
$\mathbb{U}(d)$	Unitary group of dimension d
$\mathscr{H}$	Hilbert space

$(\cdot)^*, (\cdot)^{\mathrm{T}}, (\cdot)^\dagger$	Conjugate, Transpose and Conjugate transpose
$\|\psi\rangle, \langle\psi\|$	Vector on a Hilbert space (ket) and Adjoint vector on a Hilbert space (bra)
$\langle\,\cdot\,\rangle$	Expectation value
$[\cdot,\cdot]$	Commutator, $[a,b] = ab - ba$
ρ	Density matrix on a Hilbert space
tr	Trace
$\mathsf{A}, \mathsf{B}, \ldots$	Systems are labelled with sans-serif capital letters
in, out	Inputs and outputs
in'	An apostrophe distinguishes an input that has been acted on with a map
$\mathscr{H}_\mathsf{A}$	Hilbert space associated to system A
ρ_A	Density matrix on Hilbert space A
tr_A	Partial trace over system A
$\tilde{\Psi}$	Unnormalized maximally entangled state $\tilde{\Psi} = \sum \|ii\rangle\langle jj\|$
Ψ	Normalized to unity maximally entangled state, $\Psi := \tilde{\Psi}/d$
d_A	Standard for the dimension of the Hilbert space of a system, $\dim(\mathscr{H}_\mathsf{A})$
$\mathscr{T}_k$	Set of time-steps $\{t_0, t_1, \ldots, t_{k-1}\}$
$k:0$	Set of ordered time-steps $t_0 < t_1 < \ldots < t_{k-1}$

Operations

$\mathscr{B}(\mathscr{H})$	Space of bounded linear operators acting on the Hilbert space $\mathscr{H}$
$\$(\mathscr{H})$	Space of density matrices acting on the Hilbert space $\mathscr{H}$
$A, B, \ldots$	Bounded linear operators are denoted with standard capital letters
M_i	`POVM` element
H	Hamiltonian operator
Π	Projection operator
U	Unitary operator
$\mathbb{1}$	Identity operator
SWAP	Swap operator, $\text{SWAP} := \sum \|ij\rangle\langle ji\|$
$\mathfrak{S}_{\alpha\beta}$	Defined as $\mathfrak{S}_{\alpha\beta} := \mathbb{1}_\mathsf{E} \otimes \|\alpha\rangle\langle\beta\|$ with $\|\alpha\rangle, \|\beta\rangle$ being S basis vectors
Φ	Generic map between bounded linear operator spaces
$\mathcal{U}$	Unitary map
$\mathcal{I}$	Identity map
$\mathscr{X}_t$, $\mathscr{X}_{j:i}$	Dynamical map on time parameter t, dynamical map from time-step i to j
$\mathcal{S}_i$	Swap map between system S and an ancilla at time-step t_i
$\Xi^{(n)}$	n-fold twirl map
Υ_Φ	Choi state of a map Φ
$\mathcal{A}_0, \mathcal{A}_1, \ldots$	`CPTNI` maps at time-steps $t_0, t_1, \ldots$
$\mathcal{M}$	Superchannel
$\mathcal{T}_{k:0}$	k-step process tensor; the subindex $k:0$ is dropped when clear by context

$\Upsilon_{k:0}$	Choi state of a process tensor
$\mathcal{J}$	Instrument - a collection of CP maps that sums to a CPTP map
$\Lambda_{k:0}$	Choi state of a sequence of CP operations (tester when over instruments)

Equilibration

$\overline{(\cdot)}$	Infinite time-average
$\overline{(\cdot)}^{T}$	Uniform time-average over an interval of width T
$\overline{(\cdot)}^{\mathscr{P}_T}$	Time-average over a distribution with PDF $\mathscr{P}_T$ with fuzziness T
ω	Equilibrium state
ϱ, ϖ	Initial SΓ state and SΓ equilibrium state
d_{eff}	Effective dimension (a.k.a. participation ration)
E_n	Energy levels (Hamiltonian eigenvalues)
P_n	Projector onto the eigenspace of energy E_n
$\mathcal{E}_{nm}$	Energy difference $\mathcal{E}_{nm} \propto E_n - E_m$
$\mathfrak{D}$	Number of distinct energies, $\mathfrak{D} = \lvert\text{spec}(H)\rvert$
$N(\varepsilon)$	Max number of energy gaps in any interval of size $\epsilon > 0$
$\mathcal{D}$	Dephasing map with respect to a Hamiltonian
$\mathcal{G}_T$	Partial dephasing map within a time interval of width T

Norms / distances

$\lVert\cdot\rVert_p$	Schatten p-norm
$\lVert\cdot\rVert_1$	(Schatten) 1-norm or trace norm
$\lVert\cdot\rVert$	Operator norm (max singular value)
$D(\cdot,\cdot)$	Trace distance
$S(\cdot)$	von-Neumann entropy
$S(\cdot\,\lVert\,\cdot)$	Relative entropy

Probability / Statistics

μ	Probability measure
$\mathbb{P}$	Probability distribution
$\mathfrak{P}_{j:i}$	Propagator stochastic matrix from time t_i to time t_j
$\mathbb{E}$	Expectation
h	Haar measure
$\mathfrak{h}$	Haar measure in random interaction ($U_i \neq U_j$)

t	Unitary t-design
t_ϵ	ϵ-approximate unitary t-design
$\mathscr{L}$	Lipschitz constant
$\mathscr{P}$	Probability density function
$\mathfrak{S}_n$	Symmetric group on n-elements
$\wp$	Permutation operator

APPENDIX B

Haar distributed process tensors

Here we present the derivation of the expressions presented in Chaper 4 for the expectation of quantum processes Υ and the expectation of the purity, or noisiness, of quantum processes $\mathrm{tr}\left(\Upsilon^2\right)$, according to the Haar measure.

B.1 Average processes - Constant interaction

We are considering here Υ to be unnormalized (or more precisely, normalized to $\mathrm{tr}(\Upsilon) = d_{\mathsf{S}}^{k+1}$), as presented first in the main text in Eq. (4.11). Also, here Υ stands for a k-step process $\Upsilon_{k:0}$ as done in the main text as well.

By definition,

$$\mathbb{E}_{\mathsf{h}}(\Upsilon) = \mathrm{tr}_{\mathsf{E}} \sum_{\alpha,\ldots,\gamma=1}^{d_{\mathsf{S}}} \int_{\mathbb{U}(d_{\mathsf{SE}})} U\,\mathfrak{S}_{\alpha_k\beta_k}\cdots U\,\mathfrak{S}_{\alpha_1\beta_1}\,U\,\rho\,U^\dagger\,\mathfrak{S}^\dagger_{\gamma_1\delta_1}\,U^\dagger\cdots\mathfrak{S}^\dagger_{\gamma_k\delta_k}\,U^\dagger\,d\mu_{\mathsf{h}}(U) \\ \otimes |\beta_1\alpha_1\ldots\beta_k\alpha_k\rangle\langle\delta_1\gamma_1\ldots\delta_k\gamma_k|, \tag{B.1}$$

where $\mathfrak{S}_{\alpha\beta} := \mathbb{1}_{\mathsf{E}} \otimes |\alpha\rangle\langle\beta|$. We now decompose the unitaries as $U = \sum U_{ab}|a\rangle\langle b|$ and $U^\dagger = \sum U^*_{a'b'}|b'\rangle\langle a'|$ –notice that the a and b labels refer to the whole SE space–, introducing a resolution of the identity on E within the $\mathfrak{S}$ operators as $\mathfrak{S}_{ab} = \sum_e |ea\rangle\langle eb|$, and then evaluating the $k+1$ moments of $\mathbb{U}(d_{\mathsf{SE}})$ by means of the moments of the unitary group in Eq. (4.6),

$$\begin{aligned}
&\int_{\mathbb{U}(d_{\mathsf{SE}})} U\,\mathfrak{S}_{\alpha_k\beta_k}\cdots U\,\mathfrak{S}_{\alpha_1\beta_1}\,U\rho U^\dagger\,\mathfrak{S}^\dagger_{\gamma_1\delta_1}\,U^\dagger\cdots\mathfrak{S}^\dagger_{\gamma_k\delta_k}\,U^\dagger\,d\mu_{\mathsf{h}}(U) \\
&= \sum \int_{\mathbb{U}(d_{\mathsf{SE}})} U_{i_0j_0}\cdots U_{i_kj_k}U^*_{i'_0j'_0}\cdots U^*_{i'_kj'_k}\,d\mu_{\mathsf{h}}(U)\,|i_k\rangle\langle j_0|\rho|j'_0\rangle\langle i'_k| \\
&\qquad \prod_{\ell=1}^{k}\delta_{j_\ell(e\alpha)_\ell}\delta_{(e\beta)_\ell i_{\ell-1}}\delta_{j'_\ell(e'\gamma)_\ell}\delta_{(e'\delta)_\ell i'_{\ell-1}} \\
&= \sum \sum_{\sigma,\tau\in\mathfrak{S}_{k+1}} \langle j_0|\rho|j'_0\rangle\delta_{i_0 i'_{\sigma(0)}}\delta_{j_0 j'_{\tau(0)}}\prod_{\ell=1}^{k}\delta_{i_\ell i'_{\sigma(\ell)}}\delta_{j_\ell j'_{\tau(\ell)}}\delta_{j_\ell(e\alpha)_\ell}\delta_{(e\beta)_\ell i_{\ell-1}}\delta_{j'_\ell(e'\gamma)_\ell}\delta_{(e'\delta)_\ell i'_{\ell-1}} \\
&\qquad \mathsf{Wg}(\tau\sigma^{-1})\,|i_k\rangle\langle i'_k|
\end{aligned}$$

$$= \sum_{\sigma,\tau\in\mathfrak{G}_{k+1}} \langle j'_{\tau(0)}|\rho|j'_0\rangle \delta_{i_k i'_{\sigma(k)}} \prod_{\ell=1}^{k} \delta_{(e\beta)_\ell i'_{\sigma(\ell-1)}} \delta_{(e'\delta)_\ell i'_{\ell-1}} \delta_{(e\alpha)_\ell j'_{\tau(\ell)}} \delta_{(e'\gamma)_\ell j'_\ell} \mathsf{Wg}(\tau\sigma^{-1}) \, |i_k\rangle\langle i'_k|, \tag{B.2}$$

where here $\mathfrak{G}_{k+1}$ denotes the symmetric group on $\{0,1,\ldots,k\}$, Wg is the Weingarten function with implicit dimensional argument d_{SE}, and which taking $i \to \epsilon\varsigma$ and $j \to \epsilon'\varsigma'$ to recover each E and S part explicitly, turns into

$$\begin{aligned}
&\int_{\mathbb{U}(d_{\mathsf{SE}})} U\,\mathfrak{S}_{\alpha_k\beta_k}\cdots U\,\mathfrak{S}_{\alpha_1\beta_1}\,U\,\rho\,U^\dagger\,\mathfrak{S}^\dagger_{\gamma_1\delta_1}\,U^\dagger\cdots\mathfrak{S}^\dagger_{\gamma_k\delta_k}\,U^\dagger\,d\mu(U)\\
&= \sum_{\sigma,\tau\in\mathfrak{G}_{k+1}} \langle \epsilon'_{\tau(0)}\varsigma'_{\tau(0)}|\rho|\epsilon'_0\varsigma'_0\rangle \prod_{\ell=1}^{k} \delta_{\epsilon_{\sigma(\ell-1)}\epsilon'_{\tau(\ell)}} \delta_{\epsilon_{\ell-1}\epsilon'_\ell} \delta_{\beta_\ell\varsigma_{\sigma(\ell-1)}} \delta_{\delta_\ell\varsigma_{\ell-1}} \delta_{\alpha_\ell\varsigma'_{\tau(\ell)}} \delta_{\gamma_\ell\varsigma'_\ell} \mathsf{Wg}(\tau\sigma^{-1})\\
&\qquad |\epsilon_{\sigma(k)}\varsigma_{\sigma(k)}\rangle\langle\epsilon_k\varsigma_k|,
\end{aligned} \tag{B.3}$$

and thus

$$\begin{aligned}
\mathbb{E}_{\mathsf{h}}(\Upsilon) = \sum_{\sigma,\tau\in\mathfrak{G}_{k+1}} &\langle \epsilon'_{\tau(0)}\varsigma'_{\tau(0)}|\rho|\epsilon'_0\varsigma'_0\rangle \, \mathsf{Wg}(\tau\sigma^{-1}) \, \delta_{\epsilon_{\sigma(k)}\epsilon_k} \prod_{\ell=1}^{k} \delta_{\epsilon_{\sigma(\ell-1)}\epsilon'_{\tau(\ell)}} \delta_{\epsilon_{\ell-1}\epsilon'_\ell}\\
&|\varsigma_{\sigma(k)}\varsigma_{\sigma(0)}\varsigma'_{\tau(1)}\cdots\varsigma_{\sigma(k-1)}\varsigma'_{\tau(k)}\rangle\langle\varsigma_k\varsigma_0\varsigma'_1\cdots\varsigma_{k-1}\varsigma'_k|,
\end{aligned} \tag{B.4}$$

as stated by Eq.(4.11).

