Thermalization in Quantum Information Geometry
University of Amsterdam, Institute of Physics

Submitted in partial fulfillment of the requirements for the degree of Master of Science

Daniel Chernowitz

August 2017
Abstract

In this work we approach Thermalization of discrete quantum systems from a geometrical vantage point. The aim is statistical results showing most Hamiltonians thermalize their subsystems. Fidelity, a concept in quantum information theory, relates distinguishability by measurements between states. We have numerically generated Hamiltonians according to the Haar measure and evolved the pure full system unitarily. We averaged the fidelity of subsystems between initial times and later times, finding them typically distinguishable. Conversely, the fidelity between two later times is seen to approach one for almost all systems. The inability to distinguish later times points to a steady state: the maximally mixed state. Moving on, we used Weingarten functions to integrate subsystem density matrices during evolution, over all possible Hamiltonian eigenvectors in the unitary group. This resulted in the mean subsystem density matrix, in which the initial conditions are seen to compete with the maximally mixed state over time. We observe a transition from a ‘known’ to ‘unknown’ phase after a time inversely proportional to the standard deviation in energies. However, the initial information is never completely lost. Also, polynomial functions of density matrices at any time can be integrated: we considered the variance over the unitary group and purity. In all cases, the subsystem mixing was seen to increase with bath size. Finally, we considered transformation properties of the Quantum Geometric Tensor under abelian and non-abelian gauge symmetries. The Bures metric, a natural generalization of the QGT, was found not to be gauge invariant in a degenerate spectrum.
Acknowledgements

It goes without saying that I could not have produced this work on my own. My main lifeline has been my preceptor, Vladimir Gritsev. This man has a mind for physics I can never hope to rival. If every theory is a phone line, professor Gritsev is the switchboard operator. No matter what problem I might have, he knows ten relevant articles. The only downside is that it can be hard to figure out which ideas are bad (there were many), if he can find merit in any of them. All in all, it has been my honor to stand so close to the source. Spasibo, Vladimir.

Besides this, I would like to extend gratitude to Māris Ozols. If not for our chance encounter, and him taking time to listen to my mathematical troubles, chapter 4 would have never come about. A suggestion can go a long way, and he showed me what was possible and what wasn’t. Thank you as well.
# Contents

1 Introduction to Thermalization 5  
  1.1 Isolated Quantum Systems .......................... 5  
  1.2 Quantum Geometry ................................. 7  

2 Theory and Setup 8  
  2.1 Motivation ........................................... 8  
  2.2 Setup ................................................. 9  
    2.2.1 Formal Problem ................................. 10  
  2.3 Methods .............................................. 12  
    2.3.1 Density Matrices ................................. 12  
    2.3.2 Tracing and Partial Tracing .................... 12  
  2.4 Measures of Mixing .................................. 15  
    2.4.1 Entanglement Entropy ........................... 15  
    2.4.2 Purity ........................................... 16  
  2.5 Quantum Distinguishability ......................... 16  
  2.6 Random Matrix Theory ............................... 18  
    2.6.1 Generating Unitary Matrices .................... 21  
  2.7 Rudimentary Group Theory ........................... 22  

3 Numerical Fidelity Simulations 24  
  3.1 Example Systems ...................................... 25  
  3.2 Statistics of Fidelity ................................ 28  
    3.2.1 Experiment Parameters .......................... 29  
  3.3 Results .............................................. 30  

4 Analytic Mixed State Averages 36  
  4.1 Analytic statistics of local density matrices .......... 36  
    4.1.1 Average Reduced Density Matrix ............... 37  
    4.1.2 Variance of the Reduced Density Matrix ........ 41  
  4.2 Mean Subsystem Dynamical Purity ..................... 50  

5 Quantum Differential Geometry 55  
  5.1 Manifold of Quantum States .......................... 55  
    5.1.1 Nondegenerate Case .............................. 55
Fill a glass of boiling water, and leave it alone in a room. After a while, the brightest minds cannot distinguish how exactly it was poured. It will cool to room temperature, and it might as well have been discharged freezing. There is no way to tell, because, after a while, all water looks the same. This, in essence, is the phenomenon called thermalization. It borrows its name from the thermal equilibrium attained with its surroundings. In nature thermalization is the rule and there are few exceptions. It appears any interplay with surroundings will thermalize a system. Moreover, in classical statistical mechanics, one of the celebrated results is that large collections of particles, even when isolated, maximize entropy and tend towards a thermal state. The same is even true on the smallest scales. Many quantum systems are seen to decohere, dissipate, and even with vanishingly weak coupling to the environment, thermalize. But this poses a fundamental challenge: contrary to the classical theory, quantum mechanics is inherently linear, and reversible - at the cost of an exponentially augmented state space. The reversibility means there must be a one-to-one mapping from initial states to later states by any evolution operator. Linearity precludes processes like classical chaos with a strong nonlinear dependence of later effects on initial conditions. This means an isolated system should not be capable of thermalization, or the trend to a single steady state from all of multiple, vastly diverse starting states. This begs a resolution.

1.1 Isolated Quantum Systems

One of the first to make this observation, and in 1929 a pioneer into quantum thermalization of isolated systems, was John von Neumann. His original work proved what is called the Quantum Ergodic Theorem, effectively contending that any energy eigenstate of macroscopic many-body quantum systems is in a sense thermal. He did this by exploring what is now termed 'normal typicality': For a
reasonable set of commuting macroscopic observables, the distribution of these observables evaluated on a state, restricted to an energy shell, is close to that of the micro-canonical ensemble at the same energy. In later years this result was heavily criticized as being "dynamically vacuous", and after Von Neumann’s death in 1957 largely forgotten.\[2\]. In the 1970s Michael Berry reinvigorated the research, by considering the quantum analogs of classical systems that are regular (integrable) or irregular (ergodic) \[3\]. He found that this distinction carried over to the statistics of the wave function in the quantum case, and conjectured that for many body quantum systems, the distribution of eigenstate momenta is Gaussian. He expected this to hold for classically chaotic systems only. Josh Deutsch continued this reasoning in the 1990s, leading up to Srednicki postulating that excited states of many-body wave functions appear to be plane waves with normally distributed momenta, and formulating what is known as the Eigenstate Thermalization Hypothesis or ETH \[4][5\]. This is one of the most frequently cited frameworks to describe why some systems thermalize, and notably, others don’t. The ETH is a statement about observables. It contends that for any initial state, any instantaneous observable quantity in that state will on average evolve irreversibly towards the expectation value in the micro-canonical ensemble. The observable values then only depend on the global nature of the system, and the approximate energy level. The hypothesis is predicated on a set of conditions. These are that in the basis of increasing energy eigenstates, the observable’s operator elements are close to a diagonal form varying smoothly with energy, up to a margin inversely proportional to the Hilbert space dimension. Then additionally, the observable will not stray far from this ensemble value at any later time. It holds for typical random Hamiltonians \[6\]. Since this original formulation, other have chipped away the criticism initially imparted to Von Neumann, leading to his ultimate vindication in 2010 by Goldstein et al \[7\].

The cases which do not thermalize fall mostly into two categories. They are either classically integrable (perhaps non-interacting), or live in a low-energy regime and exhibit many-body localization (MBL), manifestly the converse of plane-wave eigenstates. In order to envision MBL states, consider low-lying single particle energy eigenstates that are strongly spatially local, living on certain lattice sites, for instance. MBL is said to be observed when, after turning on interactions, the many-body wave functions remain local, in a slightly more diffuse sense. They are however not simply products of local single particle states. An initial state centered on a certain position will retain traces of its spatial distribution there for all times \[8\].

Alternatives to the ETH focus on the states themselves, and are usually formulated in terms of the occupation of energy eigenstates in the initial superposition. They argue that this occupation must be uncorrelated to the observable’s matrix element, and thus the expectation value of this operator constitutes random sampling from its elements, recovering the microcanonical ensemble. However, for many initial states with specific, out-of-equilibrium observable values, this lack of correlation is questionable \[9\].

If we do not wish to make reference to temperatures and energies, the former
concept somewhat alien to finite discrete systems, we formally only describe equilibration \cite{10}. In this case, the system moves away from its initial condition to a steady, perhaps predictable state, which in general need not be given by any particular (micro)canonical ensemble. It has already been argued appreciably in \cite{11} that we may expect statistically for a typical quantum system to equilibrate independent of its microscopic initial conditions. In the next section, we will shift gears, and discuss methods, more than problems.

### 1.2 Quantum Geometry

There is a new emerging school of thought in the field of quantum condensed matter and quantum statistical physics. Since the seminal paper by Aharonov and Bohm in 1959, increasingly geometric aspects have been found crucial for a full understanding of quantum mechanics \cite{12}\cite{13}\cite{14}. In the 1980s, many insights were achieved. Provost and Vallee are credited with the idea of constructing a metric tensor on a manifold of pure quantum ground states, parameterized by some experimental dials \cite{15}. After this, Berry again broke new ground by observing his now eponymous geometric phase, explaining anew the aforementioned Aharonov Bohm effect \cite{16}. Wilczek and Zee generalized this to non-abelian phases \cite{17}, employing work from the mathematician Uhlmann \cite{18}. This opened the gates to the *Bures metric* \cite{19}\cite{20}, the natural augmentation of the tensor of Provost and Vallee to mixed state density matrices. This metric is directly related to the *fidelity*, a quantum-informatic measure of distance inspired by distinguishability through measurements \cite{21}\cite{22}. Currently, many have turned to the powerful mathematical techniques of geometry to describe thermal ensembles and phenomena such as critical behavior. See \cite{23} by Rezakhani et al. and \cite{24} by Zanardi et al. In the first of this pair of papers, the aim is to imbue quantum states of a system with a Riemannian metric as it evolves adiabatically in time through Hilbert space. In the second, over the manifold of thermal Gibbs states at different temperatures. Also nonadiabatic responses have been studied \cite{25}.

Inspired by the elegance and effectiveness in the chronology above, this work will also make use of geometrical tools, to tackle the problem of thermalization in isolated discrete quantum systems. We hope to make statistical arguments evocative of Neumann’s, explained and expanded by Reimann \cite{10}. In the latter’s work, it is not proven that all systems thermalize all observables, as there are obvious exceptions. However, over the full space of possible macroscopic systems, one can constrain the proportion of these systems that deviates appreciably from microcanonical predictions, and the share of time they do so. These proportions turn out to scale inversely with a quantity exponential in system size. In other words, just as in the classical case, failure to thermalize is possible, but among ‘uniformly sampled systems’ exceedingly unlikely to observe in practice.
Chapter 2

Theory and Setup

As a first step into the actual physics of thermalization, we will elaborate on theory, formulas and formalisms that are useful later on. The criterion for material included in this chapter is that these results are all borrowed, in some regard standard, or basic, whereas the subsequent chapters contain more original work. The last two sections (2.6, 2.7) are rather disconnected from the narrative of the first five, but are nonetheless needed at a certain point.