B.1.1 Superchannel case

For the superchannel case, $k = 1$, we have $\mathfrak{G}_2 = \{(0,1),(0)(1)\}$ where the elements are permutations stated in *cycle* notation, representing the assignments $(0,1) : 0 \to 1 \to 0$ and $(0)(1) = \mathbf{1}^2 = 0 \to 0; 1 \to 1$, then (we write $\alpha,\beta,\gamma,\delta$ for $\varsigma_0,\varsigma_1,\varsigma'_0,\varsigma'_1$, respectively to ease the notation)

$$\begin{aligned}
\mathbb{E}_{\mathsf{h}}(\Upsilon_{1:0}) = \frac{\sum}{d_{\mathsf{S}}}\Big\{ &\left[\langle\delta|\rho_{\mathsf{S}}|\gamma\rangle|\alpha\beta\gamma\rangle\langle\beta\alpha\delta| + d_{\mathsf{E}}^2|\beta\alpha\delta\rangle\langle\beta\alpha\delta|\right]\mathsf{Wg}[\mathbf{1}^2]\\
&+ d_{\mathsf{E}}\left[\langle\delta|\rho_{\mathsf{S}}|\gamma\rangle|\beta\alpha\gamma\rangle\langle\beta\alpha\delta| + |\alpha\beta\delta\rangle\langle\beta\alpha\delta|\right]\mathsf{Wg}[(0,1)]\Big\},
\end{aligned} \tag{B.5}$$

so that with $\mathsf{Wg}[\mathbf{1}^2,d] = \frac{1}{d^2-1}$ and $\mathsf{Wg}[(0,1),d] = \frac{-1}{d(d^2-1)}$ [187] we get

$$\mathbb{E}_{\mathsf{h}}(\Upsilon_{1:0}) = \frac{1}{d_{\mathsf{E}}^2 d_{\mathsf{S}}^2 - 1}\left[\frac{d_{\mathsf{E}}^2}{d_{\mathsf{S}}}\mathbb{1}_{\mathsf{SAB}} + \frac{\mathrm{SWAP}}{d_{\mathsf{S}}}\otimes\rho_{\mathsf{S}}^{\mathrm{T}} - \frac{\mathrm{SWAP}}{d_{\mathsf{S}}}\otimes\frac{\mathbb{1}_{\mathsf{B}}}{d_{\mathsf{S}}} - \frac{\mathbb{1}_{\mathsf{SA}}}{d_{\mathsf{S}}^2}\otimes\rho_{\mathsf{S}}^{\mathrm{T}}\right], \tag{B.6}$$

where $\mathrm{SWAP} = \sum_{i,j}|ij\rangle\langle ji|$ is the usual swap operator, and hence for the corresponding purity one may verify that

$$\mathrm{tr}[\mathbb{E}_{\mathsf{h}}(\Upsilon_{1:0})^2] = \frac{2}{(d_{\mathsf{E}}^2 d_{\mathsf{S}}^2 - 1)^2}\left[\frac{1}{d_{\mathsf{S}}^3} + \mathrm{tr}(\rho_S^2)\frac{d_{\mathsf{S}}^2 - d_{\mathsf{S}} - 1}{2d_{\mathsf{S}}^2} - \frac{d_{\mathsf{E}}^2}{d_{\mathsf{S}}} + \frac{d_{\mathsf{E}}^4 d_{\mathsf{S}}}{2}\right]. \tag{B.7}$$

B.2 Average purity - Random interaction

Here Υ is normalized to unit trace, in accordance with the main text. Let $\Theta = \rho \otimes \Psi^{\otimes k}$, where we assume ρ is pure, $\mathrm{tr}\,\rho = 1$, and $\mathsf{U}_{k:0} := U_k \mathcal{S}_k \cdots U_1 \mathcal{S}_1 U_0$, we first (following the approach in section 2 of [240]) write the trace as

$$\begin{aligned}
\mathrm{tr}\left(\Upsilon^2\right) &= \mathrm{tr}\left[\left(\mathrm{tr}_{\mathsf{E}}(\mathsf{U}_{k:0}\Theta\mathcal{U}_{k:0}^\dagger)\right)^2\right] \\
&= \mathrm{tr}\left[\mathrm{tr}_{\mathsf{E}}\left((\mathbb{1}_E \otimes \Upsilon)\,\mathsf{U}_{k:0}\Theta\mathcal{U}_{k:0}^\dagger\right)\right] \\
&= \mathrm{tr}\left[\Gamma_{\mathsf{E}}(\mathsf{U}_{k:0}\Theta\mathcal{U}_{k:0}^\dagger)\,\mathsf{U}_{k:0}\Theta\mathcal{U}_{k:0}^\dagger\right] \\
&= \mathrm{tr}\left[\Theta\,\mathsf{U}_{k:0}^\dagger\Gamma_{\mathsf{E}}(\mathsf{U}_{k:0}\Theta\mathcal{U}_{k:0}^\dagger)\,\mathsf{U}_{k:0}\right],
\end{aligned} \tag{B.8}$$

where $\Gamma_{\mathsf{E}}(\cdot) \equiv \mathbb{1}_E \otimes \mathrm{tr}_{\mathsf{E}}(\cdot)$ is a CP map and can be expressed as an operator-sum, i.e. for any $\mathcal{X}$ acting on $ESA_1B_1\cdots A_kB_k$ we have $\Gamma_{\mathsf{E}}(\mathcal{X}) = \sum_{i,j=1}^{d_{\mathsf{E}}}(K_{ij}\otimes\mathbb{1})\mathcal{X}(K_{ij}\otimes\mathbb{1})^\dagger$ where the identities are on the ancillary system and $K_{ij} = |i\rangle\langle j| \otimes \mathbb{1}_S$ are the (ES system) Kraus operators with $\{|i\rangle\}_{i=1}^{d_{\mathsf{E}}}$ a given basis for E. And so we need to compute the integrals

$$\Omega_{k:0} := \int_{\mathbb{U}(d_{\mathsf{SE}})} \mathsf{U}_{k:0}^\dagger\Gamma_{\mathsf{E}}(\mathsf{U}_{k:0}\Theta\mathcal{U}_{k:0}^\dagger)\,\mathsf{U}_{k:0}\,d\mu_{\mathsf{h}}(U_0)\cdots d\mu_{\mathsf{h}}(U_k), \tag{B.9}$$

and in fact these are really in the SE part only, which has the form

$$\begin{aligned}
\omega_{k:0}^{(\rho)} := \int_{\mathbb{U}(d_{\mathsf{SE}})} & U_0^\dagger\,\mathfrak{S}_{x_1\alpha_1}\,U_1^\dagger\cdots\mathfrak{S}_{x_k\alpha_k}U_k^\dagger K_{\tilde{i}\tilde{j}}U_k\,\mathfrak{S}_{\beta_k x_k} \\
&\cdots U_1\,\mathfrak{S}_{\beta_1 x_1}\,U_0\rho U_0^\dagger\,\mathfrak{S}_{y_1\lambda_1}\,U_1^\dagger\cdots\mathfrak{S}_{y_k\lambda_k}\,U_k^\dagger K_{\tilde{i}\tilde{j}}^\dagger U_k\,\mathfrak{S}_{\theta_k y_k} \\
&\cdots U_1\,\mathfrak{S}_{\theta_1 y_1}\,U_0\,d\mu_{\mathsf{h}}(U_0)\,d\mu_{\mathsf{h}}(U_1)\cdots d\mu_{\mathsf{h}}(U)_k,
\end{aligned} \tag{B.10}$$

with summation over repeated indices x, y and i, j on the Kraus operators implied, and the ancillary part takes the form

$$\Omega_{k:0}^{(\mathsf{AB})} := (|\alpha_1\rangle\langle\beta_1| \otimes \mathbb{1}_{\mathsf{B}_1} \otimes \cdots \otimes |\alpha_k\rangle\langle\beta_k| \otimes \mathbb{1}_{\mathsf{B}_k})\Psi^{\otimes k}(|\lambda_1\rangle\langle\theta_1| \otimes \mathbb{1}_{\mathsf{B}_1} \otimes \cdots \otimes |\lambda_k\rangle\langle\theta_k| \otimes \mathbb{1}_{\mathsf{B}_k}). \tag{B.11}$$

We may then evaluate each of the integrals using

$$\int_{\mathbb{U}(d)} U^\dagger A U X U^\dagger B U\,d\mu_{\mathsf{h}}(U) = \frac{d\,\mathrm{tr}(AB) - \mathrm{tr}(A)\,\mathrm{tr}(B)}{d(d^2-1)}\,\mathrm{tr}(X)\mathbb{1} + \frac{d\,\mathrm{tr}(A)\,\mathrm{tr}(B) - \mathrm{tr}(AB)}{d(d^2-1)}X, \tag{B.12}$$

which can be seen to follow from the second moment of the unitary group (an explcit derivation can be seen e.g. in Ref. [194]), here with $d = d_{\mathsf{SE}}$, which from now on we employ, then we have

$$\begin{aligned}
&\omega_{k:0}^{(\rho)} = \frac{1}{d(d^2-1)} \\
&\Big\{\delta_{x_1y_1}\int_{\mathbb{U}(d)}\mathrm{tr}\Big[\mathfrak{S}_{\theta_1\alpha_1}\,U_1^\dagger\cdots\mathfrak{S}_{x_k\alpha_k}U_k^\dagger K_{ij}U_k\,\mathfrak{S}_{\beta_k x_k}
\end{aligned}$$

$$\cdots U_1\,\mathfrak{S}_{\beta_1\lambda_1}\,U_1^\dagger\cdots\mathfrak{S}_{y_k\lambda_k}\,U_k^\dagger K_{ij}^\dagger U_k\,\mathfrak{S}_{\theta_k y_k}\cdots U_1\Big](d\mathbb{1}-\rho)\,d\mu_{\mathsf{h}}(U_1)\cdots d\mu_{\mathsf{h}}(U_k)$$
$$+\,\delta_{x_1x_1}\delta_{y_1y_1}\int \mathrm{tr}\Big[\mathfrak{S}_{\beta_1\alpha_1}\,U_1^\dagger\cdots\mathfrak{S}_{x_k\alpha_k}U_k^\dagger K_{ij}U_k\,\mathfrak{S}_{\beta_k x_k}\cdots U_1\Big]$$
$$\mathrm{tr}\Big[\mathfrak{S}_{\theta_1\lambda_1}\,U_1^\dagger\cdots\mathfrak{S}_{y_k\lambda_k}\,U_k^\dagger K_{ij}^\dagger U_k\,\mathfrak{S}_{\theta_k y_k}\cdots U_1\Big](d\rho-\mathbb{1})\;d\mu_{\mathsf{h}}(U_1)\cdots d\mu_{\mathsf{h}}(U_k)\Big\}$$
$$=\frac{1}{d(d^2-1)}\Big[d_{\mathsf{S}}\langle e_1\alpha_1|\omega_{k:1}^{(\mathfrak{S}_{\beta_1\lambda_1})}|e_1\theta_1\rangle(d\mathbb{1}-\rho)+d_{\mathsf{S}}^2\langle e_1\alpha_1|\omega_{k:1}^{(|e_1\beta_1\rangle\langle e_1'\lambda_1|)}|e_1'\theta_1\rangle(d\rho-\mathbb{1})\Big],\quad\text{(B.13)}$$

again with sum implied over e_i's and Greek indices. Now, let us notice that

$$\omega_{k:i-1}^{(X)}=\frac{1}{d(d^2-1)}\Big[d_{\mathsf{S}}\langle e_i\alpha_i|\omega_{k:i}^{(\mathfrak{S}_{\beta_i\lambda_i})}|e_i\theta_i\rangle(d\,\mathrm{tr}(X)\mathbb{1}-X)$$
$$+\,d_{\mathsf{S}}^2\langle e_i\alpha_i|\omega_{k:i}^{(|e_i\beta_i\rangle\langle e_i'\lambda_i|)}|e_i'\theta_i\rangle(dX-\mathrm{tr}(X)\mathbb{1})\Big],\quad\text{(B.14)}$$

for $k>i-1\geq 0$. Then we get

$$\omega_{k:0}^{(\rho)}=\frac{1}{d^2(d^2-1)^2}$$
$$\Big\{d_{\mathsf{S}}\langle e_2\alpha_2|\omega_{k:2}^{(\mathfrak{S}_{\beta_2\lambda_2})}|e_2\theta_2\rangle\Big[d_{\mathsf{S}}(dd_{\mathsf{E}}^2\delta_{\alpha_1\theta_1}\delta_{\beta_1\lambda_1}-d_{\mathsf{E}}\delta_{\alpha_1\beta_1}\delta_{\lambda_1\theta_1})(d\mathbb{1}-\rho)$$
$$+\,d_{\mathsf{S}}^2(dd_{\mathsf{E}}\delta_{\alpha_1\theta_1}\delta_{\beta_1\lambda_1}-d_{\mathsf{E}}^2\delta_{\alpha_1\beta_1}\delta_{\lambda_1\theta_1})(d\rho-\mathbb{1})\Big]$$
$$+\,d_{\mathsf{S}}^2\langle e_2\alpha_2|\omega_{k:2}^{(|e_2\beta_2\rangle\langle e_2'\lambda_2|)}|e_2'\theta_2\rangle\Big[d_{\mathsf{S}}(dd_{\mathsf{E}}\delta_{\alpha_1\beta_1}\delta_{\lambda_1\theta_1}-d_{\mathsf{E}}^2\delta_{\alpha_1\theta_1}\delta_{\beta_1\lambda_1})(d\mathbb{1}-\rho)$$
$$+\,d_{\mathsf{S}}^2(dd_{\mathsf{E}}^2\delta_{\alpha_1\beta_1}\delta_{\lambda_1\theta_1}-d_{\mathsf{E}}\delta_{\alpha_1\theta_1}\delta_{\beta_1\lambda_1})(d\rho-\mathbb{1})\Big]\Big\}.\quad\text{(B.15)}$$

Before carrying on, let us notice that the case $\omega_{k:k}$ is special because here one evaluates the terms on the Kraus operators $\mathrm{tr}(K_{ij}K_{ij}^\dagger)$ and $\mathrm{tr}(K_{ij})\,\mathrm{tr}(K_{ij}^\dagger)$ summed over their indices, which leaves

$$\langle e_k\alpha_k|\omega_{k:k}^{(\mathfrak{S}_{\lambda_k\beta_k})}|e_k\theta_k\rangle=\frac{1}{d(d^2-1)}\Big[dd_{\mathsf{E}}^2(dd_{\mathsf{E}}\delta_{\alpha_k\theta_k}\delta_{\beta_k\lambda_k}-\delta_{\alpha_k\beta_k}\delta_{\lambda_k\theta_k})\quad\text{(B.16)}$$
$$+\,d^2d_{\mathsf{E}}(d_{\mathsf{S}}\delta_{\alpha_k\beta_k}\delta_{\lambda_k\theta_k}-\delta_{\alpha_k\theta_k}\delta_{\beta_k\lambda_k})\Big],\quad\text{(B.17)}$$

and

$$\langle e_k\alpha_k|\omega_{k:k}^{(|e_k\beta_k\rangle\langle e_k'\lambda_k|)}|e_k'\theta_k\rangle=\frac{1}{d(d^2-1)}\Big[dd_{\mathsf{E}}^3(d_{\mathsf{S}}\delta_{\alpha_k\theta_k}\delta_{\beta_k\lambda_k}-\delta_{\alpha_k\beta_k}\delta_{\lambda_k\theta_k})$$
$$+\,d^2(dd_{\mathsf{E}}\delta_{\alpha_k\beta_k}\delta_{\lambda_k\theta_k}-\delta_{\alpha_k\theta_k}\delta_{\beta_k\lambda_k})\Big].\quad\text{(B.18)}$$

Notice that we can get rid of the Kronecker deltas easily in the full $\Omega_{k:0}$ when summing over Greek indices, as $\delta_{\alpha\theta}\delta_{\beta\lambda}$ give rise to maximally mixed states in the ancillas while terms $\delta_{\alpha\beta}\delta_{\lambda\theta}$ give rise to identities, and there are only deltas of this kind. Thus we may simply assign $\delta_{\alpha\theta}\delta_{\beta\lambda}\to 1/d_{\mathsf{S}}$ and $\delta_{\alpha\beta}\delta_{\lambda\theta}\to 1$ when plugging the corresponding expressions in $\Omega_{k:0}$.[1]

[1] In particular this implies that the terms $d_{\mathsf{S}}\delta_{\alpha\theta}\delta_{\beta\lambda}-\delta_{\alpha\beta}\delta_{\lambda\theta}$ and $dd_{\mathsf{E}}\delta_{\alpha\beta}\delta_{\lambda\theta}-d_{\mathsf{E}}^2\delta_{\alpha\theta}\delta_{\beta\lambda}$ won't contribute to the average purity.

Furthermore, a direct consequence of this is that $\Omega_{k:0}$ will be a linear combination of 2^{k+1} tensor products between $\mathbb{1}$, ρ and Ψ, implying that (as ρ is pure, Ψ is idempotent with trace one and the trace of an outer product is the product of traces) the average purity $\mathbb{E}_{\mathfrak{h}}[\mathrm{tr}(\Upsilon^2)]$ will be a sum of the scalar terms in $\omega^{(\rho)}_{k:0}$ together with an extra overall factor of $d-1$. Let us denote by

$$A = d_{\mathsf{E}}(d_{\mathsf{E}}^2-1), \quad B = d_{\mathsf{E}}^2 - d_{\mathsf{E}}^2 = 0, \quad C = d_{\mathsf{E}}^2(d_{\mathsf{S}} - 1/d_{\mathsf{S}}), \quad D = d_{\mathsf{E}}(dd_{\mathsf{E}} - 1/d_{\mathsf{S}}), \tag{B.19}$$

the terms that appear in expression (B.15) after having taken $\delta_{\alpha\theta}\delta_{\beta\lambda} \to 1/d_{\mathsf{S}}$ and $\delta_{\alpha\beta}\delta_{\lambda\theta} \to 1$. Doing similarly for the $\omega_{k:k}$ case,

$$\mathcal{A} = dd_{\mathsf{E}}^2(d_{\mathsf{E}}^2 + d_{\mathsf{S}}^2 - 2), \quad \mathcal{B} = dd_{\mathsf{E}}(d^2-1), \tag{B.20}$$

we find that[2]

$$\mathbb{E}_{\mathfrak{h}}[\mathrm{tr}(\Upsilon^2)] = \frac{d_{\mathsf{S}}^k(d-1)}{[d(d^2-1)]^{k+1}}\left[\mathcal{A}\,A^{k-1} + d_{\mathsf{S}}\mathcal{B}\left(CA^{k-2} + C\sum_{i=1}^{k-2} d_{\mathsf{S}}^i D^i A^{k-i-2} + d_{\mathsf{S}}^{k-1}D^{k-1}\right)\right], \tag{B.21}$$

where the series has to be expanded before this can be evaluated; by means of the geometric series, $\sum_{i=1}^{n} x^i = \frac{x(x^n-1)}{x-1}$, i.e. with

$$A^{k-2}\sum_{i=1}^{k-2}\left(\frac{d_{\mathsf{S}}D}{A}\right)^i = Dd_{\mathsf{S}}\left(\frac{d_{\mathsf{S}}^{k-2}D^{k-2} - A^{k-2}}{d_{\mathsf{S}}D - A}\right), \tag{B.22}$$

then plugging in the dimensional values (with $d = d_{\mathsf{SE}}$) of all the constants we can simplify the purity to

$$\mathbb{E}_{\mathfrak{h}}[\mathrm{tr}(\Upsilon^2)] = \frac{(d_{\mathsf{SE}}+1)\left(d_{\mathsf{SE}}^2-1\right)^k + \left(d_{\mathsf{E}}^2-1\right)^{k+1}}{d_{\mathsf{E}}(d_{\mathsf{SE}}+1)\left(d_{\mathsf{SE}}^2-1\right)^k}, \tag{B.23}$$

which simplifies to the expression in Eq. (4.27).

B.3 Average purity - Constant interaction

As in the previous Appendix B.2, here Υ is normalized to unit trace. We may write the integral in the ES part, as in Eq.(B.10) now with same unitary (throughout we take sums over repeated indices),

$$\omega^{(\rho)}_{k:0} := \int_{\mathbb{U}(d_{\mathsf{SE}})} U^\dagger\,\mathfrak{S}_{x_1\alpha_1}\,U^\dagger\cdots\mathfrak{S}_{x_k\alpha_k}U_k^\dagger K_{\tilde{i}\tilde{j}} U\,\mathfrak{S}_{\beta_k x_k}\cdots U\,\mathfrak{S}_{\beta_1 x_1}\,U\rho U^\dagger\,\mathfrak{S}_{y_1\lambda_1}\,U^\dagger$$
$$\cdots\mathfrak{S}_{y_k\lambda_k}\,U^\dagger K_{\tilde{i}\tilde{j}}^\dagger U\,\mathfrak{S}_{\theta_k y_k}\cdots U\,\mathfrak{S}_{\theta_1 y_1}\,U\,d\mu_{\mathfrak{h}}(U), \tag{B.24}$$

We decompose the unitaries in the whole SE space as $U = \sum U_{ab}|a\rangle\langle b|$ and $U^\dagger = \sum U^*_{a'b'}|b'\rangle\langle a'|$ (i.e. the a, b labels here refer to the whole SE space) and we enumerate the labels of the ones to

[2] A way to deduce this is to sub-label each factor $A, B, \cdots$ by the i index of the term $\langle e_i\alpha_i|\omega^{X_i}|e_i^{(\prime)}\theta_i\rangle$ that they come from and then substituting recursively in Eq. (B.15) as if one were already evaluating the whole purity. The vanishing of B simplifies this process a great deal.

the left of ρ in Eq. (B.24) from 0 to k (priming, $'$, the adjoint ones) and the remaining from $k+1$ to $2k+1$, we also do this in increasing order for the adjoint unitaries and decreasing order for the original unitaries so that they match the order of the $\mathfrak{S}$ operators, i.e. the unitary components will originally appear as $U^*_{i'_0 j'_0}\cdots U^*_{i'_k j'_k} U_{i_k j_k}\cdots U_{i_0 j_0} U^*_{i'_{k+1} j'_{k+1}}\cdots U^*_{i'_{2k+1} j'_{2k+1}} U_{i'_{2k+1} j'_{2k+1}}\cdots U_{i'_{k+1} j'_{k+1}}$, and we then rearrange them to the form of Eq. (4.6) just keeping track of the correct order in the operator part, for which we introduce resolutions of identity on E into each $\mathfrak{S}$ operator as $\mathfrak{S}_{ab} = \sum |ea\rangle\langle eb|$ labeled by e for the ones to the left of ρ in Eq. (B.24), priming the ones between adjoint unitaries, and by ϵ the corresponding ones to the right, also priming the ones between unitaries.

This then leads to

$$\begin{aligned}
\omega^{(\rho)}_{k:0} &= \int_{\mathbb{U}(d_{\mathsf{SE}})} U_{i_0 j_0}\cdots U_{i_{2k+1} j_{2k+1}} U^*_{i'_0 j'_0}\cdots U^*_{i'_{2k+1} j'_{2k+1}}\, d\mu_{\mathsf{h}}(U)\, \langle i'_k|K_{\tilde{\imath}\tilde{\jmath}}|i_k\rangle\langle j_0|\rho|j'_{k+1}\rangle\langle i'_{2k+1}|K^\dagger_{\tilde{\imath}\tilde{\jmath}}|i_{2k+1}\rangle \\
&\qquad \prod_{\ell=1}^{k} \delta_{i_{\ell-1}(ex)_\ell}\delta_{j_\ell(e\beta)_\ell}\delta_{i_{\ell+k}(\epsilon y)_\ell}\delta_{j_{\ell+k+1}(\epsilon\theta)_\ell}\delta_{i'_{\ell-1}(e'x)_\ell}\delta_{j'_\ell(e'\alpha)_\ell}\delta_{i'_{\ell+k}(\epsilon' y)_\ell}\delta_{j'_{\ell+k+1}(\epsilon'\lambda)_\ell}|j'_0\rangle\langle j_{k+1}| \\
&= \langle i'_k|K_{\tilde{\imath}\tilde{\jmath}}|i_k\rangle\langle j_0|\rho|j'_{k+1}\rangle\langle i'_{2k+1}|K^\dagger_{\tilde{\imath}\tilde{\jmath}}|i_{2k+1}\rangle \sum_{\sigma,\tau\in\mathfrak{S}_{2k+2}} \mathsf{Wg}(\tau\sigma^{-1}) \prod_{n=0}^{2k+1} \delta_{i_n i'_{\sigma(n)}}\delta_{j_n j'_{\tau(n)}}\, |j'_0\rangle\langle j_{k+1}| \\
&\qquad \prod_{\ell=1}^{k} \delta_{i_{\ell-1}(ex)_\ell}\delta_{j_\ell(e\beta)_\ell}\delta_{i_{\ell+k}(\epsilon y)_\ell}\delta_{j_{\ell+k+1}(\epsilon\theta)_\ell}\delta_{i'_{\ell-1}(e'x)_\ell}\delta_{j'_\ell(e'\alpha)_\ell}\delta_{i'_{\ell+k}(\epsilon' y)_\ell}\delta_{j'_{\ell+k+1}(\epsilon'\lambda)_\ell} \\
&= \sum_{\sigma,\tau\in\mathfrak{S}_{2k+2}} \mathsf{Wg}(\tau\sigma^{-1})\langle i'_k|K_{\tilde{\imath}\tilde{\jmath}}|i'_{\sigma(k)}\rangle\langle j'_{\tau(0)}|\rho|j'_{k+1}\rangle\langle i'_{2k+1}|K^\dagger_{\tilde{\imath}\tilde{\jmath}}|i'_{\sigma(2k+1)}\rangle|j'_0\rangle\langle j'_{\tau(k+1)}| \\
&\qquad \prod_{\ell=1}^{k} \delta_{i'_{\sigma(\ell-1)}(ex)_\ell}\delta_{j'_{\tau(\ell)}(e\beta)_\ell}\delta_{i'_{\sigma(\ell+k)}(\epsilon y)_\ell}\delta_{j'_{\tau(\ell+k+1)}(\epsilon\theta)_\ell}\delta_{i'_{\ell-1}(e'x)_\ell}\delta_{j'_\ell(e'\alpha)_\ell}\delta_{i'_{\ell+k}(\epsilon' y)_\ell}\delta_{j'_{\ell+k+1}(\epsilon'\lambda)_\ell}.
\end{aligned} \tag{B.25}$$