2.1 Motivation

As mentioned in the introduction, quantum mechanical time evolution is unitary. This means no information can be lost, because we can always in principle reverse the evolution and obtain the state at any time prior, so long as we know the state at present and know the Hamiltonian $H$. This follows from the fact that any unitary operator has an inverse, in this case generated by $-H$. Imagine we have a system evolving according to:

$$\hat{U}(t) := e^{-iHt}$$ (2.1)

An operator translating states in time. In this work we have set $\hbar = 1$. As a Gedanken experiment, start the system in either of two states, $|\psi\rangle$ or $|\phi\rangle$. Assume $|\langle \psi |\phi \rangle| < 1$, then we may think of there being some distance in Hilbert space $\mathcal{H}$ between them. We will later learn that $|\langle \psi |\phi \rangle|$ is called the fidelity between these states, and it is a genuinely geometrical concept. Continuing, it is immediate that this distance cannot grow over time. We work in the Schrödinger picture, so time is kept on the states. Then:

$$|\psi_t\rangle \equiv \hat{U}(t) |\psi\rangle, \quad |\phi_t\rangle \equiv \hat{U}(t) |\phi\rangle$$ (2.2)

The distance at a later time is given by:

$$|\langle \psi_t |\phi_t \rangle | = |\langle \psi |\hat{U}^{\dagger}(t)\hat{U}(t)|\phi \rangle | = |\langle \psi |\phi \rangle |$$ (2.3)
The inner product of Hilbert space vectors is conserved by the action of a Hermitian Hamiltonian. In the jargon of classical chaos theory, the Lyapunov exponent is zero [5].

However, in practice quantum mechanical systems do exhibit a kind of thermalization. For many such systems, the properties at late times seem independent of most of the specifics of the initial conditions, save perhaps some global quantities such as total energy [26][1][10][9][11]. The resolution, spearheaded by Von Neumann in the 1920s, lies in a shift of focus away from states, to operators. The quantum state offers the best possible description of the system, given that it cannot be fully known due to the Heisenberg uncertainty relation. However, a state is a mathematical object, and an experimenter does not deal with the state directly. He or she performs measurements on the system, and these are codified by operators evaluated on the states. Moreover, we deal only with those operators we choose to represent our preferred observables, the ones we can access and in which we are interested. This restriction will allow us to understand the mechanism of losing information in quantum mechanics. The moral will be that the information of course is not lost, it is just very inaccessible for an experimenter with measurement apparatus that live in a certain kind of space, corresponding to operators that are natural only in a certain basis.

2.2 Setup

In order to formalize these claims, we first construct some standard machinery [27][11]. For the rest of this work, we may envision our full quantum system to be partitioned into two parts. There is the subsystem, $S$, and the much larger environment, or bath, $B$. The intuition should be that an experimenter has access to $S$, and not to $B$. However, he or she cannot prevent the two from interacting. Moreover, from the perspective of the full system $S+B$, the division may be highly arbitrary and unnatural. There is simply some rule designating certain degrees of freedom to correspond to $S$, and the rest to $B$. Think of a set of coupled quantum spins, of which a subset are selected for analysis. But also some specific Fourier modes or particles, or the inhabitants of some submanifold of real space will do. See figure 2.1 for an illustration.

As both $S$ and $B$ are quantum, when isolated, they are in some quantum states $|\psi_S\rangle \in \mathcal{H}_S$ and $|\psi_B\rangle \in \mathcal{H}_B$, respectively. The full coupled Hilbert space is the tensor product of the constituent spaces:

$$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B \quad (2.4)$$

And given any two bases of the constituents, one can immediately construct a basis of the full system:

$$\mathcal{H}_S = \text{span}(\{|\phi^S_j\rangle\}), \quad \mathcal{H}_B = \text{span}(\{|\phi^B_k\rangle\}), \quad \mathcal{H} = \text{span}(\{|\phi^S_j\rangle \otimes |\phi^B_k\rangle\}) \quad (2.5)$$
2.2.1 Formal Problem

Turning now to the isolated subsystem, forget the coupling to the bath momentarily. We take $S$ to start at $t = 0$ in a well defined state $|\psi_S\rangle$, evolving according to a unitary operator $\hat{U}_S(t) := e^{-iH_S t} = \hat{U}_S$, we will suppress the argument for now. Say, at time $t = 0$, we know some property of the system, an observable $\hat{O}_S$, describing some measurement we can perform. The intuition should be that this observable is local, in a sense that will become clear below. For definiteness assume $|\psi_S\rangle$ is an eigenstate of $\hat{O}_S$, so there is no uncertainty of the outcome. Then our initial information, or condition, is the value $x$:

$$x = \langle \psi_S | \hat{O}_S | \psi_S \rangle$$  \hspace{1cm} (2.6)

One could generalize to an optimal set of commuting operators $\{\hat{O}_S^k\}$, of which the expectation values $x_k$ form the maximal available initial information. We will illustrate the single operator case.

This approach means our initial information is given by observable values, not states, and losing it corresponds to the inability to perform measurements that output the value. When we are further along in time, the state is $\hat{U}_S |\phi_S\rangle$. We can find the expected instantaneous observable by taking the inner product $\langle \phi_S | \hat{U}_S^\dagger \hat{O}_S \hat{U}_S | \phi_S \rangle$, which may differ from the experiment outcome: assume $[\hat{H}_S, \hat{O}_S] \neq 0$. Instead, we are interested in recovering the initial conditions from the state at this later time, so we will need a different operator. The proposed form is evident: $\hat{U}_S \hat{O}_S \hat{U}_S^\dagger$, which corresponds to evolving the state backwards to the initial time, measuring the state, and evolving forward again. It is manifestly Hermitian, so a well defined observable, and its action is the following:

$$\langle \psi_S | \hat{U}_S^\dagger \hat{U}_S \hat{O}_S \hat{U}_S^\dagger | \psi_S \rangle = \langle \psi_S | \hat{O}_S | \psi_S \rangle \equiv x$$  \hspace{1cm} (2.7)
Although this is theoretically possible, it is impractical. The operator’s form is very sensitive to the exact time of evaluation and value of $\hat{H}_S$. It is generally difficult to construct. Still, there is no conceptual mechanism for information loss. We must restore the bath.

In what follows, we will always assume the subsystem and bath at initial time are in a product state. In other words: there is no initial entanglement. This will allow us to gauge to what extent the Hamiltonian generates it. The full system at $t = 0$ is in the state $|\psi\rangle = |\psi_S\rangle |\psi_B\rangle$, unwritten tensor multiplication implied.

In order to find our initial condition, we must also augment our observable:

$$\hat{O} := \hat{O}_S \otimes \mathbb{1}_B \quad (2.8)$$

This is our working definition of a local operator. It only depends on a few ‘close’ degrees of freedom: those contained in $S$. We call $\hat{O}$ local, when it is trivial on $B$: it does not act on $B$ or depend on it. We can now formalize the notion that there are ‘preferred observables’: after partitioning $S$ and $B$, we are restricted to local operators. The ones accessible to us can measure a subset of the degrees of freedom fully, but have no access to the rest.

Within this framework, the initial condition on $S$ is still well posed. Observe, in this larger Hilbert space, for any $x$,

$$\langle \psi | \hat{O} | \psi \rangle = \langle \psi_S | \hat{O}_S | \psi_S \rangle \langle \psi_B | \mathbb{1}_B | \psi_B \rangle = x \cdot 1 = x \quad (2.9)$$

Owing to the product nature of the state and operator, and normalization of all states. We can still describe the initial information the same way.

Subsequently we turn on time evolution $\hat{U}$ driven by a general $\hat{H}$ on $S + B$. At later time, the full system is in some state:

$$\hat{U} |\psi\rangle = \sum_{jk} C_{jk} |\phi^j_S\rangle |\phi^k_B\rangle \neq |\psi^S\rangle |\psi^B\rangle \quad (2.10)$$

In general, the state will be a superposition over many products of component vectors, which cannot be written as a product state through any basis transformation. This constitutes entanglement between $S$ and $B$. For most states, the Schmidt number (minimal number of terms in the sum above) is larger than one [14]. From here, endeavoring to recover initial information, we may attempt to reproduce the trick of equation 2.7. But we will fail on one count: $\hat{U} \hat{O} \hat{U}^\dagger$ is not local. It too will be an intricate superposition of tensored operators, not trivial on the bath. We cannot reverse coupled time evolution with a local operator. There is one exception: for $\hat{H} = \hat{H}_S \otimes \mathbb{1}_B + \mathbb{1}_S \otimes \hat{H}_B$, $\hat{U} = \hat{U}_S \otimes \hat{U}_B$, and all product forms are preserved. Without interaction, there is no entanglement, and there can be no information loss or thermalization.

We get a sense that what we call local is subjective, a consequence of what basis we use to describe our system. Trivially, there is always a basis which exhibits no thermalization, the basis of energy eigenstates. Their occupation is conserved in time, but the eigenstates do not in general respect the partition into $S$ and $B$: they are not simple products of subsystem and bath states, but
are entangled. They would be product states without interaction. The more the energy eigenstates resemble our lab coordinates, the better behaved (more accessible and predictable) our system will appear to us. Some systems will entangle and thermalize more than others, and one wonders what is typical. Such questions deserve quantification, which we will attempt in later sections.

2.3 Methods

In the previous section, the idea was elucidated of only having a local, partial access to a global state. This begs the question: what is our best local description of the subsystem \( S \)? What relates the information on \( S \) that can be obtained without access to \( B \)?

2.3.1 Density Matrices

For context, we remind the reader of the density matrix formalism. For any state in bra-ket notation, we may assign a density matrix [14]:

\[
|\psi\rangle \mapsto \rho_0 := |\psi\rangle \langle \psi| \tag{2.11}
\]

It is an operator, because it takes states to states -in this case a projection operator. Furthermore, it is Hermitian and positive semidefinite: Its eigenvalues are nonnegative.

So far, we have been certain, in a classically probabilistic sense, of what state the system occupies: \(|\psi\rangle\) with probability one. Therefore, \(\rho_0\) is a pure state, which is always the case when obtained as in 2.11. By contrast, if one is uncertain of the exact occupied state, one might wish to account for a discrete probability distribution over different states. In order to be concrete, say a \(p_j\) chance to be in state \(|\psi_j\rangle\). The corresponding density matrix of such a mixed state is the following:

\[
\rho := \sum_j p_j |\psi_j\rangle \langle \psi_j| \tag{2.12}
\]

It is equivalently an observable that measures the probability to be in each of the states \(|\psi_j\rangle\).