We now replace $i' \to \varepsilon\varsigma$ and $j' \to \varepsilon'\varsigma'$ in order to split the E and S parts explicitly, leaving

$$\begin{aligned}
&\omega^{(\rho)}_{k:0} \\
&= \sum_{\sigma,\tau\in\mathfrak{S}_{2k+2}} \mathsf{Wg}(\tau\sigma^{-1})\langle\varepsilon_k\varsigma_k|K_{\tilde{\imath}\tilde{\jmath}}|\varepsilon_{\sigma(k)}\varsigma_{\sigma(k)}\rangle\langle\varepsilon'_{\tau(0)}\varsigma'_{\tau(0)}|\rho|\varepsilon'_{k+1}\varsigma'_{k+1}\rangle\langle\varepsilon_{2k+1}\varsigma_{2k+1}|K^\dagger_{\tilde{\imath}\tilde{\jmath}}|\varepsilon_{\sigma(2k+1)}\varsigma_{\sigma(2k+1)}\rangle \\
&\qquad \prod_{\ell=1}^{k} \delta_{(\varepsilon\varsigma)_{\sigma(\ell-1)}(ex)_\ell}\delta_{(\varepsilon'\varsigma')_{\tau(\ell)}(e\beta)_\ell}\delta_{(\varepsilon\varsigma)_{\sigma(\ell+k)}(\epsilon y)_\ell}\delta_{(\varepsilon'\varsigma')_{\tau(\ell+k+1)}(\epsilon\theta)_\ell}\delta_{(\varepsilon\varsigma)_{\ell-1}(e'x)_\ell} \\
&\qquad\qquad \delta_{(\varepsilon'\varsigma')_\ell(e'\alpha)_\ell}\delta_{(\varepsilon\varsigma)_{\ell+k}(\epsilon' y)_\ell}\delta_{(\varepsilon'\varsigma')_{\ell+k+1}(\epsilon'\lambda)_\ell}|\varepsilon'_0\varsigma'_0\rangle\langle\varepsilon'_{\tau(k+1)}\varsigma'_{\tau(k+1)}| \\
&= \sum_{\sigma,\tau\in\mathfrak{S}_{2k+2}} \mathsf{Wg}(\tau\sigma^{-1})\langle\varepsilon'_{\tau(0)}\varsigma'_{\tau(0)}|\rho|\varepsilon'_{k+1}\varsigma'_{k+1}\rangle|\varepsilon'_0\varsigma'_0\rangle\langle\varepsilon'_{\tau(k+1)}\varsigma'_{\tau(k+1)}|\delta_{\varepsilon_k\varepsilon_{\sigma(2k+1)}}\delta_{\varepsilon_{\sigma(k)}\varepsilon_{2k+1}} \\
&\qquad \prod_{\ell=1}^{k} \delta_{\varepsilon_{\sigma(\ell-1)}\varepsilon'_{\tau(\ell)}}\delta_{\varepsilon_{\sigma(\ell+k)}\varepsilon'_{\tau(\ell+k+1)}}\delta_{\varepsilon_{\ell-1}\varepsilon'_\ell}\delta_{\varepsilon_{\ell+k}\varepsilon'_{\ell+k+1}}\delta_{\varsigma'_\ell\alpha_\ell} \\
&\qquad\qquad \delta_{\varsigma'_{\tau(\ell)}\beta_\ell}\delta_{\varsigma'_{\tau(\ell+k+1)}\theta_\ell}\delta_{\varsigma'_{\ell+k+1}\lambda_\ell} \prod_{n=0}^{2k+1} \delta_{\varsigma_{\sigma(n)}\varsigma_n}.
\end{aligned} \tag{B.26}$$

The ancillary part is simply $\Omega_{k:0}^{(\mathsf{AB})} = d_{\mathsf{S}}^{-k}|\alpha_1\beta_1\ldots\alpha_k\beta_k\rangle\langle\theta_1\lambda_1\ldots\theta_k\lambda_k|$, so we get for the analogue of Eq. (B.9),

$$\begin{aligned}\Omega_{k:0} &= \omega_{k:0}^{(\rho)} \otimes \Omega_{k:0}^{(AB)} \\ &= \frac{1}{d_{\mathsf{S}}^k} \sum_{\sigma,\tau\in\mathfrak{G}_{2k+2}} \mathsf{Wg}(\tau\sigma^{-1})\delta_{\varepsilon_k\varepsilon_{\sigma(2k+1)}}\delta_{\varepsilon_{\sigma(k)}\varepsilon_{2k+1}} \\ &\quad |\varepsilon'_0\varsigma'_0\varsigma'_1\varsigma'_{\tau(1)}\cdots\varsigma'_k\varsigma'_{\tau(k)}\rangle\langle\varepsilon'_{\tau(k+1)}\varsigma'_{\tau(k+1)}\varsigma'_{\tau(k+2)}\varsigma'_{k+2}\cdots\varsigma'_{\tau(2k+1)}\varsigma'_{2k+1}| \\ &\quad \langle\varepsilon'_{\tau(0)}\varsigma'_{\tau(0)}|\rho|\varepsilon'_{k+1}\varsigma'_{k+1}\rangle \prod_{n=0}^{2k+1}\delta_{\varsigma_{\sigma(n)}\varsigma_n}\prod_{\ell=1}^{k}\delta_{\varepsilon_{\sigma(\ell-1)}\varepsilon'_{\tau(\ell)}}\delta_{\varepsilon_{\sigma(\ell+k)}\varepsilon'_{\tau(\ell+k+1)}}\delta_{\varepsilon_{\ell-1}\varepsilon'_\ell}\delta_{\varepsilon_{\ell+k}\varepsilon'_{\ell+k+1}},\end{aligned} \tag{B.27}$$

and thus

$$\begin{aligned}\mathbb{E}_{\mathsf{h}}[\mathrm{tr}(\Upsilon^2)] &= \mathrm{tr}[(\rho\otimes\Psi^{\otimes k})\,\Omega_{k:0}] \\ &= \frac{1}{d_{\mathsf{S}}^{2k}} \sum_{\sigma,\tau\in\mathfrak{G}_{2k+2}} \mathsf{Wg}(\tau\sigma^{-1})\langle\varepsilon'_{\tau(0)}\varsigma'_{\tau(0)}|\rho|\varepsilon'_{k+1}\varsigma'_{k+1}\rangle\langle\varepsilon'_{\tau(k+1)}\varsigma'_{\tau(k+1)}|\rho|\varepsilon'_0\varsigma'_0\rangle\delta_{\varepsilon_k\varepsilon_{\sigma(2k+1)}}\delta_{\varepsilon_{2k+1}\varepsilon_{\sigma(k)}} \\ &\quad \prod_{\substack{\ell=1\\ \ell\neq k+1}}^{2k+1}\delta_{\varepsilon_{\ell-1}\varepsilon'_\ell}\delta_{\varepsilon_{\sigma(\ell-1)}\varepsilon'_{\tau(\ell)}}\delta_{\varsigma'_\ell\varsigma'_{\tau(\ell)}}\prod_{n=0}^{2k+1}\delta_{\varsigma_{\sigma(n)}\varsigma_n},\end{aligned} \tag{B.28}$$

as stated in Eq. (4.28), where we make the definitions

$$\rho_{\tau(0);k+1}\rho_{\tau(k+1);0} = \langle\varepsilon'_{\tau(0)}\varsigma'_{\tau(0)}|\rho|\varepsilon'_{k+1}\varsigma'_{k+1}\rangle\langle\varepsilon'_{\tau(k+1)}\varsigma'_{\tau(k+1)}|\rho|\varepsilon'_0\varsigma'_0\rangle \tag{B.29}$$

$$\tilde{\Delta}_{k,\sigma,\tau}^{(d_{\mathsf{E}})} = \delta_{e_k e_{\sigma(2k+1)}}\delta_{e_{2k+1}e_{\sigma(k)}}\prod_{\substack{\ell=1\\ \ell\neq k+1}}^{2k+1}\delta_{e_{\ell-1}e'_\ell}\delta_{e_{\sigma(\ell-1)}e'_{\tau(\ell)}}, \quad \tilde{\tilde{\Delta}}_{k,\sigma,\tau}^{(d_{\mathsf{S}})} = \prod_{\substack{\ell=1\\ \ell\neq k+1}}^{2k+1}\delta_{s'_\ell s'_{\tau(\ell)}}\prod_{n=0}^{2k+1}\delta_{s_n s_{\sigma(n)}}$$

$$\Delta_{k,\sigma,\tau}^{(d_{\mathsf{E}},d_{\mathsf{S}})} = \tilde{\Delta}_{k,\sigma,\tau}^{(d_{\mathsf{E}})}\tilde{\tilde{\Delta}}_{k,\sigma,\tau}^{(d_{\mathsf{S}})} \tag{B.30}$$

B.3.1 Small subsystem limit

In the $d_{\mathsf{E}}\to\infty$ limit the only term that remains is the one with

$$\sigma = \tau = (0,k+1)(1,k+2)\cdots(k,2k+1), \tag{B.31}$$

as expressed in cycle notation, which simply means $\sigma(0) = k+1 = \tau(0)$, $\sigma(k+1) = 0 = \tau(k+1)$ for the first $(0,k+1)$, and similarly for the rest of assignments. This then leads to the correct limit, as

$$\tilde{\tilde{\Delta}}_{k,\sigma,\tau}^{(d_{\mathsf{S}})} \to \sum_{\substack{\varsigma_i^{(\prime)}=1\\ \varsigma'\neq\{\varsigma'_0,\varsigma'_{k+1}\}}}^{d_{\mathsf{S}}}\prod_{\substack{\ell=1\\ \ell\neq k+1}}^{2k+1}\delta_{\varsigma'_\ell\varsigma'_{\tau(\ell)}}\prod_{n=0}^{2k+1}\delta_{\varsigma_{\sigma(n)}\varsigma_n} = {d_{\mathsf{S}}}^{2k+1}, \tag{B.32}$$

$$\tilde{\Delta}_{k,\sigma,\tau}^{(d_{\mathsf{E}})} \to \sum_{\substack{\varepsilon_i^{(\prime)}=1\\ \varepsilon'\neq\{\varepsilon'_0,\varepsilon'_{k+1}\}}}^{d_{\mathsf{E}}}\delta_{\varepsilon_k\varepsilon_{\sigma(2k+1)}}\delta_{\varepsilon_{2k+1}\varepsilon_{\sigma(k)}}\prod_{\substack{\ell=1\\ \ell\neq k+1}}^{2k+1}\delta_{\varepsilon_{\ell-1}\varepsilon'_\ell}\delta_{\varepsilon_{\sigma(\ell-1)}\varepsilon'_{\tau(\ell)}} = {d_{\mathsf{E}}}^{2k+2}, \tag{B.33}$$

while keeping $\sigma\tau^{-1} = \mathbf{1}^{2k+2}$, i.e. with the argument in the Wg function an identity (or fixed point), which generates the least denominator powers because $\#\mathbf{1}^n = \#[(1)(2)\cdots(n)] = n$, i.e. the identity generates the most number of cycles. When considering other permutations for σ and τ we see that the d_{E} powers in the numerator can only decrease while those in the denominator can only increase (as the identity is the permutation that generates the most cycles) so indeed in the large subsystem limit this is the only pair of permutations that survive; the terms in ρ yield traces when summed over basis vectors and thus give factors of one,

$$\sum_{\varepsilon'_{k+1},\varsigma'_{k+1}} \langle \varepsilon'_{k+1}\varsigma'_{k+1}|\rho|\varepsilon'_{k+1}\varsigma'_{k+1}\rangle = \sum_{\varepsilon'_0,\varsigma'_0} \langle \varepsilon'_0\varsigma'_0|\rho|\varepsilon'_0\varsigma'_0\rangle = \operatorname{tr}\rho = 1. \tag{B.34}$$

This leads then to the conclusion

$$\mathbb{E}_{\mathsf{h}}[\operatorname{tr}(\Upsilon^2)] \sim \frac{d_{\mathsf{S}}^{2k+1} d_{\mathsf{E}}^{2k+2}}{d_{\mathsf{S}}^{2k}} \mathsf{Wg}(\mathbf{1}^{2k+2}) = d_{\mathsf{S}} d_{\mathsf{E}}^{2k+2} \frac{1}{(d_{\mathsf{E}} d_{\mathsf{S}})^{2k+2}} = \frac{1}{d_{\mathsf{S}}^{2k+1}}, \quad \text{when } d_{\mathsf{E}} \to \infty. \tag{B.35}$$

B.3.2 Long time limit

For the case when $k \to \infty$, the Wg function behaves dominantly as in the small subsystem limit and the only pair of permutations that identify all dimension powers in the numerator without dependence of k are identities $\sigma = \tau = \mathbf{1}^{2k+2}$. In this case

$$\sum_{\substack{\varsigma_i^{(\prime)}=1 \\ \varsigma' \neq \{\varsigma'_0,\varsigma'_{k+1}\}}}^{d_{\mathsf{S}}} \prod_{\substack{\ell=1 \\ \ell\neq k+1}}^{2k+1} \delta_{\varsigma'_\ell \varsigma'_\ell} \prod_{n=0}^{2k+1} \delta_{\varsigma_n \varsigma_n} = {d_{\mathsf{S}}}^{4k+2}, \tag{B.36}$$

$$\sum_{\substack{\varepsilon_i^{(\prime)}=1 \\ \varepsilon' \neq \{\varepsilon'_0,\varepsilon'_{k+1}\}}}^{d_{\mathsf{E}}} \delta_{\varepsilon_k \varepsilon_{2k+1}} \delta_{\varepsilon_{2k+1}\varepsilon_k} \prod_{\substack{\ell=1 \\ \ell\neq k+1}}^{2k+1} \delta_{\varepsilon_{\ell-1}\varepsilon'_\ell} \delta_{\varepsilon_{\ell-1}\varepsilon'_\ell} = {d_{\mathsf{E}}}^{2k+1}, \tag{B.37}$$

giving

$$\mathbb{E}_{\mathsf{h}}[\operatorname{tr}(\Upsilon^2)] \sim \frac{d_{\mathsf{S}}^{4k+2} d_{\mathsf{E}}^{2k+1}}{d_{\mathsf{S}}^{2k}} \operatorname{tr}(\rho^2) \mathsf{Wg}(\mathbf{1}^{2k+2}) = \frac{d_{\mathsf{S}}^{2k+2} d_{\mathsf{E}}^{2k+1}}{(d_{\mathsf{E}} d_{\mathsf{S}})^{2k+2}} = \frac{1}{d_{\mathsf{E}}}, \quad \text{when } k \to \infty, \tag{B.38}$$

as the fiducial state ρ is pure, $\operatorname{tr}(\rho^2) = 1$, by assumption.

APPENDIX C

Sampling from the unitary group

Here we present detail for how the numerical sampling and therefore the plot in Fig. 4.3 of Chapter 4 was obtained. We follow Ref. [66], which contains a complete discussion and reasoning behind the algorithm to numerically sample unitaries from the Haar measure.

Specifically, we computed the bound on the non-Markovianity $\mathcal{N}_1$ given by the trace distance D between a process Υ sampled at random and the maximally noisy process $\mathbb{1}/d_{\mathsf{S}}^{2k+1}$, i.e. $D\left(\Upsilon, \mathbb{1}/d_{\mathsf{S}}^{2k+1}\right) \geq \mathcal{N}_1$. This amounts to sampling unitaries $U_i \sim \mu_{\mathsf{h}}$ of dimension $d_{\mathsf{ES}} \times d_{\mathsf{ES}}$, with which the full process is defined, and then computing its trace distance with respect to the maximally mixed state of dimension d_{S}^{2k+1}. This is repeated a given number of times n in order to obtain an arithmetic average, E_n, which we compare with the upper-bound $\mathcal{B}_k \geq \mathbb{E}_{\mathfrak{h}}[\mathcal{N}_1]$. These must satisfy $\mathrm{E}_n \leq \mathcal{B}_k$, given that $\mathcal{B}_k$ is in turn an upper bound on the uniform (Haar) average trace distance between a process and the maximally noisy process.

The algorithm in Ref. [66] to obtain a uniformly distributed $d \times d$ unitary matrix is as follows, with each step explained below:

i. Generate a $d \times d$ complex matrix Z with independent identically distributed (i.i.d.) standard normal complex random variables.
ii. Perform a QR decomposition of Z.
iii. Define a diagonal matrix $\Lambda = \mathrm{diag}\left(\frac{r_1}{|r_1|}, \cdots, \frac{r_d}{|r_d|}\right)$, where r_i are the diagonal elements of R.
iv. The matrix $U = Q\Lambda$ is distributed according to the Haar measure.

The first step is straightforward, where here "*standard normal complex*" refers to a complex random variable with real and imaginary parts being independent normally distributed random variables with mean zero and variance $1/2$. This means the components Z_{ij} are normally distributed with a PDF given by $\mathscr{P}(Z_{ij}) = \frac{1}{\pi}\exp\left(-|Z_{ij}|^2\right)$. Since these components are statistically independent, the joint distribution for Z is simply the product of distributions so that $\mathscr{P}(Z) = \frac{1}{\pi^{d^2}}\exp\left[-\mathrm{tr}(ZZ^\dagger)\right]$. This distribution describes the so-called *Ginibre ensemble*, and the variable Z induces a probability measure $d\mu_{\mathsf{G}}$ on $\mathrm{GL}(\mathbb{C}, d)$, the set of $d \times d$ invertible matrices with ordinary matrix multiplication. Importantly, this measure on the Ginibre ensemble is left and right invariant under unitary transformations, $d\mu_{\mathsf{G}}(UZ) = d\mu_{\mathsf{G}}(ZV) = d\mu_{\mathsf{G}}(Z)$ for any fixed $U, V \in \mathbb{U}(d)$.

For step two, any matrix $Z \in \mathrm{GL}(\mathbb{C}, d)$ can be decomposed as $Z = QR$ with $Q \in \mathbb{U}(d)$ and R an upper-triangular and invertible matrix [48]. This is called the QR decomposition of Z. This means that, because the matrix R is invertible, we can write $Q = ZR^{-1}$ and we get a unitary random matrix. The QR decomposition is a standard routine in most symbolic softwares such as Matlab or Mathematica or in packages like NumPy for Python.

The last step is related to the non-uniqueness of the QR decomposition. As discussed in Ref. [66], the unitary Q is not quite Haar-distributed because of this non-uniqueness. That is, for any diagonal unitary matrix $\Lambda \in \mathbb{U}(d)$, we have $QR = (Q\Lambda)(\Lambda^\dagger R) = Q'R'$ with Q' unitary and R' upper-triangular. This is remedied with $\Lambda = \mathrm{diag}\left(\frac{r_1}{|r_1|}, \cdots, \frac{r_d}{|r_d|}\right)$ because it forces R to have positive diagonal entries.

Thus, for a k-step random interaction process, i.e. with $U_i \neq U_j$, in the i^{th} run, we generate $k+1$ Haar-distributed $d_{\mathsf{SE}} \times d_{\mathsf{SE}}$ unitary matrices, $\left\{U_0^{(i)}, U_1^{(i)}, \ldots, U_k^{(i)}\right\}$, according to the algorithm above and compute the arithmetic average

$$\mathrm{E}_n = \frac{1}{n}\sum_{i=1}^{n} D\left(\Upsilon^{(i)}, \mathbb{1}/d_{\mathsf{S}}^{2k+1}\right), \tag{C.1}$$

for a given number n of trials, where $\Upsilon^{(i)} = \mathrm{tr}_{\mathsf{E}}\left[\mathcal{U}_k^{(i)}\mathcal{S}_k \cdots \mathcal{U}_1^{(i)}\mathcal{S}_1\mathcal{U}_0^{(i)}(\rho \otimes \Psi)\right]$ with $\mathcal{U}_\ell^{(i)}(\cdot) = U_\ell^{(i)}(\cdot)U_\ell^{(i)\,\dagger}$, is the random process in the i^{th} trial for any fiducial pure state ρ.

Finally, the greatest problem to be overcome is the large-dimensional nature of the Choi state, which prior to partial tracing of the environment, is a $d_{\mathsf{E}} d_{\mathsf{S}}^{2k+1}$ dimensional square matrix. While we directly computed Eq. (C.1), depending on the purpose of the calculation,it might be worth it to try and render this problem efficient through other computational or numerical techniques.

Bibliography

1. Figueroa-Romero, P., Pollock, F. A. & Modi, K. "*Equilibration on average in quantum processes with finite temporal resolution*", arXiv:1905.08469 [quant–ph] (2019).
2. Figueroa-Romero, P., Modi, K. & Pollock, F. A. "*Almost Markovian processes from closed dynamics*". *Quantum* vol. 3, 136 (2019).
3. Figueroa-Romero, P., Pollock, F. A. & Modi, K. "*Markovianization by design*", arXiv:2004.07620 [quant–ph] (2020).
4. Von Neumann. "*Mathematical Foundations of Quantum Mechanics*". English translation by Robert T. Beyer (Princeton University Press, 1955).
5. Peres, A. "*Quantum Theory: Concepts and Methods*" (Springer Netherlands, 2006).
6. Nielsen, M., Nielsen, M., Chuang, I. & Chuang, I. "*Quantum Computation and Quantum Information*" (Cambridge University Press, 2000).
7. Bengtsson, I. & Życzkowski, K. "*Geometry of Quantum States: An Introduction to Quantum Entanglement*" (Cambridge University Press, 2006).
8. Breuer, H. & Petruccione, F. "*The Theory of Open Quantum Systems*" (Oxford University Press, 2002).
9. Gleason, A. M. "*Measures on the Closed Subspaces of a Hilbert Space*" (ed Hooker, C. A.) 123–133 (Springer Netherlands, Dordrecht, 1975).
10. Busch, P. "*Quantum States and Generalized Observables: A Simple Proof of Gleason's Theorem*". *Phys. Rev. Lett.* vol. 91, 120403 (12 2003).
11. Watrous, J. "*The Theory of Quantum Information*" (Cambridge University Press, 2018).
12. Einstein, A., Podolsky, B. & Rosen, N. "*Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?*" *Phys. Rev.* vol. 47, 777–780 (10 1935).
13. Schrödinger, E. "*Discussion of Probability Relations between Separated Systems*". *Math. Proc. Cambridge* vol. 31, 555–563 (1935).
14. Goold, J., Huber, M., Riera, A., del Rio, L. & Skrzypczyk, P. "*The role of quantum information in thermodynamics—a topical review*". *J. Phys. A-Math. Theor.* vol. 49, 143001 (2016).

15. Schrödinger, E. "*Energieaustausch nach der Wellenmechanik*". *Ann. Phys.* vol. 388, 956–968 (1927).

16. Schrödinger, E. "*Collected Papers on Wave Mechanics*" (AMS Chelsea Pub., 2003).

17. Von Neumann, J. "*Proof of the ergodic theorem and the H-theorem in quantum mechanics*". *Eur. Phys. J. H.* vol. 35, 201–237 (2010).

18. Gogolin, C. & Eisert, J. "*Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems*". *Rep. Prog. Phys.* vol. 79, 056001 (2016).

19. D'Alessio, L., Kafri, Y., Polkovnikov, A. & Rigol, M. "*From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics*". *Adv. Phys.* vol. 65, 239–362 (2016).

20. Bocchieri, P. & Loinger, A. "*Ergodic Foundation of Quantum Statistical Mechanics*". *Phys. Rev.* vol. 114, 948–951 (4 1959).

21. Peres, A. "*Ergodicity and mixing in quantum theory. I*". *Phys. Rev. A* vol. 30, 504–508 (1 1984).

22. Jensen, R. V. & Shankar, R. "*Statistical Behavior in Deterministic Quantum Systems with Few Degrees of Freedom*". *Phys. Rev. Lett.* vol. 54, 1879–1882 (17 1985).

23. Deutsch, J. M. "*Quantum statistical mechanics in a closed system*". *Phys. Rev. A* vol. 43, 2046–2049 (4 1991).

24. Srednicki, M. "*Chaos and quantum thermalization*". *Phys. Rev. E* vol. 50, 888–901 (2 1994).

25. Srednicki, M. "*The approach to thermal equilibrium in quantized chaotic systems*". *J. Phys. A-Math. Gen.* vol. 32, 1163–1175 (1999).

26. Rigol, M. & Srednicki, M. "*Alternatives to Eigenstate Thermalization*". *Phys. Rev. Lett.* vol. 108, 110601 (11 2012).

27. Murthy, C. & Srednicki, M. "*Bounds on Chaos from the Eigenstate Thermalization Hypothesis*". *Phys. Rev. Lett.* vol. 123, 230606 (23 2019).