2.3.2 Tracing and Partial Tracing

We are already familiar with the trace, a linear mapping from operators to scalars. For any basis \(|\phi_j\rangle\) of \(\mathcal{H}\), and any operator \(\hat{Z} \in \mathcal{HS}\), the Hilbert-Schmidt space of operators on \(\mathcal{H}\), we define:

\[
\text{Tr} \hat{Z} = \sum_j \langle \phi_j | \hat{Z} | \phi_j \rangle \tag{2.13}
\]

Although formulated in terms of a basis, the trace is a basis independent value. It can be thought of as an integral over the Hilbert space. As \(\rho\) is an
operator also, we can take the trace $\text{Tr}(\rho)$, and we will find that it is always unit, as probabilities sum to 1. The following also holds due to 2.11 and the cyclic property of the trace:

$$\langle \psi | \hat{Z} | \psi \rangle = \text{Tr}(\rho_0 \hat{Z}) \quad (2.14)$$

So there is an equivalent formalism to obtain expectations, which generalizes straightforwardly to mixed states:

$$\text{Tr}(\rho \hat{Z}) = \text{Tr} \left( \sum_j p_j |\psi_j\rangle \langle \psi_j | \hat{Z} \right) = \sum_j p_j \langle \psi_j | \hat{Z} | \psi_j \rangle \quad (2.15)$$

By linearity of the trace, and of expectations, this becomes the expectation over the classical ensemble of expectations on quantum states: our best estimate of the observable.

A natural specification of the trace, is the partial trace: instead of integrating over the whole system, we only integrate out the bath (or subsystem, for that matter). This is done in practice by choosing a basis of $B$: $\{ |\phi^j_B \rangle \}$, and summing similar to 2.13, but now it’s a mapping to a smaller Hilbert space:

$$\text{Tr}_B \hat{Z} := \sum_j \langle \phi^j_B | \hat{Z} | \phi^j_B \rangle \quad (2.16)$$

Performing the partial trace on a pure density matrix $\rho_0$ in $\mathcal{HS}$ is a natural way to obtain a mixed density matrix on a constituent Hilbert-Schmidt space. Notably, the one not traced over.

$$\text{Tr}_B (\rho_0) = \rho_S \in \mathcal{HS}_S \quad (2.17)$$

This is our best description of $S$: a mixed state. The usefulness of the partial trace stems from the following. For any product state $|\psi\rangle = |\psi_S\rangle |\psi_B\rangle$ in 2.11 and operator of the form 2.8:

$$\langle \psi_S | \hat{O}_S | \psi_S \rangle = \text{Tr}(\hat{O}_S \rho_S) = \text{Tr}(\hat{O} \rho_0) = \langle \psi | \hat{O} | \psi \rangle \quad (2.18)$$

By linearity the last three expressions are valid for expectations over entangled states $|\psi\rangle$, without the ill-posed first equality in 2.18, as there is no $|\psi_S\rangle$.

Note that also $\text{Tr}_B (\hat{O}) \propto \hat{O}_S$. In fact, a basis independent definition of partial trace is that it is the unique linear operator satisfying:

$$\text{tr}_B(W_S \otimes X_B) = tr(X_B)W_S \quad \forall X_B \in \mathcal{HS}_B, W_S \in \mathcal{HS}_S \quad (2.19)$$

For an illustration of trace and partial trace, we elicit the simplest nontrivial example of a composite Hilbert space: when $S$ and $B$ are spin-$\frac{1}{2}$ particles, or qubits. Then each lives in

$$\mathcal{H}_S = \text{span}(\{|\uparrow\rangle, |\downarrow\rangle\}) = \mathcal{H}_B \quad (2.20)$$
A state in the full $\mathcal{H}$ will in general be a superposition of all combinations of subsystem configurations.

$$|\psi\rangle = C_0 |\uparrow\uparrow\rangle + C_1 |\uparrow\downarrow\rangle + C_2 |\downarrow\uparrow\rangle + C_3 |\downarrow\downarrow\rangle \in \mathcal{H}_S \otimes \mathcal{H}_B$$ (2.21)

Then treating the bra-vectors as column coordinates, and ket-vectors as rows,

$$\rho = |\psi\rangle \langle \psi| = |C_0|^2 |\uparrow\uparrow\rangle \langle \uparrow\uparrow| + C_0 C_1^* |\uparrow\downarrow\rangle \langle \uparrow\downarrow| + \ldots$$ (2.22)

naturally reshapes to a matrix, see the figure 2.2. Identify $a = C_0^2$, $b = C_0 C_1^*$ etc. from 2.22 in the figures. The partial trace is also drawn for this example in figure 2.3. Note that although this graphic represents the operations in a specific basis, the trace is basis independent, and the partial trace is independent of the bases of the constituent spaces.

We will work in the supposition that full systems are always pure. In the author’s view, there is no such thing as physical mixed states, which is instead a mathematical tool to express one’s lack of complete knowledge. Subsystems may be mixed, from an experimenters perspective, as full information on them requires access to the bath. This view is evidenced by [28].

Combining the Schrödinger equation on the $|\psi\rangle \sim \rho_0$ with the partial trace, we may construct the time evolution for density matrices.

$$\frac{d}{dt} \rho_S(t) = \text{Tr}_B (-i[H, \rho_0(t)])$$ (2.23)

This is called the *Liouville-Von Neumann equation* [27].
2.4 Measures of Mixing

Although touched upon in the previous section, there is more to be said than the binary division into conditions of mixed and pure. We envision a continuum of varying mixed states, and would like to delineate between cases with measures. We will define two.

2.4.1 Entanglement Entropy

The first is a well known concept, the entropy. One instinctively thinks of the amount of disarray. Many fields have their own definition of entropy, and in this field it is in a statistical sense, the uncertainty of a mixed state. Also known as the Von Neumann or entanglement entropy [14]:

$$S(\rho) := - \text{Tr} (\rho \ln \rho)$$  \hspace{1cm} (2.24)

Indeed, it is weakly linked to the classical entropy that counts microstates. This is no secret. Von Neumann famously quipped to Claude Shannon, the formula’s inventor: “You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.” [29]

Involving a matrix logarithm, 2.24 is only defined through diagonalization to the matrix’s eigenbasis. For a Hermitian operator $\hat{Z} \in \mathcal{H}S$ with dimension $d$, this always exists for some $W \in \mathcal{U}(d)$ -more on this in section 2.6- and in that basis any function $f$ is performed $f(\hat{Z}) = f(W \hat{D} W^\dagger) = W f(\hat{D}) W^\dagger$. The function is simply evaluated element-wise on the entries of the diagonal $D$. Due to the trace in 2.24, the expression is ultimately independent of the basis $W$ in which the state is represented.

The entropy is zero if and only if $\rho$ is pure. In its eigenbasis, $(\rho_0)_{j,k} = \delta_{j,1} \delta_{k,1}$ and $S(\rho_0)$ is understood as the limit

$$S(\rho_0) = \lim_{j,k} ((\rho_0 \ln \rho_0)_{j,k}) \delta_{j,k} = -1 \cdot 1 \cdot \ln(1) - (d-1) \cdot \lim_{x \to 0} (x \ln(x)) = 0 \quad (2.25)$$

The maximally mixed state is $\rho = \mathbb{1}/d$, and $S(\mathbb{1}/d) = \ln(d)$ also maximizes the entropy. This entropy has more attractive features, but these will suffice for our purposes. In the following chapters, we will sometimes wish to compare entropies in different size Hilbert spaces. For this, the normalized entropy is useful:

$$S_1(\rho) := \frac{S(\rho)}{S(\mathbb{1}/d)} = \frac{\text{Tr} (\rho \ln \rho)}{\ln(d)}$$  \hspace{1cm} (2.26)

Normalizing by the maximum entropy, $S_1(\rho) \leq 1$ for any space.
2.4.2 Purity

Moving on to an inverted measure, the purity \([30]\). The emphasis is that it measures how pure the density matrix is, not how mixed or entropic.

\[
\gamma(\rho) := \text{Tr} (\rho^2)
\]  

(2.27)

The smaller the purity, the higher the degree of mixing. \(\gamma(\rho) = 1\) is equivalent to \(\rho\) being pure. Our example in 2.11 gives \(\rho_0^2 = |\psi\rangle \langle \psi| |\psi\rangle \langle \psi| = \rho_0\) which has unit trace. An arbitrary mixed state defined in 2.12 has \(\gamma(\rho) = \sum_j p_j^2 \leq 1\), because squaring probabilities cannot increase them. The maximally mixed \(\rho = \mathbb{I}/d\) has:

\[
\gamma(\mathbb{I}/d) = 1/d
\]  

(2.28)

So \(\gamma(\rho) \in [1/d, 1]\). This function is computationally useful as it does not require diagonalization.

2.5 Quantum Distinguishability

Moving on, an underlying theme in the above has been distinguishability of quantum states, specifically obtained under different initial conditions. A number of useful tools shall be introduced here and in chapter 5 that formalize quantum distance. The first we encounter, a powerful concept, is the fidelity, which relates the resemblance between two states of the same system \([14]\). Mentioned already at the beginning of the section, for pure states it is a simple inner product modulus.

\[
F (|\psi\rangle, |\phi\rangle) := | \langle \psi | \phi \rangle |
\]  

(2.29)

Some properties:

- Due to the absolute signs, the fidelity is always real, positive, and is symmetric in its arguments.
- Due to normalization of Hilbert space vectors, \(F \in [0, 1]\).
- The fidelity is basis independent, as long as both states are in the same basis.
- Lastly, a fidelity of 1 means \(|\psi\rangle\) and \(|\phi\rangle\) are equal up to a \(U(1)\) gauge transformation factor.

Though not directly evident, these properties carry over when we generalize to mixed states \([22]\), on which the fidelity is also defined, thusly:

\[
F (\rho, \eta) := \text{Tr} \sqrt{\sqrt{\eta} \rho \sqrt{\eta}}
\]  

(2.30)
In order to verify that the definitions agree in the limiting case, first set one of the arguments pure: \( \eta = |\phi\rangle \langle \phi| = \eta^2 \) [21]:

\[
F(\rho, |\phi\rangle \langle \phi|) = \text{Tr} \sqrt{\sqrt{|\phi\rangle \langle \phi| \rho \sqrt{|\phi\rangle \langle \phi|}}} = \text{Tr} \sqrt{|\phi\rangle \langle \phi| \rho |\phi\rangle \langle \phi|} = \sqrt{|\phi\rangle \langle \phi|} \text{Tr} \sqrt{|\phi\rangle \langle \phi|} = \sqrt{|\phi\rangle \langle \phi|}
\]

From here, also specify \( \rho = |\psi\rangle \langle \psi| \):

\[
F(|\psi\rangle \langle \psi|, |\phi\rangle \langle \phi|) = \sqrt{|\phi\rangle \langle \psi| \langle \psi| \langle \phi|} = |\langle \psi| \phi|\rangle|
\]

We recover expression 2.29. An additional property of the mixed state fidelity ties it to the pure version as follows:

\[
F(\rho_S, \eta_S) = \max_{|\psi\rangle, |\phi\rangle} |\langle \psi| \phi|\rangle|
\]

Where the states \( |\psi\rangle, |\phi\rangle \) are called *purifications* of \( \rho_S \) and \( \eta_S \), i.e. states in an doubled Hilbert space such that tracing out one space reduces them to \( \rho_S \) and \( \eta_S \):

\[
|\psi\rangle, |\phi\rangle \in \mathcal{H}_S \otimes \mathcal{H}_S, \quad \text{Tr}_S \left( |\psi\rangle \langle \psi| \right) = \rho_S, \quad \text{Tr}_S \left( |\phi\rangle \langle \phi| \right) = \eta_S
\]

So the mixed fidelity is the maximal pure fidelity over possible purifications of the arguments [14].