28. Boltzmann, L. "*Further Studies on the Thermal Equilibrium of Gas Molecules*", 262–349 (2003).

29. Kittel, C., Charles Kittel, H., Charles, K., Kroemer, H. & Herbert, K. "*Thermal Physics*" (W. H. Freeman, 1980).

30. Bocchieri, P. & Loinger, A. "*Quantum Recurrence Theorem*". *Phys. Rev.* vol. 107, 337–338 (2 1957).

31. Peres, A. "*Recurrence Phenomena in Quantum Dynamics*". *Phys. Rev. Lett.* vol. 49, 1118–1118 (15 1982).

32. Wallace, D. "*Recurrence Theorems: a unified account*", arXiv:1306.3925 [quant–ph] (2013).

33. Bhattacharyya, K. & Mukherjee, D. "*On estimates of the quantum recurrence time*". *J. Chem. Phys.* vol. 84, 3212–3214 (1986).

34. Gimeno, V. & Sotoca, J. M. "*Upper bounds for the Poincaré recurrence time in quantum mixed states*". *J. Phys. A-Math. Theor.* vol. 50, 185302 (2017).

35. Gogolin, C., Müller, M. P. & Eisert, J. "*Absence of Thermalization in Nonintegrable Systems*". *Phys. Rev. Lett.* vol. 106, 040401 (4 2011).

36. Besicovitch, A. S. "*On Generalized Almost Periodic Functions*". *P. Lond. Math. Soc.* vol. s2-25, 495–512 (1926).

37. Bhatia, R. "*Matrix Analysis*" (Springer New York, 2013).

38. Lindblad, G. "*Completely positive maps and entropy inequalities*". *Commun. Math. Phys.* vol. 40, 147–151 (1975).

39. Gibbs, J. "*Elementary Principles in Statistical Mechanics*" (Dover Publications, 2014).

40. Jaynes, E. T. "*Information Theory and Statistical Mechanics*". *Phys. Rev.* vol. 106, 620–630 (4 1957).

41. Jaynes, E. T. "*Information Theory and Statistical Mechanics. II*". *Phys. Rev.* vol. 108, 171–190 (2 1957).

42. Rigol, M., Dunjko, V., Yurovsky, V. & Olshanii, M. "*Relaxation in a completely integrable many-body quantum system: An ab initio study of the dynamics of the highly excited states of 1D lattice hard-core bosons*". *Phys. Rev. Lett.* vol. 98, 050405 (5 2007).

43. Reimann, P. "*Foundation of Statistical Mechanics under Experimentally Realistic Conditions*". *Phys. Rev. Lett.* vol. 101, 190403 (19 2008).

44. Linden, N., Popescu, S., Short, A. J. & Winter, A. "*Quantum mechanical evolution towards thermal equilibrium*". *Phys. Rev. E* vol. 79, 061103 (6 2009).

45. Short, A. J. "*Equilibration of quantum systems and subsystems*". *New. J. Phys.* vol. 13, 053009 (2011).

46. Neuenhahn, C. & Marquardt, F. "*Thermalization of interacting fermions and delocalization in Fock space*". *Phys. Rev. E* vol. 85, 060101 (6 2012).

47. Calixto, M. & Romera, E. "*Inverse participation ratio and localization in topological insulator phase transitions*". *J. Stat. Mech.-Theory E.* vol. 2015, P06029 (2015).

48. Horn, R. & Johnson, C. "*Matrix Analysis*" (Cambridge University Press, 2012).

49. Reimann, P. & Kastner, M. "*Equilibration of isolated macroscopic quantum systems*". *New. J. Phys.* vol. 14, 043020 (2012).

50. Farrelly, T., Brandão, F. G. S. L. & Cramer, M. "*Thermalization and Return to Equilibrium on Finite Quantum Lattice Systems*". *Phys. Rev. Lett.* vol. 118, 140601 (14 2017).

51. Müller, M. P., Adlam, E., Masanes, L. & Wiebe, N. "*Thermalization and Canonical Typicality in Translation-Invariant Quantum Lattice Systems*". *Commun. Math. Phys.* vol. 340, 499–561 (2015).

52. Brandão, F. G. & Cramer, M. "*Equivalence of Statistical Mechanical Ensembles for Non-Critical Quantum Systems*", arXiv:1502.03263 [quant–ph] (2015).

53. Tasaki, H. "*On the Local Equivalence Between the Canonical and the Microcanonical Ensembles for Quantum Spin Systems*". *J. Stat. Phys.* vol. 172, 905–926 (2018).

54. Short, A. J. & Farrelly, T. C. "*Quantum equilibration in finite time*". *New. J. Phys.* vol. 14, 013063 (2012).

55. Hemmer, P. C., Maximon, L. C. & Wergeland, H. "*Recurrence Time of a Dynamical System*". *Phys. Rev.* vol. 111, 689–694 (3 1958).

56. García-Pintos, L. P., Linden, N., Malabarba, A. S. L., Short, A. J. & Winter, A. "*Equilibration Time Scales of Physically Relevant Observables*". *Phys. Rev. X* vol. 7, 031027 (3 2017).

57. De Oliveira, T. R., Charalambous, C., Jonathan, D., Lewenstein, M. & Riera, A. "*Equilibration time scales in closed many-body quantum systems*". *New. J. Phys.* vol. 20, 033032 (2018).

58. Heveling, R., Knipschild, L. & Gemmer, J. "*Comment on "Equilibration Time Scales of Physically Relevant Observables"*". *Phys. Rev. X* vol. 10, 028001 (2 2020).

59. Wilming, H., de Oliveira, T. R., Short, A. J. & Eisert, J. "*Equilibration Times in Closed Quantum Many-Body Systems*" (eds Binder, F., Correa, L. A., Gogolin, C., Anders, J. & Adesso, G.) 435–455 (Springer International Publishing, Cham, 2018).

60. Yukalov, V. "*Equilibration and thermalization in finite quantum systems*". *Laser Phys. Lett.* vol. 8, 485–507 (2011).

61. Goldstein, S., Lebowitz, J. L., Mastrodonato, C., Tumulka, R. & Zanghì, N. "*Normal typicality and von Neumann's quantum ergodic theorem*". *P. Roy. Soc. A-Math. Phy.* vol. 466, 3203–3224 (2010).

62. Lloyd, S. "*Black Holes, Demons and the loss of Coherence: how complex systems get information, and what they do with it*" (Rockefeller University, 1988).

63. Goldstein, S., Lebowitz, J. L., Tumulka, R. & Zanghì, N. "*Canonical Typicality*". *Phys. Rev. Lett.* vol. 96, 050403 (5 2006).

64. Gemmer, J., Michel, M. & Mahler, G. "*Quantum Thermodynamics: Emergence of thermodynamic behavior within composite quantum systems*" (Springer Berlin Heidelberg, 2009).

65. Popescu, S., Short, A. J. & Winter, A. "*Entanglement and the foundations of statistical mechanics*". *Nat. Phys.* vol. 2, 754–758 (2006).

66. Mezzadri, F. "*How to generate random matrices from the classical compact groups*". *Notices of the AMS* vol. 54, 592–604 (2007).

67. Życzkowski, K. & Kus, M. "*Random unitary matrices*". *J. Phys. A-Math. Theor.* vol. 27, 4235–4245 (1994).

68. Halmos, P. "*Measure Theory*" (Springer New York, 2013).

69. Roberts, D. A. & Yoshida, B. "*Chaos and complexity by design*". *J. High Energy Phys.* vol. 2017, 121 (2017).

70. Ledoux, M. "*The Concentration of Measure Phenomenon*" (American Mathematical Society, 2005).

71. Gläßle, T. "*Almost all pure quantum states are almost maximally entangled*". *Selected topics in Mathematical Physics: Quantum Information Theory* (2013).

72. Page, D. N. "*Average entropy of a subsystem*". *Phys. Rev. Lett.* vol. 71, 1291–1294 (9 1993).

73. Hayden, P., Leung, D. W. & Winter, A. "*Aspects of Generic Entanglement*". *Commun. Math. Phys.* vol. 265, 95–117 (2006).

74. Masanes, L., Roncaglia, A. J. & Acín, A. "*Complexity of energy eigenstates as a mechanism for equilibration*". *Phys. Rev. E* vol. 87, 032137 (3 2013).

75. Modi, K. "*Operational approach to open dynamics and quantifying initial correlations*". *Sci. Rep.* vol. 2, 581 (2012).

76. Accardi, L. "*Nonrelativistic quantum mechanics as a noncommutative Markof process*". *Adv. Math.* vol. 20, 329–366 (1976).

77. Accardi, L., Frigerio, A. & Lewis, J. T. "*Quantum Stochastic Processes*". *Publ. Res. I. Math. Sci.* vol. 18, 97–133 (1982).

78. Breuer, H.-P., Laine, E.-M., Piilo, J. & Vacchini, B. "*Colloquium: Non-Markovian dynamics in open quantum systems*". *Rev. Mod. Phys.* vol. 88, 021002 (2 2016).

79. Milz, S., Kim, M. S., Pollock, F. A. & Modi, K. "*Completely Positive Divisibility Does Not Mean Markovianity*". *Phys. Rev. Lett.* vol. 123, 040401 (4 2019).

80. Jordan, T. F. & Sudarshan, E. C. G. "*Dynamical Mappings of Density Operators in Quantum Mechanics*". *J. Math. Phys.* vol. 2, 772–775 (1961).

81. Sudarshan, E. C. G., Mathews, P. M. & Rau, J. "*Stochastic Dynamics of Quantum-Mechanical Systems*". *Phys. Rev.* vol. 121, 920–924 (3 1961).

82. Kraus, K. "*General state changes in quantum theory*". *Ann. Phys.* vol. 64, 311–335 (1971).

83. Milz, S., Pollock, F. A. & Modi, K. "*An Introduction to Operational Quantum Dynamics*". *Open Syst. Inf. Dyn.* vol. 24, 1740016 (2017).

84. Choi, M.-D. "*Positive Linear Maps on C*-Algebras*". *Canadian J. Math.* vol. 24, 520–529 (1972).

85. Pechukas, P. "*Reduced Dynamics Need Not Be Completely Positive*". *Phys. Rev. Lett.* vol. 73, 1060–1062 (8 1994).

86. Alicki, R. "*Comment on "Reduced Dynamics Need Not Be Completely Positive"*". *Phys. Rev. Lett.* vol. 75, 3020–3020 (16 1995).

87. Pechukas, P. "*Pechukas Replies:*" *Phys. Rev. Lett.* vol. 75, 3021–3021 (16 1995).

88. Wood, C. J. "*Non-completely positive maps: properties and applications*", arXiv:0911.3199 [quant–ph] (2008).

89. Milz, S. C. "*On the Operational Theory of General (Quantum) Stochastic Processes*". 2019.

90. Taranto, P. "*Memory effects in quantum processes*". *Int. J. Quantum. Inf.* vol. 18, 1941002 (2020).

91. Cappellini, V., Sommers, H.-J. & Życzkowski, K. "*Subnormalized states and trace-nonincreasing maps*". *J. Math. Phys.* vol. 48, 052110 (2007).

92. Stinespring, W. F. "*Positive Functions on C*-Algebras*". *P. Am. Math. Soc.* vol. 6, 211–216 (1955).

93. Kraus, K., Böhm, A., Dollard, J. & Wootters, W. "*States, effects, and operations*" (Springer-Verlag, 1983).

94. Coecke, B. & Kissinger, A. "*Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*" (Cambridge University Press, 2017).

95. Wood, C. J., Biamonte, J. D. & Cory, D. G. "*Tensor networks and graphical calculus for open quantum systems*". *Quantum Inf. Comput.* vol. 15, 759–811 (2015).

96. Jamiołkowski, A. "*Linear transformations which preserve trace and positive semidefiniteness of operators*". *Rep. Math. Phys.* vol. 3, 275–278 (1972).

97. Choi, M.-D. "*Completely positive linear maps on complex matrices*". *Linear Algebra Appl.* vol. 10, 285–290 (1975).

98. Kuah, A., Modi, K., Rodríguez-Rosario, C. A. & Sudarshan, E. C. G. "*How state preparation can affect a quantum experiment: Quantum process tomography for open systems*". *Phys. Rev. A* vol. 76, 042113 (4 2007).

99. Preskill, J. "*Lecture Notes for Ph219/CS219: Quantum Information*" (Oct. 2018).

100. Carmichael, H. & de Bruxelles, U. "*An Open Systems Approach to Quantum Optics*". **v. 18** (Springer Berlin Heidelberg, 1993).

101. Blanchard, P., Giulini, D., Joos, E., Kiefer, C. & Stamatescu, I. "*Decoherence: Theoretical, Experimental, and Conceptual Problems: Proceedings of a Workshop Held at Bielefeld Germany, 10–14 November 1998*" (Springer Berlin Heidelberg, 2000).

102. Schlosshauer, M. "*Decoherence: and the Quantum-To-Classical Transition*" (Springer Berlin Heidelberg, 2007).

103. Alicki, R. & Lendi, K. "*Quantum Dynamical Semigroups and Applications*" (Springer Berlin Heidelberg, 2007).

104. Franke, V. A. "*On the general form of the dynamical transformation of density matrices*". *Theor. Math. Phys+.* vol. 27, 406–413 (1976).

105. Gorini, V., Kossakowski, A. & Sudarshan, E. C. G. "*Completely positive dynamical semigroups of N-level systems*". *J. Math. Phys.* vol. 17, 821–825 (1976).

106. Lindblad, G. "*On the generators of quantum dynamical semigroups*". *Commun. Math. Phys.* vol. 48, 119–130 (1976).

107. Chuang, I. L. & Nielsen, M. A. "*Prescription for experimental determination of the dynamics of a quantum black box*". *J. Mod. Optic.* vol. 44, 2455–2467 (1997).

108. Nielsen, M. A., Knill, E. & Laflamme, R. "*Complete quantum teleportation using nuclear magnetic resonance*". *Nature* vol. 396, 52–55 (1998).

109. Childs, A. M., Chuang, I. L. & Leung, D. W. "*Realization of quantum process tomography in NMR*". *Phys. Rev. A* vol. 64, 012314 (1 2001).

110. Mitchell, M. W., Ellenor, C. W., Schneider, S. & Steinberg, A. M. "*Diagnosis, Prescription, and Prognosis of a Bell-State Filter by Quantum Process Tomography*". *Phys. Rev. Lett.* vol. 91, 120402 (12 2003).

111. Altepeter, J. B. *et al.* "*Ancilla-Assisted Quantum Process Tomography*". *Phys. Rev. Lett.* vol. 90, 193601 (19 2003).

112. Weinstein, Y. S. *et al.* "*Quantum process tomography of the quantum Fourier transform*". *J. Chem. Phys.* vol. 121, 6117–6133 (2004).

113. O'Brien, J. L. *et al.* "*Quantum Process Tomography of a Controlled-NOT Gate*". *Phys. Rev. Lett.* vol. 93, 080502 (8 2004).

114. Myrskog, S. H., Fox, J. K., Mitchell, M. W. & Steinberg, A. M. "*Quantum process tomography on vibrational states of atoms in an optical lattice*". *Phys. Rev. A* vol. 72, 013615 (1 2005).

115. Howard, M. *et al.* "*Quantum process tomography and Linblad estimation of a solid-state qubit*". *New. J. Phys.* vol. 8, 33–33 (2006).

116. Chow, J. M. *et al.* "*Randomized Benchmarking and Process Tomography for Gate Errors in a Solid-State Qubit*". *Phys. Rev. Lett.* vol. 102, 090502 (9 2009).

117. Li, C.-F., Tang, J.-S., Li, Y.-L. & Guo, G.-C. "*Experimentally witnessing the initial correlation between an open quantum system and its environment*". *Phys. Rev. A* vol. 83, 064102 (6 2011).

118. Gessner, M. *et al.* "*Local detection of quantum correlations with a single trapped ion*". *Nat. Phys.* vol. 10, 105–109 (2014).

119. Štelmachovič, P. & Bužek, V. "*Dynamics of open quantum systems initially entangled with environment: Beyond the Kraus representation*". *Phys. Rev. A* vol. 64, 062106 (6 2001).

120. Rodríguez-Rosario, C. A., Modi, K., Kuah, A.-m., Shaji, A. & Sudarshan, E. C. G. "*Completely positive maps and classical correlations*". *J. Phys. A-Math. Theor.* vol. 41, 205301 (2008).

121. Rodríguez-Rosario, C. A., Modi, K. & Aspuru-Guzik, A. "*Linear assignment maps for correlated system-environment states*". *Phys. Rev. A* vol. 81, 012313 (1 2010).

122. Jordan, T. F., Shaji, A. & Sudarshan, E. C. G. "*Dynamics of initially entangled open quantum systems*". *Phys. Rev. A* vol. 70, 052110 (5 2004).

123. Shaji, A. & Sudarshan, E. "*Who's afraid of not completely positive maps?*" *Phys. Lett. A* vol. 341, 48–54 (2005).

124. Jordan, T. F., Shaji, A. & Sudarshan, E. C. G. "*Mapping the Schrödinger picture of open quantum dynamics*". *Phys. Rev. A* vol. 73, 012106 (1 2006).

125. Carteret, H. A., Terno, D. R. & Życzkowski, K. "*Dynamics beyond completely positive maps: Some properties and applications*". *Phys. Rev. A* vol. 77, 042113 (4 2008).

126. Modi, K. "*Preparation of States in Open Quantum Mechanics*". *Open Syst. Inf. Dyn.* vol. 18, 253–260 (2011).

127. Xu, G. F., Liu, L. J. & Tong, D. M. "*Effect of preparation procedures on the system's entanglement evolution*". *Eur. Phys. J. D.* vol. 66, 236 (2012).

128. Majeed, M. & Chaudhry, A. Z. "*Effect of initial system–environment correlations with spin environments*". *Eur. Phys. J. D.* vol. 73, 16 (2019).

129. Peres, A. "*What's Wrong with These Observables?*" *Foundations of Physics* vol. 33, 1543–1547 (2003).

130. Ringbauer, M. *et al.* "*Characterizing Quantum Dynamics with Initial System-Environment Correlations*". *Phys. Rev. Lett.* vol. 114, 090402 (9 2015).

131. Blok, M. S. *et al.* "*Manipulating a qubit through the backaction of sequential partial measurements and real-time feedback*". *Nat. Phys.* vol. 10, 189–193 (2014).

132. Shrapnel, S., Costa, F. & Milburn, G. "*Quantum Markovianity as a supervised learning task*". *Int. J. Quantum. Inf.* vol. 16, 1840010 (2018).

133. Giarmatzi, C. & Costa, F. "*Witnessing quantum memory in non-Markovian processes*", arXiv:1811.03722 [quant–ph] (2018).

134. Winick, A., Wallman, J. J. & Emerson, J. "*Phenomenological measure of quantum non-Markovianity*", arXiv:1901.00267 [quant–ph] (2019).

135. Morris, J., Pollock, F. A. & Modi, K. "*Non-Markovian memory in IBMQX4*", arXiv:1902.07980 [quant–ph] (2019).

136. Taranto, P., Milz, S., Pollock, F. A. & Modi, K. "*Structure of quantum stochastic processes with finite Markov order*". *Phys. Rev. A* vol. 99, 042108 (4 2019).

137. Taranto, P., Pollock, F. A., Milz, S., Tomamichel, M. & Modi, K. "*Quantum Markov Order*". *Phys. Rev. Lett.* vol. 122, 140401 (14 2019).

138. Guo, Y. *et al.* "*Experimental Demonstration of Quantum Processes with Finite Memory*", arXiv:2003.14045 [quant–ph] (2020).

139. Shapiro, M. & Brumer, P. "*Quantum Control of Molecular Processes*" (Wiley, 2012).

140. Altafini, C. & Ticozzi, F. "*Modeling and Control of Quantum Systems: An Introduction*". *IEEE T. Automat. Contr.* vol. 57, 1898–1917 (2012).

141. Feller, W. "*An Introduction to Probability Theory and Its Applications*". *An Introduction to Probability Theory and Its Applications* **v. 1** (Wiley, 1968).

142. Tao, T. "*An Introduction to Measure Theory*" (American Mathematical Society, 2011).

143. Kolmogorov, A., Bharucha-Reid, A. & Morrison, N. "*Foundations of the Theory of Probability: Second English Edition*" (Dover Publications, 2018).

144. Hänggi, P. & Thomas, H. "*Time evolution, correlations, and linear response of non-Markov processes*". *Z. Phys. B Cond. Mat.* vol. 26, 85–92 (1977).

145. Hanggi, P., Thomas, H., Grabert, H. & Talkner, P. "*Note on time evolution of non-Markov processes*". *J. Stat. Phys.* vol. 18, 155–159 (1978).

146. Vacchini, B., Smirne, A., Laine, E.-M., Piilo, J. & Breuer, H.-P. "*Markovianity and non-Markovianity in quantum and classical systems*". *New. J. Phys.* vol. 13, 093004 (2011).

147. Hänggi, P. & Thomas, H. "*Stochastic processes: Time evolution, symmetries and linear response*". *Phys. Rep.* vol. 88, 207–319 (1982).

148. Dynkin, E., Kovary, T. & Brown, D. "*Theory of Markov Processes*" (Dover Publications, 2006).

149. Stroock, D. "*An Introduction to Markov Processes*" (Springer Berlin Heidelberg, 2013).

150. Van Kampen, N. "*Remarks on Non-Markov Processes*". en. *Braz. J. Phys.* vol. 28, 90–96 (1998).

151. Pollock, F. A., Rodríguez-Rosario, C., Frauenheim, T., Paternostro, M. & Modi, K. "*Non-Markovian quantum processes: Complete framework and efficient characterization*". *Phys. Rev. A* vol. 97, 012127 (1 2018).

152. Pollock, F. A., Rodríguez-Rosario, C., Frauenheim, T., Paternostro, M. & Modi, K. "*Operational Markov Condition for Quantum Processes*". *Phys. Rev. Lett.* vol. 120, 040405 (4 2018).

153. Milz, S., Pollock, F. A. & Modi, K. "*Reconstructing non-Markovian quantum dynamics with limited control*". *Phys. Rev. A* vol. 98, 012108 (1 2018).

154. Milz, S., Sakuldee, F., Pollock, F. A. & Modi, K. "*Kolmogorov extension theorem for (quantum) causal modelling and general probabilistic theories*". *Quantum* vol. 4, 255 (2020).

155. Lindblad, G. "*Non-Markovian quantum stochastic processes and their entropy*". *Commun. Math. Phys.* vol. 65, 281–294 (1979).

156. Oreshkov, O., Costa, F. & Brukner, C. "*Quantum correlations with no causal order*". *Nat. Commun.* vol. 3, 1092 (2012).

157. Costa, F. & Shrapnel, S. "*Quantum causal modelling*". *New J. Phys.* vol. 18, 063032 (2016).

158. Chiribella, G., D'Ariano, G. M. & Perinotti, P. "*Theoretical framework for quantum networks*". *Phys. Rev. A* vol. 80 (2009).

159. Hardy, L. "*The operator tensor formulation of quantum theory*". *Philos. T. R. Soc. A* vol. 370, 3385–3417 (2012).

160. Hardy, L. "*Operational General Relativity: Possibilistic, Probabilistic, and Quantum*", arXiv:1608.06940 [gr–qc] (2016).

161. Kretschmann, D. & Werner, R. F. "*Quantum channels with memory*". *Phys. Rev. A* vol. 72, 062323 (6 2005).

162. Caruso, F., Giovannetti, V., Lupo, C. & Mancini, S. "*Quantum channels and memory effects*". *Rev. Mod. Phys.* vol. 86, 1203–1259 (4 2014).

163. Chiribella, G., D'Ariano, G. M. & Perinotti, P. "*Transforming quantum operations: Quantum supermaps*". *Europhys. Lett.* vol. 83, 30004 (2008).

164. Chiribella, G., D'Ariano, G. M., Perinotti, P. & Valiron, B. "*Quantum computations without definite causal structure*". *Phys. Rev. A* vol. 88, 022318 (2 2013).

165. Oreshkov, O. & Giarmatzi, C. "*Causal and causally separable processes*". *New. J. Phys.* vol. 18, 093020 (2016).

166. Portmann, C., Matt, C., Maurer, U., Renner, R. & Tackmann, B. "*Causal Boxes: Quantum Information-Processing Systems Closed Under Composition*". *IEEE T. Inform. Theory* vol. 63, 3277–3305 (2017).

167. Milz, S., Pollock, F. A., Le, T. P., Chiribella, G. & Modi, K. "*Entanglement, non-Markovianity, and causal non-separability*". *New. J. Phys.* vol. 20, 033033 (2018).

168. Strasberg, P. "*Operational approach to quantum stochastic thermodynamics*". *Phys. Rev. E* vol. 100, 022127 (2 2019).

169. Strasberg, P. "*Repeated Interactions and Quantum Stochastic Thermodynamics at Strong Coupling*". *Phys. Rev. Lett.* vol. 123 (2019).

170. Rivas, Á., Huelga, S. F. & Plenio, M. B. "*Quantum non-Markovianity: characterization, quantification and detection*". *Rep. Prog. Phys.* vol. 77, 094001 (2014).

171. Rivas, Á., Huelga, S. F. & Plenio, M. B. "*Entanglement and Non-Markovianity of Quantum Evolutions*". *Phys. Rev. Lett.* vol. 105, 050403 (5 2010).

172. Hou, S., Yi, X., Yu, S. & Oh, C. "*Alternative non-Markovianity measure by divisibility of dynamical maps*". *Phys. Rev. A* vol. 83, 062115 (6 2011).