Why are we interested in the fidelity specifically? It is uniquely motivated by the focus on operators posited above. Ideally, 'distance' in the Hilbert space is directly related to distinguishability by measurements. In order for this to be more formal, we refer to the theory of Positive Operator Value Measures (POVM) [30]. We recollect the fact that a quantum measurement has no predictable outcome in general. Instead, any possible eigenstate or eigen-subspace of the measurement-operator can result, with probability according to the Born rule. A certain measurement \( \hat{M} \) having a number of possible outcomes labelled by \( j \) induces a POVM: a set of operators \( \{ \hat{K}_j \} \) such that \( \sum_j \hat{K}_j = \hat{1} \), and the probability to find the \( j \)th outcome on state \( \rho \) is \( P(j) = \text{Tr}(\rho \hat{K}_j) \). By linearity total probability is one. The experimenter deals with these probabilities, and their distribution is his or her tool to map the preceding state, by means of *quantum tomography*.

The next ingredient is a result from classical probability theory: the *Bhattacharyya coefficient*, \( \cos \Delta \). For two discrete classical probability distributions \( \{ p_j \} \) and \( \{ q_j \} \), the overlap between the two statistical samples is given by:

\[
\cos \Delta = \sum_j \sqrt{p_j q_j}
\]
Where $\Delta$ is called the 'angle of divergence'. It is clear that $\cos \Delta = 1 \Leftrightarrow \{q_j\} = \{p_j\}$. The more distinct the distributions, the smaller the cosine, and thus the larger the angle of divergence. This angle is the most sensitive measure of the ability to distinguish the classical distributions.

As measurement on quantum states induces a classical probability distribution, a natural question to ask, for two differing states, is the following: What divergence between classical distributions resulting from these states is maximal over all possible measurements? Equivalently, what is the minimum Bhattacharyya coefficient on the $P(j)'s$, taken over all POVMs? The answer was proven to be the fidelity, 2.30, in [31].

$$F(\rho, \eta) = \min_{\{K_j\}} \sum_j \sqrt{\text{Tr}(\rho K_j) \text{Tr}(\eta K_j)}$$

(2.36)

Thus the fidelity directly relates an experimenters ability to distinguish states, using the ideal measurement for that purpose. We will later use the fidelity as the preferred identifier of 'change' to the quantum state. The larger $F$, the harder to tell the states apart, or the more similar they look.

In chapter 5, we will elaborate on differential concepts closely related to fidelity: the Quantum Geometric Tensor, and the Bures metric.

### 2.6 Random Matrix Theory

In discrete, finite Hilbert spaces, operators are represented by matrices. When we discuss statistics of these operators, of quantum systems, the unavoidable language is that of Random Matrix Theory (RMT). The idea is to treat a matrix as a random variable, living in a sample space, and following some probability distribution, just like any scalar variable. Strongly coupled real world Hamiltonian systems are often found to be well approximated by RMT, using only minimal knowledge of the actual inner workings of the models.

We will borrow from Eynard’s lecture notes [32].

Random matrices may be viewed as statistical variables drawn from a Random Matrix Ensemble (RME). The motivating example will be the ensemble of Hermitian matrices $H \in \mathcal{E}$. This is because Hamiltonians are Hermitian, and we will semantically equate a Hamiltonian to a specific quantum system.

In order to calculate probabilities in a continuous sample space, one needs a probability density function $\mathcal{P}(H)$, and a measure $dH$. The following is evident notation for the probability to find $H$ in some subset of the full space, or manifold, $\mathcal{E}$:

$$P(H \in Y \subset \mathcal{E}) = \int_Y \mathcal{P}(H) dH$$

(2.37)

And we will assume the measure and density are normalized:

$$\int_{\mathcal{E}} \mathcal{P}(H) dH = 1$$

(2.38)
The spectral theorem guarantees a Hermitian matrix has a complete set of linearly independent eigenvectors. A fact that will be used often in this work, all $d \times d$ Hermitian matrices $H$ can be diagonalized as follows:

$$H = V \Lambda V^\dagger, \quad V \in \mathcal{U}(d), \quad \Lambda = \text{diag}(E_1, E_2, \ldots, E_d) \in \mathbb{R}^d$$ (2.39)

Where $\Lambda$ is a diagonal matrix holding eigenvalues of $H$, evocatively denoted $E_i$ for energy, and $V$ is unitary matrix of size $d$ carrying the eigenvectors of $H$ in its columns. In this case $\mathcal{U}(d)$ is what we call the symmetry group, of the RME, a Lie group. Intuitively, we wish for the particular $V$ not to influence the probability of $H$ being drawn: thus it is the symmetric under the (left or right) action of $\mathcal{U}(d)$. Define the mapping $(\Lambda, V) \mapsto H$ through 2.39, then $(\Lambda, V)$ is equiprobable to $(\Lambda, VW)$ and $(\Lambda, WV)$ for any $W \in \mathcal{U}(d)$. This reflects the earlier stated sentiment that there is no preferred basis.

Moving on, we examine the objects $dH$ and $\mathcal{P}(H)$. The probability is conventionally written as a Boltzmann weight of a potential, which in turn is the trace of a polynomial in $H$:

$$\mathcal{P}(H) := \frac{1}{Z} e^{-V(H)} = \frac{1}{Z} e^{\text{Tr}(\sum_{n=0}^{d} g_n H^n)}$$ (2.40)

$Z$ is the partition function, and normalizes the full probability as in 2.38. According to the Cayley-Hamilton theorem, the matrix powers $\{H^n\}_{n=1}^d$ are linearly dependent, thus so are their traces by linearity. This means we may curtail the polynomial at order $d$. Observe that for $V \in \mathcal{U}(d)$, $\text{Tr}(V \Lambda V^\dagger)^n = \text{Tr}(V \Lambda^n V^\dagger) = \text{Tr}(\Lambda^n)$. Combine this with linearity of the trace, and we find that potentials of this form are independent of the basis $V$, as desired: $\mathcal{P}(H) = \mathcal{P}(\Lambda)$. To begin with, we usually take $g_2 = g$, other $g_n = 0$, in which case the potential is Gaussian. This Gaussian distribution and unitary symmetry lend their names to the Gaussian Unitary Ensemble or GUE, $\mathcal{E}$. In fact, in this work, we will not be much concerned with the potential, as it only influences the distribution of energies of the Hamiltonian. It has been treated here for sake of completeness. The interested reader may look into the saddle point method as one way of reducing the task of integrating over this potential [32].

We now turn our attention to the most subtle component: the measure, $dH$. This measure simultaneously tracks the $\frac{1}{2}d(d+1)$ independent real and $\frac{1}{2}d(d-1)$ imaginary components of a Hermitian matrix satisfying $H^\dagger = H$. It turns out that a good model to reproduce physical behavior is to treat each of these $\frac{1}{2}d(d+1) + \frac{1}{2}d(d-1) = d^2$ variables as real independent dimensions in Euclidean space $\mathbb{R}^{d^2}$, and see to what this means when they are arranged into a matrix. This is called the Lebesgue measure:

$$dH = \prod_{j=1}^{d} dH_{jj} \prod_{j<k} d\text{Re}H_{jk} d\text{Im}H_{jk}$$ (2.41)

In order to recover our basis invariance, we wish to change to the coordinates in equation 2.39 and find a form:
\[ dH = d(VAV^\dagger) = J(\Lambda)d\Lambda dV \quad (2.42) \]

Where the Lebesgue measure splits into a product form of a measure on the eigenvalues \( d\Lambda \) and an (invariant) measure \( dV \) on the eigenvectors. A coordinate transformation is accompanied by a Jacobian, which here is explicitly independent of \( V \) by assumption. The only choice for \( dV \) is the Haar measure, the unique uniform measure on a compact group, such as \( U(d) \), invariant under the action of the group. It satisfies \( \int_{U(d)} dV = 1 \), thus is a probability measure, and \( d(WV) = d(VW) = dV \), \( \forall W \in U(d) \) [33]. In order to find \( J(\Lambda) \), we equate 2.41 and 2.42. From the latter:

\[ d(VAV^\dagger) = dV\Lambda V^\dagger + Vd\Lambda V^\dagger + V\Lambda dV^\dagger \quad (2.43) \]

As we assume the measure \( dH \) is independent of \( V \), we may calculate it around \( V = \mathbb{1} \). This dramatically simplifies the equality. We also employ:

\[ 0 = d\mathbb{1} = d(V^\dagger V) = dV^\dagger V + V^\dagger dV \Leftrightarrow dV^\dagger = -V^\dagger dVV^\dagger \quad (2.44) \]

Where again for \( V = \mathbb{1} \), \( dV^\dagger = -dV \). Then:

\[ dH = dV\Lambda \mathbb{1}^\dagger + \mathbb{1}d\Lambda \mathbb{1}^\dagger + \mathbb{1}\Lambda dV^\dagger = d\Lambda + [dV, \Lambda] \quad (2.45) \]

From this expression, it is straightforward to read off the components \( dH_{jk} \). On the diagonal:

\[ \text{diag}(dH) = d\Lambda \Leftrightarrow dH_{jj} = dE_j \quad (2.46) \]

Off the diagonal:

\[ dH - \text{diag}(dH) = [dV, \Lambda] \Leftrightarrow dH_{jk} = (E_j - E_k)dV_{jk} \quad (2.47) \]

In expression 2.41, the imaginary and real parts multiply each other, resulting in:

\[ dH = \prod_{j=1}^{d} dE_j \prod_{j<k}^{d} (E_j - E_k)^2 d\text{Re}V_{jk} d\text{Im}V_{jk} \quad (2.48) \]

We now recognize the terms in 2.42:

\[ d\Lambda = \prod_{j=1}^{d} dE_j, \quad J(\Lambda) = \prod_{j<k}^{d} (E_j - E_k)^2, \quad dV = \prod_{j<k}^{d} d\text{Re}V_{jk} d\text{Im}V_{jk} \quad (2.49) \]

The form of \( J(\Lambda) \) is the square of the so called Vandermonde determinant of the eigenvalues of \( H \). It causes eigenvalue repulsion, a mutual interaction of the energies on top of their global, independent distribution determined by the potential. This phenomenon often observed in strongly coupled and interacting
systems, such as atomic nuclei of high atom number. All eigenstates are sensitive to the shape of all other eigenstates, which begets so called Wigner-Dyson energy gap statistics, where the probability of finding a gap $\Delta E$ around zero vanishes: hence repulsion. Contrast this to Poisson statistics, symptomatic of integrable systems with many decoupled degrees of freedom. The composite, product structure of the energy eigenstates allows for many energy crossings, and the vanishing energy gaps are the most likely.

The moral of this section is that we may consider the distribution of the eigenvalues of $H$, scaling with $P(\Lambda)d(\Lambda)d\Lambda$, as an independent degree of freedom from the distribution of its eigenvectors, according to the Haar measure $dV$. For instance, we will later integrate over the unitary group, leaving the distribution over eigenvalues untouched. In this way, the GUE is like an ensemble of spheres. Equation 2.39 is then a decomposition into spherical coordinates, with an angular component $V$ and radial $\Lambda$.