173. Wolf, M. M., Eisert, J., Cubitt, T. S. & Cirac, J. I. "*Assessing Non-Markovian Quantum Dynamics*". *Phys. Rev. Lett.* vol. 101, 150402 (15 2008).

174. Usha Devi, A. R., Rajagopal, A. K. & Sudha. "*Open-system quantum dynamics with correlated initial states, not completely positive maps, and non-Markovianity*". *Phys. Rev. A* vol. 83, 022109 (2 2011).

175. Devi, A. R. U., Rajagopal, A. K., Shenoy, S. & Rendell, R. W. "*Interplay of Quantum Stochastic and Dynamical Maps to Discern Markovian and Non-Markovian Transitions*". *J. Quantum Inf. Sci.* vol. 02, 47–54 (2012).

176. Chruściński, D. & Maniscalco, S. "*Degree of Non-Markovianity of Quantum Evolution*". *Phys. Rev. Lett.* vol. 112, 120404 (12 2014).

177. Breuer, H.-P., Laine, E.-M. & Piilo, J. "*Measure for the Degree of Non-Markovian Behavior of Quantum Processes in Open Systems*". *Phys. Rev. Lett.* vol. 103, 210401 (21 2009).

178. Chruściński, D., Kossakowski, A. & Rivas, Á. "*Measures of non-Markovianity: Divisibility versus backflow of information*". *Phys. Rev. A* vol. 83, 052128 (5 2011).

179. Sakuldee, F., Milz, S., Pollock, F. A. & Modi, K. "*Non-Markovian quantum control as coherent stochastic trajectories*". *J. Phys. A-Math. Theor.* vol. 51, 414014 (2018).

180. Hoeffding, W. "*On Probabilities of Large Deviations*" (eds Fisher, N. I. & Sen, P. K.) 473–490 (Springer New York, New York, NY, 1994).

181. Cover, T. & Thomas, J. "*Elements of Information Theory*" (Wiley, 2012).

182. Hiai, F. & Petz, D. "*The proper formula for relative entropy and its asymptotics in quantum probability*". *Commun. Math. Phys.* vol. 143, 99–114 (1991).

183. Vedral, V., Plenio, M. B., Jacobs, K. & Knight, P. L. "*Statistical inference, distinguishability of quantum states, and quantum entanglement*". *Phys. Rev. A* vol. 56, 4452–4455 (6 1997).

184. Yukalov, V. "*Decoherence and equilibration under nondestructive measurements*". *Ann. Phys.* vol. 327, 253–263 (2012).

185. Weingarten, D. "*Asymptotic behavior of group integrals in the limit of infinite rank*". *J. Math. Phys.* vol. 19, 999–1001 (1978).

186. Collins, B. "*Moments and cumulants of polynomial random variables on unitarygroups, the Itzykson-Zuber integral, and free probability*". *Int. Math. Res. Notices* vol. 2003, 953 (2003).

187. Gu, Y. "*Moments of Random Matrices and. Weingarten Functions*". MA thesis (Queen's University, Ontario, Canada, 2013).

188. Collins, B. & Śniady, P. "*Integration with Respect to the Haar Measure on Unitary, Orthogonal and Symplectic Group*". *Commun. Math. Phys.* vol. 264, 773–795 (2006).

189. Puchała, Z. & Miszczak, J. A. "*Symbolic integration with respect to the Haar measure on the unitary groups*". *Bull. Pol. Acad. Sci. Tech. Sci.* vol. 65, 21 (2017).

190. Guhr, T., Müller–Groeling, A. & Weidenmüller, H. A. "*Random-matrix theories in quantum physics: common concepts*". *Phys. Rep.* vol. 299, 189–425 (1998).

191. Mehta, M. "*Random Matrices*" (Elsevier Science, 2004).

192. Milman, V. D. & Schechtman, G. "*Asymptotic Theory of Finite Dimensional Normed Spaces*". *Lecture Notes in Mathematics* **1200** (Springer-Verlag, 1986).

193. Boucheron, S., Lugosi, G. & Massart, P. "*Concentration Inequalities: A Nonasymptotic Theory of Independence*" (OUP Oxford, 2013).

194. Zhang, L. "*Matrix integrals over unitary groups: An application of Schur-Weyl duality*", arXiv:1408.3782 [quant–ph] (2014).

195. Ginory, A. & Kim, J. "*Weingarten calculus and the IntHaar package for integrals over compact matrix groups*". *J. Symb. Comput.* (2019).

196. Epstein, J. M. & Whaley, K. B. "*Quantum speed limits for quantum-information-processing tasks*". *Phys. Rev. A* vol. 95, 042314 (4 2017).

197. Deza, M. & Deza, E. "*Encyclopedia of Distances*" (Springer Berlin Heidelberg, 2009).

198. Taranto, P., Modi, K. & Pollock, F. A. "*Emergence of a fluctuation relation for heat in nonequilibrium Landauer processes*". *Phys. Rev. E* vol. 97, 052111 (5 2018).

199. Tamascelli, D., Smirne, A., Huelga, S. F. & Plenio, M. B. "*Nonperturbative Treatment of non-Markovian Dynamics of Open Quantum Systems*". *Phys. Rev. Lett.* vol. 120, 030402 (3 2018).

200. Luchnikov, I. A., Vintskevich, S. V., Ouerdane, H. & Filippov, S. N. "*Simulation Complexity of Open Quantum Dynamics: Connection with Tensor Networks*". *Phys. Rev. Lett.* vol. 122, 160401 (16 2019).

201. Anza, F., Gogolin, C. & Huber, M. "*Eigenstate Thermalization for Degenerate Observables*". *Phys. Rev. Lett.* vol. 120, 150603 (15 2018).

202. Pollock, F. A. & Modi, K. "*Tomographically reconstructed master equations for any open quantum dynamics*". *Quantum* vol. 2, 76 (2018).

203. Knill, E. "*Approximation by Quantum Circuits*". *arXiv:quant-ph/9508006* (1995).

204. Low, R. A. "*Large deviation bounds for k-designs*". *P. Roy. Soc. A-Math.Phy.* vol. 465, 3289–3308 (2009).

205. Nakata, Y., Hirche, C., Koashi, M. & Winter, A. "*Efficient Quantum Pseudorandomness with Nearly Time-Independent Hamiltonian Dynamics*". *Phys. Rev. X* vol. 7, 021006 (2 2017).

206. Nakata, Y., Hirche, C., Morgan, C. & Winter, A. "*Unitary 2-designs from random X- and Z-diagonal unitaries*". *J. Math. Phys.* vol. 58, 052203 (2017).

207. Dankert, C., Cleve, R., Emerson, J. & Livine, E. "*Exact and approximate unitary 2-designs and their application to fidelity estimation*". *Phys. Rev. A* vol. 80, 012304 (1 2009).

208. Wallman, J. J. & Flammia, S. T. "*Randomized benchmarking with confidence*". *New. J. Phys.* vol. 16, 103032 (2014).

209. Emerson, J., Alicki, R. & Życzkowski, K. "*Scalable noise estimation with random unitary operators*". *J. Opt. B: Quantum S. O.* vol. 7, S347–S352 (2005).

210. Gross, D., Audenaert, K. & Eisert, J. "*Evenly distributed unitaries: On the structure of unitary designs*". *J. Math. Phys.* vol. 48, 052104 (2007).

211. Harrow, A. W. & Low, R. A. "*Random Quantum Circuits are Approximate 2-designs*". *Commun. Math. Phys.* vol. 291, 257–302 (2009).

212. Nakata, Y. & Murao, M. "*Diagonal unitary 2-designs and their implementations by quantum circuits*". *Int. J. Quantum Inf.* vol. 11, 1350062 (2013).

213. Webb, Z. "*The Clifford group forms a unitary 3-design*". *Quantum Inf. Comput.* vol. 16, 1379–1400 (2016).

214. Zhu, H. "*Multiqubit Clifford groups are unitary 3-designs*". *Phys. Rev. A* vol. 96, 062336 (6 2017).

215. Brandão, F. G. S. L., Harrow, A. W. & Horodecki, M. "*Local Random Quantum Circuits are Approximate Polynomial-Designs*". *Commun. Math. Phys.* vol. 346, 397–434 (2016).

216. Gilchrist, A., Langford, N. K. & Nielsen, M. A. "*Distance measures to compare real and ideal quantum processes*". *Phys. Rev. A* vol. 71, 062310 (6 2005).

217. Sacchi, M. F. "*Optimal discrimination of quantum operations*". *Phys. Rev. A* vol. 71, 062340 (6 2005).

218. Taranto, P., Pollock, F. A. & Modi, K. "*Memory Strength and Recoverability of Non-Markovian Quantum Stochastic Processes*", arXiv:1907.12583 [quant–ph] (2019).

219. Chiribella, G., D'Ariano, G. M. & Perinotti, P. "*Memory Effects in Quantum Channel Discrimination*". *Phys. Rev. Lett.* vol. 101, 180501 (18 2008).

220. Gutoski, G. "*On a measure of distance for quantum strategies*". *J. Math. Phys.* vol. 53, 032202 (2012).

221. Aharonov, D., Kitaev, A. & Nisan, N. "*Quantum Circuits with Mixed States*", 20–30 (Association for Computing Machinery, Dallas, Texas, USA, 1998).

222. Paulsen, V. "*Completely Bounded Maps and Operator Algebras*" (Cambridge University Press, 2003).

223. Nakata, Y., Hirche, C., Morgan, C. & Winter, A. "*Decoupling with random diagonal unitaries*". *Quantum* vol. 1, 18 (2017).

224. De Vega, I. "*The quantum dynamical map of the spin boson model*", arXiv:2001.04236 [quant–ph] (2020).

225. Cerrillo, J. & Cao, J. "*Non-Markovian Dynamical Maps: Numerical Processing of Open Quantum Trajectories*". *Phys. Rev. Lett.* vol. 112, 110401 (11 2014).

226. Guarnieri, G., Smirne, A. & Vacchini, B. "*Quantum regression theorem and non-Markovianity of quantum dynamics*". *Phys. Rev. A* vol. 90, 022110 (2 2014).

227. Knipschild, L. & Gemmer, J. "*Modern concepts of quantum equilibration do not rule out strange relaxation dynamics*". *Phys. Rev. E* vol. 101, 062205 (6 2020).

228. Banchi, L., Burgarth, D. & Kastoryano, M. J. "*Driven Quantum Dynamics: Will It Blend?*" *Phys. Rev. X* vol. 7, 041015 (4 2017).

229. Brenes, M., Pappalardi, S., Goold, J. & Silva, A. "*Multipartite Entanglement Structure in the Eigenstate Thermalization Hypothesis*". *Phys. Rev. Lett.* vol. 124, 040605 (4 2020).

230. Luchnikov, I. A., Vintskevich, S. V., Grigoriev, D. A. & Filippov, S. N. "*Machine Learning Non-Markovian Quantum Dynamics*". *Phys. Rev. Lett.* vol. 124 (2020).

231. Berk, G. D., Garner, A. J. P., Yadin, B., Modi, K. & Pollock, F. A. "*Resource theories of multi-time processes: A window into quantum non-Markovianity*", arXiv:1907.07003 [quant–ph] (2019).

232. Jørgensen, M. R. & Pollock, F. A. "*Exploiting the Causal Tensor Network Structure of Quantum Processes to Efficiently Simulate Non-Markovian Path Integrals*". *Phys. Rev. Lett.* vol. 123, 240602 (24 2019).

233. Giovannetti, V. & Burgarth, D. "*Improved Transfer of Quantum Information Using a Local Memory*". *Phys. Rev. Lett.* vol. 96 (2006).

234. Clivaz, F. *et al.* "*Unifying Paradigms of Quantum Refrigeration: A Universal and Attainable Bound on Cooling*". *Phys. Rev. Lett.* vol. 123, 170605 (17 2019).

235. Knill, E. *et al.* "*Randomized benchmarking of quantum gates*". *Phys. Rev. A* vol. 77, 012307 (1 2008).

236. Proctor, T., Rudinger, K., Young, K., Sarovar, M. & Blume-Kohout, R. "*What Randomized Benchmarking Actually Measures*". *Phys. Rev. Lett.* vol. 119, 130502 (13 2017).

237. Dirkse, B., Helsen, J. & Wehner, S. "*Efficient unitarity randomized benchmarking of few-qubit Clifford gates*". *Phys. Rev. A* vol. 99, 012315 (1 2019).

238. Harper, R. & Flammia, S. T. "*Fault-Tolerant Logical Gates in the IBM Quantum Experience*". *Phys. Rev. Lett.* vol. 122, 080504 (8 2019).

239. Helsen, J., Xue, X., Vandersypen, L. M. K. & Wehner, S. "*A new class of efficient randomized benchmarking protocols*". *NPJ Quantum Inf.* vol. 5 (2019).

240. Zhang, L. & Xiang, H. "*Average entropy of a subsystem over a global unitary orbit of a mixed bipartite state*". *Quantum Inf. Process.* vol. 16, 112 (2017).