### 2.6.1 Generating Unitary Matrices

A crucial step in later simulations will prove to be the ability to sample unitary matrices of arbitrary dimension $d$, uniformly according to the Haar measure, or from the Circular Unitary Ensemble [32]. An algorithm was explained by Ozols in [34], following [35].

The starting point is the so called Ginibre Ensemble, which is computationally available readily. One generates $d^2$ independent identically distributed standard normal complex random variables $\{z_{jk}\}$.

$$f(z_{jk}) = \frac{e^{-|z_{jk}|^2}}{\pi}$$ (2.50)

These in turn can be generated by treating $|z_{jk}|^2$ as an exponentially distributed random variable with parameter $\lambda = 1$. Then the complex phase is chosen uniformly at random from $[0, 2\pi]$, and $z_{jk}$ is decomposed using polar coordinates on $\mathbb{C}$.

Once these are obtained, one arranges them in a matrix $Z \in \mathbb{C}^{d \times d}$, which is distributed according to the Ginibre ensemble, of which the joint distribution is:

$$f_G(Z) = \prod_{j,k=1}^{d} f(z_{jk}) = \frac{1}{\pi^{d^2}} \exp \left( - \sum_{j,k=1}^{d} |z_{jk}|^2 \right) = \frac{1}{\pi^{d^2}} \exp \left( - \text{Tr}(Z^\dagger Z) \right)$$ (2.51)

It is immediately clear that this distribution is invariant under left -also right- action of the unitary group $V \in U(d)$:

$$f_G(VZ) = \frac{1}{\pi^{d^2}} \exp \left( - \text{Tr}((VZ)^\dagger(VZ)) \right) = f_G(Z)$$ (2.52)

In the ensemble, the subset of $Z$ that are singular is measure zero. This can be seen as follows. After generating any number $< d$ of columns, the chance to
generate another column inside their span is vanishingly small. Thus we may
assume $Z$ is invertible and admits a so called $QR$-decomposition.

$$Z = QR, \quad Q \in U(d), \quad R \in T'(n) \quad (2.53)$$

Here, $T'(n)$ is the set of set of invertible $d \times d$ upper-triangular complex ma-
trices. The QR decomposition of a matrix can be computed readily by standard
Gram-Schmidt orthogonalization algorithms. In general, it is not unique. For
any diagonal $D \in U(d),$

$$QR = QD \overline{D} R = (QD)(\overline{D} R) = Q'R' \quad (2.54)$$

is also a decomposition of $Z$. In order to guarantee uniqueness, we additional-
ly impose $R \in T(n)$, the specification of $T'(n)$ to have real positive diagonal
entries. This can be done by setting

$$D = \text{diag} \left( \frac{r_{11}}{|r_{11}|}, \frac{r_{22}}{|r_{22}|}, \ldots, \frac{r_{dd}}{|r_{dd}|} \right) \quad (2.55)$$

Then $R' = D^R R \in T(n)$ from 2.54 and the $Q' = QD$ is unique, and the
mapping $Z \mapsto (Q', R')$ is well defined.

It remains to be shown that our candidate $Q'$ is uniformly drawn according to
the Haar measure. But this is immediate from 2.52: $VZ = VQ'R' \mapsto (VQ', R'')$ is
equally likely to be generated as $Z = Q'R' \mapsto (Q', R')$, for any possible $V$.
As group action is closed, $U(d)Q' = U(d): \{ VQ' \}$ is the entire group. Each $Q'$
is then equally likely. A more rigorous explanation can be found in the source,
[34].

2.7 Rudimentary Group Theory

We shortly recall from [36] a number of rote applications of discrete group theory,
specifically pertaining to the symmetric group $S_q$, necessitated by chapter 4.

The symmetric group $S_q$, is the group of permutations on $q$ symbols. An
element of this group, $\sigma$, is a mapping from a set of $q$ distinct, ordered elements
$I := \{ i_1, i_2, \ldots, i_q \}$ to itself. This is symbolically denoted $S_q \ni \sigma : I \mapsto I$. There
are $q!$ such permutations, so that is the cardinality of the finite group.

The basic building blocks are cycles: $(a_1 a_2 \ldots a_r)$ is an $r$-cycle, with each
$a_j \in I$. This cycle corresponds to the mapping $a_j \mapsto a_{j+1}$ for $j = \{ 1, 2, \ldots, r-1 \}$, and the last $a_r \mapsto a_1$. The unused elements $I \setminus \{ a_j \}$ are mapped to themselves. Clearly, $r \leq q$.

For disjoint $\{ a_j \}$ and $\{ b_j \}$, the cycles $(a_1 a_2 \ldots a_r)$ and $(b_1 b_2 \ldots b_s)$ can be
composed, or multiplied -the group action- to $(a_1 a_2 \ldots a_r)(b_1 b_2 \ldots b_s)$. This is
simply the mapping defined by carrying out both constituent mappings, and the
multiplication is commutative. By contrast, for $\{ a_j \} \cap \{ b_j \} \neq \emptyset$, the mul-
tiplication is consecutive application of the cycles, starting with the rightmost
written one. It is in general not commutative. For all $a_j = b_k$, the total map-
ing is $b_{k-1} \mapsto b_k = a_j \mapsto a_{j+1}$ or $b_{k-1} \mapsto a_{j+1}$. This generalizes readily to
elements $\sigma, \tau \in S_q$ which are themselves composed of multiple cycles. They are composed as $\sigma \tau \in S_q$. Any element can be reduced to a representation of a product of disjunct cycles: in which each symbol only features once among all the constituent cycles. Moreover, this representation is unique. After all, each symbol is only mapped to and from once.

There is a unique element $\text{Id} \in S_q$, the identity, which takes all symbols in $I$ to themselves, and $\sigma \text{Id} = \text{Id} \sigma = \sigma \forall \sigma \in S_q$. Furthermore, any element $\sigma$ has a unique inverse $\sigma^{-1}$, satisfying $\sigma \sigma^{-1} = \sigma^{-1} \sigma = \text{Id}$. For an isolated cycle, $(a_1 a_2 \ldots a_r)$, the inverse is $(a_r a_{r-1} \ldots a_1)$, as it takes $a_j \mapsto a_{j+1} \mapsto a_j$, and similarly for $a_r$. Disjunct cycles commute, so this generalizes to the inverse of composite elements in cycle representation, by inverting the constituent cycles.

Once a representation of $\sigma \in S_q$ has been reduced to a product of disjunct cycles $\sigma = (a_{1,1} a_{1,2} \ldots a_{1,r_1})(a_{2,1} a_{2,2} \ldots a_{2,r_2}) \ldots (a_{s,1} a_{s,2} \ldots a_{s,r_s})$, we say it is in the conjugacy class $r_1, r_2, \ldots, r_s$. This is because any conjugation $\tau \sigma \tau^{-1}$ consists of cycles of the same length, and any such cycle in the class can be obtained from any other for appropriate choice of $\tau \in S_q$.

We are finally equipped to derive the results in the main body of this work.
Chapter 3

Numerical Fidelity Simulations

We now move to the first results of this thesis, numerical calculations. They will aid us in getting a grasp of the behavior of quantum systems. Following the setup of section 2.2, we imagine a number of degrees of freedom for the subsystem S and bath B. Numerics call for concreteness, so for the remainder of this section we set the building block of all discussed Hilbert spaces to be qubits, or spin $\frac{1}{2}$-particles. This is customary in many quantum-information contexts, as the qubit is the smallest unit of quantum information. Also it gives the Hilbert space the maximally composite structure, maximizing the possible amount of tracing relative to total dimension. This is useful because simulations on classical computers of high-dimensional quantum systems quickly become hard to handle. What follows may be redesigned with any Hilbert space, so long as its dimension isn’t prime. We assume the predictions generalize to systems that are not composed of qubits, but e.g. qudits: particles of larger spin, or any other composite discrete Hilbert space for that matter.

The aim of this chapter is to observe thermalization, or at least equilibration, in most quantum (sub)systems, determined by their driving full system Hamiltonian. The most important statistic will be the fidelity, defined in 2.30. As mentioned, the fidelity quantifies measurable distinguishability between states. If a subsystem tangibly moves away from its initial state, the fidelity between the initial state and the later-time state must be less than unity. Conversely, if the subsystem reaches a steady state after some time, the fidelity between the subsystem at two times after this point should approach unity. The strategy is thus to generate a random Hamiltonian, check the fidelity between different times, and repeat. With enough such Hamiltonians, a trend is expected to emerge.
3.1 Example Systems

In order to obtain insight, an example of fidelity over time is plotted in the figures 3.2 and 3.3. For a number of different sized baths ($n_B = 4, 5$ or $6$ spins) and subsystems ($n_S = 1$ or $2$ spins) we have generated one interacting Hamiltonian with eigenvectors $V$ from the Haar measure, and eigenvalues according to a flat individual distribution on $[-\pi, \pi]$, but following the mutual repulsion of the Vandermonde determinant in equation 2.49. These systems start evolution in a product state, so the subsystem is pure, and allowed to evolve according to the Schrödinger equation. All units are such that $\hbar = 1$. A schematic of the setup is included in figure 3.1.

What function do we display exactly? $F$ takes two arguments, so we have chosen one of them fixed, in essence a slice of the full picture, in order to have a 2D graph. The expressions $F_0 := F(\rho_S(0), \rho_S(t))$ and $F_1 := F(\rho_S(1), \rho_S(t))$ are plotted over $t$. Of course, the time $t = 1$ is arbitrary, but empirically for the energies used, it is enough to be considerably far from the initial state, corroborated by figure 3.2.

It is already apparent in figure 3.2 that the larger the subsystem, the more it can deviate from the initial state. There is, in a sense, more 'room' to get away, as the Hilbert space is larger. Next, compare with a not-initial state in figure 3.3.

As expected this figure has value one at $t = 1$, the fidelity between a state and itself is maximal. After this time, the value hovers close to the maximum: the subsystem doesn’t deviate far from its value at time 1. By the same token, we can plot the time dependent entropy $S$ (from 2.24) and purity $\gamma$ (from 2.27) for similar systems. See the figures 3.5 and 3.4.

Pursuant to theory, the entanglement entropy of a pure state is 0. After this, the entropy plateaus around a value correlated to the system size. In fact, the maximum entropy for a $n_S$-spin subsystem is $\ln(d_S) = \ln(2^{n_S}) = n_S \cdot \ln 2$, so it appears these systems quickly maximize disorder. This conclusion is echoed by the purity.

Note again that the minimum purity, $1/d_S = n_S^{-2}$ corresponds to maximum entropy. As a test of theory, random noninteracting Hamiltonians $H = H_S \otimes 1_B + 1_S \otimes H_B$, were constructed and used to evolve the system (not plotted).
Figure 3.2: Example: $F(\rho_S(0),\rho_S(t))$ plotted over time $t$, for 6 random spin systems, consisting of different size subsystem and bath. Units are such that $\hbar = 1$.

Figure 3.3: Example: $F(\rho_S(1),\rho_S(t))$ plotted over time $t$, for 6 random spin systems. At $t = 1$, the fidelity is taken between the state and itself, and is thus exactly unit for all systems.
Figure 3.4: Example: $S(ρ_S(t))$ plotted over time $t$, for 6 random spin systems. Initially in a product state, the subsystem entropy is zero. It quickly increases, approaching a value dependent on $n_S$.

Figure 3.5: Example: $\gamma(ρ_S(t))$ plotted over time $t$, for 6 random spin systems. Initially in a product state, the subsystem purity is 1. It quickly moves away, approaching a value dependent on $n_S$. 
Here entropy remained zero for all times, as one would expect.

A final note: each of the figures 3.2, 3.3, 3.4, 3.5 were created with different sets of random Hamiltonians, they are not intended to be compared directly.

3.2 Statistics of Fidelity

Let us now be more specific and rigorous. The numerical experiments detailed below have a common structure. The system lives in the variable \( n \)-spin Hilbert space \( \mathcal{H}_S \otimes \mathcal{H}_B \). A Hamiltonian is generated, with eigenvectors \( V \) uniformly drawn from the Haar measure, and eigenvalues (energies) arbitrary. We start, for definiteness, in the product state \( |1⟩ = |1_S⟩ \otimes |1_B⟩ \) of the computational basis in 2.5. This may be thought of as all spins pointing up, as in figure 3.1. It is conventionally the first basis vector, and is a product state as it can be written trivially as \( |(↑↑...)_S⟩ \otimes |(↑↑...)_B⟩ \), using the all up states of the components. The composite system is evolved forward to other times \( |ψ(t)⟩ = \hat{U} |1⟩ \) according to the Schrödinger equation, 2.1. When we wish to evaluate the fidelity on the subsystem, the bath is traced out from the full system pure state, as in 2.17:

\[
\rho_S(t) = \text{Tr}_B (|ψ(t)⟩⟨ψ(t)|)
\] (3.1)

Done for two different times, this may be used as arguments in definition 2.30. We will construct two main statistics of the fidelity. For the first, consider the expression:

\[
⟨F_0⟩ := \lim_{T \to \infty} \frac{1}{T} \int_0^T dt F(\rho_S(0), \rho_S(t))
\] (3.2)

This is the all-time \( t \) average of the fidelity of the subsystem between time 0, and itself at time \( t \). It tracks how far the subsystem deviates from its starting state. As mentioned before and forecast by figure 3.2, we expect this value to be less than one for almost any system. It is clear from the construction that \( \rho_S(0) = |1_S⟩⟨1_S| \), a pure state. Then equation 2.31 tells us the \( F_0 \) is simply:

\[
F_0 \equiv F(\rho_S(0), \rho_S(t)) = \sqrt{⟨1_S | \rho_S(t) | 1_S⟩} = \sqrt{\rho_S(t)_{1,1}}
\] (3.3)

In index notation, it is simply the square root of the first diagonal element. This can be substituted into 3.2 to expedite the computation. The second interesting formula is:

\[
⟨F_∞⟩ := \lim_{T \to \infty} \frac{1}{T^2} \int_0^T dt \int_0^T du F(\rho_S(u), \rho_S(t))
\] (3.4)

It is the all time average over \( t \) and \( u \) of the fidelity between the subsystem at times \( t \) and \( u \). This relates to the degree of equilibration, how well we can distinguish the state at two later times. By our reasoning and evaluation of figure 3.3 we expect this to approach one for almost any system, implying thermalization.
Analogous to the calculation of \( \langle F_0 \rangle \), one can average the normalized entanglement entropy, defined in 2.26, over time:

\[
\langle S_1 \rangle := \lim_{T \to \infty} \frac{1}{T} \int_0^T dt S_1 (\rho_S(t)) \tag{3.5}
\]

In the simulation, the choice was made to generate a unitary matrix \( V \) to represent the eigenvectors of \( H \) as in subsection 2.6.1. Concerning the eigenvalues \( \Lambda \): time always multiplies energy in the time-dependent Schrödinger equation. Owing to the nature of time in expressions 3.2 and 3.4 as a dummy variable -it is always integrated out- the results should be energy independent. We only formally preclude any degenerate energy gaps, i.e. \( E_h - E_j = E_k - E_l \iff (h,j) = (k,l) \), which otherwise would cause resonance and amplified Poincaré recurrence. In the general case without symmetries, due to the fully coupled nature of the world, this is a physically reasonable assumption. Then \( \langle F_0 \rangle \) and \( \langle F_\infty \rangle \) are essentially only functions of the eigenvectors of the Hamiltonian, the energy eigenstates, arranged as the columns of \( V \). Under this rationale, we simplified the computational burden by sampling the energies independently at random from a uniform \([-\pi, \pi]\) distribution. This does not reflect the interdependence caught by the RMT-prescribed Vandermonde determinant in section 2.6, but that should not systematically affect the final estimates.

### 3.2.1 Experiment Parameters

A sequence of simulations were run, varying the amount \( n_B \) of bath, and \( n_S \) of subsystem spins. Hilbert spaces consisting of \( n = n_B + n_S \) equals 3, 4, 5, 6, 7 and 8 total spins, subsequently divided into a subsystem of \( n_S \) equalling 1 or 2 and a bath comprising the rest were considered. For each of these 12 situations, 4000 random Hamiltonians were generated and on each Hamiltonian, the following quantities were approximated by a Monte Carlo integral over time:

- \( \langle F_0 \rangle \)
- \( \langle F_\infty \rangle \)
- The temporal variance in \( F_\infty \): \( \sigma_\infty^2 \)
- Time average entanglement entropy \( \langle S_1 \rangle \)

We might also have included \( \sigma_0^2 \), the variance in \( F_0 \), but it did not prove very insightful on first estimates and was very similar to \( \sigma_\infty^2 \).

Sampling 150 random times uniformly from \( t \in [0, 1000] \) constituted an ordered dataset of density matrices for one Hamiltonian. Calculating \( F_0 \) as in 3.3 for each of these times, and taking the average over this vector, served as a proxy for \( \langle F_0 \rangle \). Similarly, calculating the fidelity between each subsequent density matrix in the set, and averaging that, is how we achieved an estimate of \( \langle F_\infty \rangle \). Taking the variance of the vector resulted in \( \sigma_\infty^2 \). Though this may not be the best, unbiased estimator for the actual variance, it gives ample sense of

29
Figure 3.6: Distribution of 4000 random Hamiltonians: their time average fidelity to the initial pure state, $\langle F_0 \rangle$, for a system of $n = 3, 4, 5$ spins, and $n_S = 2$ spins. The surface under each plot is normalized to one.

the spread of $F_\infty$ over time. Lastly, the entropy $\langle S \rangle$ was obtained analogously to $\langle F_0 \rangle$, collecting all of these averages formed a single data point. These data points are displayed tallied into selected histograms in the rest of the chapter. More may be found in the appendix.

### 3.3 Results

This section will predominantly consist of figures accompanied by clarification. The horizontal scale of the distributions of fidelity is heavily dependent on the amount of total spins. For a ‘large system’ (6, 7 or 8 total spins $n$), the spread is found to be an order of magnitude smaller than with a ‘small system’, ($n$ of 3, 4 or 5). For clarity, the two cases will be plotted separately, allowing natural axes.

The first result is $\langle F_0 \rangle$ (equation 3.2) for a small system of which two spins are taken to constitute the subsystem. See figure 3.6. The rest of the two-spin subsystem cases, as part of the larger system, is plotted in 3.7.

The analogous plots, using $n_S = 1$, are displaced to appendix A, figures A.1, A.2. It is clear from these four figures that $\langle F_0 \rangle$ tends to converge on a value depending on $n_S$, and the degree to which the distribution clusters increases with bath size. For $n_s = 2$, $\langle F_0 \rangle \xrightarrow{n_B \to \infty} \frac{1}{2}$, for $n_s = 1$, $\langle F_0 \rangle \xrightarrow{n_B \to \infty} \frac{1}{\sqrt{2}}$. This is consistent with the following supposition. The state $\rho_S(t \neq 0)$ is generally close to the maximally mixed state $1_S/d_S = 1_S \cdot 2^{-n_S}$. The larger the bath, the closer.
Figure 3.7: Distribution of 4000 random Hamiltonians: their time average fidelity to the initial pure state, $\langle F_0 \rangle$, for a system of $n = 6, 7, 8$ spins, and $n_S = 2$ spins. The surface under each plot is normalized to one.

As statistically more information is moved to the bath, the uncertainty on the subsystem grows, capping off at $I_S/d_S$. Using formula 3.3, and observing that the first (indeed any) diagonal element of the maximally mixed state is $2^{-n_S}$, we have:

$$\lim_{n_B \to \infty} \langle F_0 \rangle = \sqrt{2^{-n_S}} \quad (3.6)$$

This confirms the prediction voiced in the previous section.

We move on to the next quantity: the distribution of $\langle F_\infty \rangle$ (equation 3.4). The case of a two-spin subsystem is plotted in figures 3.8 and 3.9. For the one-spin subsystem, see appendix A, figures A.3 and A.4.

From these figures, it is clear that for a random system, the average fidelity between different times does indeed have a large chance to be close to one. Moreover, the clustering of the distribution increases with bath size, and with the ratio of bath to system. This is a confirmation of the forecast that almost all systems, for close to any initial conditions, will become indistinguishable in time: a case for thermalization being the norm. The initial state will likely comprise many energy eigenstates and will be mixed randomly from the perspective of any subsystem. Conversely, states of the full system for which a small subsystem is out of this 'equilibrium' must be rare. In the field of quantum dissipation, in fact an infinite bath size is necessary to observe thermalization. Caldeira and Leggett first formulated a bath consisting of an infinite number of harmonic oscillators. With these infinite degrees of freedom, Poincaré recurrence, or the coalescing
Figure 3.8: Distribution of 4000 random Hamiltonians: their time average fidelity between all times, $\langle F_\infty \rangle$, for a system of $n = 3, 4, 5$ spins, and $n_S = 2$ spins. The surface under each plot is normalized to one.

Figure 3.9: Distribution of 4000 random Hamiltonians: their time average fidelity between all times, $\langle F_\infty \rangle$, for a system of $n = 6, 7, 8$ spins, and $n_S = 2$ spins. The surface under each plot is normalized to one.
of the phases of the system to eventually reproduce the initial conditions, is postponed ad infinitum [37].

Supplementing this result is the temporal variance $\sigma^2_\infty$. In figure 3.10, it is seen to be very small for a large system with a two-spin subsystem. The small system variant is in appendix A, figure A.5. The small variance means not only is the average fidelity close to one, it also stays in the neighborhood of one for most times. Moreover, there is a theoretical argument making this the only possibility. $\langle F_\infty \rangle$ lies near one, the maximum, for most systems. Using Markov’s inequality on $1 - F \geq 0$, it is inescapable that the fidelity must be close to one at nearly all times. A large portion of time with low fidelity could not be compensated in this average.

These graphs can be somewhat opaque. The state of the system at different times ‘looks’ the same, resembling some steady state. Although hypothesizing above about the maximally mixed state, we do not know for certain what state it is. The distribution of the time-averaged entropy sheds light on this definitively. See figure 3.11 for the time average normalized entanglement entropy $\langle S_1 \rangle$ from 3.5, for a subsystem of $n_S = 2$ spins, and all systems $n$. The case of $n_S = 1$ is removed to to appendix A, figure A.6

This confirms our suspicion induced by figure 3.4 that the average subsystem moves toward the maximally mixed state, the state that maximizes entropy. The degree to which it adheres scales with the ratio of the bath to the subsystem. I.e. a larger bath increases the probability to find the subsystem close to $\mathbb{I}_S$ for a random coupled time evolution. This effect may be called quantum decoherence:

![Figure 3.10: Distribution of 4000 random Hamiltonians: their temporal variance of the fidelity between different times, $\sigma^2_\infty$, for a system of $n = 6, 7, 8$ spins, and $n_S = 2$ spins. The surface under each plot is normalized to one.](image)
due to coupling with the environment, the information of the mutual phase between subsystem states is lost [27].

It is interesting to note that the peaks of the distributions in figures 3.11 and A.6 agree neatly with an analytic result by Sen [38]. Hilbert space, being a projective space, is compact. It is possible to draw a vector from it uniformly at random according to the Haar measure [14]. If one does this for a composite \( d = d_S \cdot d_B \) dimensional system, and traces out the \( d_B \) degrees of freedom, the resulting density matrix will on average have the entanglement entropy given by:

\[
\langle S(\rho_S) \rangle_{d_s, d_B} = \left( \sum_{j=d_B+1}^{d_s} \frac{1}{j} \right) - \frac{d_S - 1}{2d_B} = \Psi(d+1) - \Psi(d_B+1) - \frac{d_S - 1}{2d_B} \quad (3.7)
\]

Here, \( \Psi(z) \) is the digamma function. The normalized (equation 2.26) values of this average entropy \( \langle S_1(\rho_S) \rangle_{d_s, d_B} \) corresponding to the system dimensions in figures 3.11 and A.6 are tabulated in Table 1. We learn that for most times and systems, the state \( |\psi(t)\rangle \) is essentially uniformly random in \( \mathcal{H} \), as far as entanglement entropy is concerned.
<table>
<thead>
<tr>
<th>$n_S$</th>
<th>$n = 3$</th>
<th>$n = 4$</th>
<th>$n = 5$</th>
<th>$n = 5$</th>
<th>$n = 6$</th>
<th>$n = 7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.74</td>
<td>0.87</td>
<td>0.93</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>2</td>
<td>0.34</td>
<td>0.67</td>
<td>0.83</td>
<td>0.92</td>
<td>0.96</td>
<td>0.98</td>
</tr>
</tbody>
</table>

The results above results agree qualitatively, in particular, with less geometric calculations such as those by Reimann [10]. However, it is difficult to make quantitative statements, due to the arbitrary magnitude of constants like the macroscopic measurement accuracy, which has no real analog in the language of this work. Also [11] draws similar conclusions, both in results and in philosophy of the nature of our ignorance.

The simulations above were carried out in Python. The main scripts, accompanied by usage instructions, are to be found in appendix B.
Chapter 4

Analytic Mixed State Averages

In this chapter, we will use geometrical, analytic arguments to quantify thermalization of the typical quantum system.

4.1 Analytic statistics of local density matrices

The physical setup is the same as in the previous chapters. We construct a Hamiltonian with unspecified eigenvalues \( \{E_i\}, \ i \in \{1, 2, \ldots, d\} \), and eigenvectors defined by the unitary matrix \( V \), drawn from the Haar measure on the unitary group \( U(d) \). Considering qubits again, one may take \( d = 2^{(n_S + n_B)} = \dim \mathcal{H} \), as it represents the dimension of \( n_S \) system qubits (spins) and \( n_B \) bath qubits, which will be coupled together by the Hamiltonian. Then \( d_B = 2^{n_B}, d_S = 2^{n_S} \) are the dimensions of the component Hilbert spaces. However, these results are more general, and will hold for any \( d_S, d_B \) sized component systems totalling \( d = d_S \cdot d_B \) full system dimensions. In a diagonalized form, from section 2.6:

\[
H = V \Lambda V^\dagger, \quad V \in U(d), \quad \Lambda = \text{diag}(E_1, E_2, \ldots, E_d)
\]  

(4.1)

Then, following the standard construction, we have \( H \) evolve the coupled system and bath, starting in a product state \( |1\rangle = |1_S\rangle \otimes |1_B\rangle \) at \( t = 0 \). This constitutes one of our assumptions, that there is no initial entanglement between the artificially partitioned \( S \) and \( B \). We have chosen the basis of the full Hilbert space \( \mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B \) such that our (random) initial state is indeed the first basis vector, in turn defined to be the tensor product of the component first basis vectors. The full basis is \( \{|k\rangle\} \). We may always do this by means of local unitary basis transformations, which can then be absorbed into \( V \). By default:

\[
|\psi(t)\rangle = e^{-iHt} |1\rangle, \quad \rho_S(t) = \text{Tr}_B(|\psi(t)\rangle \langle \psi(t)|)
\]

(4.2)
We cannot at present perform the integrals from chapter 3, for instance the fidelity between $\rho$ at different times, due to the diagonalization necessary in the definition of the matrix root. There are identities to integrate over the unitary group, as long as the integrand is of a certain symmetric form, as in the Harish-Chandra-Itzykson-Zuber-formula [39]. Then we must ask slightly different questions.

4.1.1 Average Reduced Density Matrix

We will take a different route from the previous, proving a related result. We aim to find the elements of $\rho_S(t)$, averaged uniformly over the possible $V$. We start by explicitly writing the components of $|\psi(t)\rangle := \sum_k \psi_k(t) |k\rangle$, which are found by

$$\psi_k(t) = \langle k| V e^{-i\Lambda t} V^\dagger |1\rangle$$

The components of the full density matrix $\rho(t) = |\psi(t)\rangle \langle \psi(t)|$ are given by

$$\rho_{kl}(t) = \langle k| V e^{-i\Lambda t} V^\dagger |l\rangle$$

Now it must be observed that each index is actually a multi-index $j \simeq (j_S, j_B)$, owing to the tensor product structure of the Hilbert space. By formula 2.4, the full basis is formed by combinations of the component bases. In order to calculate the reduced density matrix, one must contract only the index of the bath from 1 to $d_B$. We will make the multi-indexed nature explicit only where needed, on the outer sides of $\rho(t)$. In components:

$$\rho_S(t)_{k_S g_S} = \sum_{k_B=1}^{d_B} \sum_{j,m=1}^d V_{k_S,k_B} e^{-iE_j t} V^\dagger_{j,m} e^{iE_m t} V^\dagger_{m(g_S,k_B)}$$

This expression reveals that the time and energy and time dependence is just a multiplicative factor on each term of the sum. We invoke an identity applicable using Weingarten functions, following the techniques of Benoit Collins [40]. Products of individual elements of $V$ may be integrated over $U(d)$. Without going into too much detail of the derivation of this identity, which involves representation theory, we will simply state the result:

$$\int_{U(d)} dV V_{i_1,j_1} \ldots V_{i_q,j_q} V^\dagger_{j_1',i_1'} \ldots V^\dagger_{j_q',i_q'} =$$

$$\sum_{\sigma,\tau \in S_q} \delta_{i_1,i_1'} \ldots \delta_{i_q,i_q'} \delta_{i_1,j_1'} \ldots \delta_{i_q,j_q'} W g(d, \sigma \tau^{-1})$$

In this expression, $S_q$ is the symmetric group on $q$ symbols, and $W g(d, \sigma)$ is on of a fixed set of Weingarten functions on $d$ dimensions and permutation.
\(\sigma\). Which particular function of the set is used depends only on the conjugacy class (cycle lengths) of \(\sigma\). All are quotients of polynomials in \(d\). Thankfully, they have been tabulated by Brouwer and Beenakker for arbitrary \(d\) and for permutations up to \(q = 5\), in [41], so all need be done is to sum over \(\sigma, \tau\) in the symmetric group, and multiply the Kronecker deltas by the Weingarten function corresponding to the class of \(\sigma \tau^{-1}\). For the usage of permutation cycles, conjugation, inverses, and classes we refer to section 2.7, and for more background, to any introductory text on group theory, e.g. [36].

In general, not all delta functions will be satisfied, and they will kill many terms in the sum of 4.5. In our case, for the average \(\rho_S(t)\), it is clear that \(q = 2\), so the possible permutations are \(\tau, \sigma \in \{\text{Id}, (12)\}\). We can also populate the indices \(i_1, i_2, j_1, j_2, i'_1, i'_2, j'_1, j'_2\) in 4.6. Reordering the scalar, indexed terms, our expression now takes the form:

\[
\int_{U(d)} dV \rho_S(t) = \sum_{\sigma, \tau \in S_2} R_{\sigma, \tau} \quad (4.7)
\]

Where we imply the following index equalities:

\[
(i_1, i_2) = ((k_S, k_B), 1); \quad (i'_1, i'_2) = (1, (g_S, k_B))
\]
\[
(j_1, j_2) = (j, m); \quad (j'_1, j'_2) = (j, m) \quad (4.8)
\]

There are four combinations of \(\sigma, \tau\) in the sum in 4.7. We may write:

\[
\int_{U(d)} dV \rho_S(t) = \sum_{\sigma, \tau \in S_2} R_{\sigma, \tau} \quad (4.9)
\]

The \(R_{\sigma, \tau}\)’s involve two contributing Weingarten functions for \(S_2\). These are, symbolically dependent on cycle length as the 2nd argument:

\[
Wg(d, 1^2) = \frac{1}{d^2 - 1}, \quad Wg(d, 2) = \frac{-1}{d(d^2 - 1)} \quad (4.10)
\]

For \(\sigma = \tau = \text{Id}, \sigma \tau^{-1} \sim 1^2\), the contribution \(R_{\sigma, \tau}\) becomes:

\[
R_{\text{Id}, \text{Id}} = \sum_{k_B=1}^{d_B} \sum_{j, m=1}^{d} \delta_{k_B}(k_S, k_B) \delta_{j, m} Wg(d, 1^2) e^{i(E_m - E_j)t} \quad (4.11)
\]

\[
= \frac{\chi(t)}{d^2 - 1} \delta_{k_S, 1} \delta_{g_S, 1}
\]
The first deltas kill all terms over $k_B$ except the first, as only multi-index $(k_S, k_B) = (1, 1) = 1$ as a single index. Here we have also defined function $\chi(t)$ to carry the time and energy dependence:

$$\chi(t) := \sum_{j, m=1}^d e^{i(E_m - E_j)t} = d + \sum_{j<m} 2\cos((E_m - E_j)t)$$  \hspace{1cm} (4.12)

Observe that $\chi(t = 0) = d^2$ and empirically, the function quickly drops to its time average $d$, in time $t = O(\sigma_E^{-1})$, around the inverse standard deviation in energies. See the figures 4.1 and that beside it, for an example.

The $\chi(t)$ plotted are constructed with energies uniform in $[-\pi, \pi]$, but generated by the techniques of RMT, thus with mutual energy repulsion. For a small number of dimensions, i.e. $d = 5$, the harmonic character of $\chi(t)$ is still prevalent, Poincaré recurrence is rapid, and the function is erratic. Increasing the number to $d = 100$, we see that $\chi(t)$ is more predictable, and seems ‘damped’ due to the strong eventual decoherence of the phases, causing destructive interference. This $d$ may seem like a large number, but 7 qubits already have a $2^7 = 128$-dimensional Hilbert space. These are not experimentally intractable systems. From some (admittedly subjective) trials, it appears the neat decaying behavior depends on the eigenvalue-repulsion, under which rapidly oscillating cosines are suppressed. For independent random energies from the same support, the damped character of the function is less pronounced, though to make this suspicion rigorous, more study is needed.

The next term is for $(\sigma, \tau) = (Id, (12)), \sigma \tau^{-1} \sim 2$:

$$R_{Id,(12)} = \sum_{k_B=1}^{d_B} \sum_{j, m=1}^d \delta_{(k_S, k_B), 1}\delta_{1,(g_S, k_B)}\delta_{j, m} \delta_{m, j} W(g(d, 2) e^{i(E_m - E_j)t})$$

$$= -\frac{d}{d(d^2 - 1)} \delta_{k_S, 1} \delta_{g_S, 1}$$  \hspace{1cm} (4.13)
Here, \( m = j \) is forced by the delta function, leaving the exponent equal to unity for all times. Moving on, \((\sigma, \tau) = ((12), Id)\):

\[
R_{(12),Id} = \sum_{k_B=1}^{d_B} \sum_{j,m=1}^{d} \delta_{(k_B),(g_S,k_B)} \delta_{j,1} \delta_{m,m} W g(d,2) e^{i(E_m - E_j) t}
\]

\[
= -\frac{d_B \chi(t)}{d(d^2 - 1)} \delta_{k_S,g_S}
\]

(4.14)

Here, the delta in the multi-indices is satisfied for every dummy \( k_B \), contributing a factor \( d_B \), but only on diagonal entries in the system’s density matrix \( k_S = g_S \). The last combination, \((\sigma, \tau) = ((12), (12))\), \( \sigma \tau^{-1} \sim 1^2 \):

\[
R_{(12),(12)} = \sum_{k_B=1}^{d_B} \sum_{j,m=1}^{d} \delta_{(k_B),(g_S,k_B)} \delta_{j,1} \delta_{m,j} W g(d,1^2) e^{i(E_m - E_j) t}
\]

\[
= \frac{d_B \cdot d}{d^2 - 1} \delta_{k_S,g_S}
\]

(4.15)

Summing the four terms \( R_{\bullet, \bullet} \) in 4.9 we obtain the full integral.

\[
\hat{\rho}_S(t) := \int_{U(d)} dV \rho_S(t) = \left( \frac{\chi(t)}{d^2 - 1} + \frac{-d}{d(d^2 - 1)} \right) \delta_{k_S,1} \delta_{g_S,1} + \left( \frac{-d_B \chi(t)}{d(d^2 - 1)} + \frac{d_B d}{d^2 - 1} \right) \delta_{k_S,g_S} = \left( \frac{\chi(t) - 1}{d^2 - 1} \right) |1_S\rangle \langle 1_S| + \left( \frac{d_B(d^2 - \chi(t))}{d(d^2 - 1)} \right) 1_S
\]

(4.16)

Elements off the main diagonal are zero, and on the diagonal, the first element \((k_S, g_S) = (1, 1)\) is unique.

First a consistency check: The integrand \( \hat{\rho}_S \) has trace 1 at all times, and the measure is normalized such that \( \int_{U(d)} dV = 1 \). Furthermore integrating and tracing are linear and must commute. So

\[
1 = \int_{U(d)} 1 dV = \int_{U(d)} \text{Tr}(\hat{\rho}_S) dV = \text{tr}(\int_{U(d)} \hat{\rho}_S dV) = \text{Tr}(\hat{\rho}(t)).
\]

Let us verify:

\[
\text{Tr}(\hat{\rho}_S(t)) = \left( \frac{\chi(t) - 1}{d^2 - 1} \right) + d_S \left( \frac{d_B(d^2 - \chi(t))}{d(d^2 - 1)} \right) = \frac{\chi(t) - 1 + d^2 - \chi(t)}{d^2 - 1} = 1
\]

(4.17)

Indeed, using the fact that \( d_S \cdot d_B = d \). Next, we consider the behavior for \( t = 0 \), or averaged over all time. As mentioned above, \( \chi(0) = d^2 \). Substituting this in 4.16, the diagonal terms vanish, and the first element term becomes unit, as we would expect: at time \( t = 0 \), no matter the Hamiltonian, the unevolved system is pure, in the first basis state, \( \hat{\rho}_S(0) = \rho_S(0) = |1_S\rangle \langle 1_S| \). Conversely,
averaged over time, the cosines in 4.12 vanish, and \( \langle \chi \rangle_t = d \). As \( \bar{\rho}_S(t) \) is linear in \( \chi(t) \), the time average \( \langle \bar{\rho}_S(t) \rangle_t \) is obtained by substituting \( \chi(t) \mapsto d \).

\[
\langle \bar{\rho}_S(t) \rangle_t = \left( \begin{array}{c} \frac{d - 1}{d^2 - 1} \end{array} \right) |1_S \rangle \langle 1_S| + \left( \begin{array}{c} \frac{d_B(d^2 - d)}{d(d^2 - 1)} \end{array} \right) 1_S = \left( \begin{array}{c} \frac{1}{d + 1} \end{array} \right) |1_S \rangle \langle 1_S| + \left( \begin{array}{c} \frac{d_B}{d + 1} \end{array} \right) 1_S
\]

(4.18)

Which is close to the maximally mixed state with equal entries on each diagonal. We conjecture that \( \chi(t) \) spends most time close to its time average. Then information of the initial conditions persist in the anomalous extra occupation of the first element forever. One may think of a phase transition of sorts between \( t = 0 \) and \( t \to \infty \), moving from a 'known' to 'unknown' phase, with the shift occurring around \( \sigma_E^{-1} \).

From here interestingly, using \( d = d_B \cdot d_S \), it is evident that taking \( d_B \to \infty \), leaving \( d_S \) fixed, we indeed obtain the maximally mixed state for the time and system averaged density matrix.

\[
\lim_{d_B \to \infty} \langle \bar{\rho}_S(t) \rangle_t = \lim_{d_B \to \infty} \left[ \left( \begin{array}{c} \frac{1}{d_B d_S + 1} \end{array} \right) |1_S \rangle \langle 1_S| + \left( \begin{array}{c} \frac{d_B}{d_B d_S + 1} \end{array} \right) 1_S \right] = \frac{1}{d_S} 1_S
\]

(4.19)

The density matrix is a mathematical object, it cannot be observed directly. In order to obtain a more physical result, one would like to evaluate a local operator \( \hat{O}_S \) on \( \rho_S \), as in equation 2.18. By linearity of expectation (trace and integral commute), this is straightforward:

\[
\hat{O}_S(t) := \int_{U(d)} dV \text{Tr} \left( \hat{O}_S \rho_S(t) \right) = \text{Tr} \left( \hat{O}_S \int_{U(d)} dV \rho_S(t) \right) = \text{Tr} \left( \hat{O}_S \bar{\rho}_S(t) \right) = \left( \begin{array}{c} \chi(t) - 1 \\ d^2 - 1 \end{array} \right) |1_S \rangle \langle 1_S| + \left( \begin{array}{c} d_B(d^2 - \chi(t)) \\ d(d^2 - 1) \end{array} \right) \text{Tr}(\hat{O}_S)
\]

(4.20)

Where we have substituted 4.16. Integrating over all systems has left little structure of the operator, a contribution from the initial condition, and a contribution from the diagonal, mixed ensemble average.

### 4.1.2 Variance of the Reduced Density Matrix

The next logical step is to calculate the variance of the matrix elements of \( \rho(t) \) over the unitary group. This will lend credence to the average. For a small variance, we may assume most systems behave like their average. Unfortunately, this is expression quartic in \( V \) as well as in \( V^\dagger \), so \( q = 4 \), and the number of
terms contributing to the integral scales as \((q!)^2\), exploding to 576 combinatorial pairs \(\sigma, \tau \in S_4\). We will touch upon the important steps of the calculation, most of the manipulation was performed by computer.

The variance for a complex quantity is given by:

\[
\text{var}(\rho_S(t)) := \int_{U(d)} dV |\rho_S(t)|^\circ 2 - \left| \int_{U(d)} dV \rho_S(t) \right|^\circ 2  \tag{4.21}
\]

In this expression, modulus squared is taken element-wise, a Hadamard power, not as a matrix, for which the circle serves as a reminder. The second term is simply the Hadamard square of \(\bar{\rho}_S(t)\). The first, we elaborate:

\[
\int_{U(d)} dV |\rho_S(t)|^\circ 2 = \int_{U(d)} dV \left( \sum_{k_B=1}^d \sum_{j,m=1}^d V_{(k_S,k_B)} V_{m_{(g_S,g_B)}} V_{1_j} V_{m_{(g_S,g_B)}}^\dagger e^{i(E_m - E_j)t} \right)^\circ 2
\]

\[
= \int_{U(d)} dV \sum_{k_B,g_B=1}^d \sum_{j,m,l,h=1}^d V_{(k_S,k_B)} V_{1_m} V_{1_l} V_{m_{(g_S,g_B)}} V_{1_h} \times V_{j_1} V_{m_{(g_S,g_B)}}^\dagger V_{l_{(g_S,g_B)}}^\dagger e^{i(E_m + E_l - E_j - E_h)t}
\]

\[
= \int_{U(d)} dV \sum_{k_B,g_B=1}^d \sum_{j,m,l,h=1}^d V_{(k_S,k_B)} V_{1_m} V_{1_l} V_{m_{(g_S,g_B)}} V_{1_h} \times V_{j_1} V_{m_{(g_S,g_B)}}^\dagger V_{l_{(g_S,g_B)}}^\dagger e^{i(E_m + E_l - E_j - E_h)t}
\]

\[
\quad = \int_{U(d)} dV \sum_{k_B,g_B=1}^d \sum_{j,m,l,h=1}^d V_{(k_S,k_B)} V_{1_m} V_{1_l} V_{m_{(g_S,g_B)}} V_{1_h} \times V_{j_1} V_{m_{(g_S,g_B)}}^\dagger V_{l_{(g_S,g_B)}}^\dagger e^{i(E_m + E_l - E_j - E_h)t} \tag{4.22}
\]

Now invoking Collins’ identity in 4.6, a peculiar pattern emerges: The full range indices \(j, m, l, h\) decouple from the split ones \(k_S, k_B, g_S, g_B\). Defining multi-indices \(I := (i_1, i_2, i_3, i_4)\) and similar for \(I', J, J'\):

\[
I = ((k_S, k_B), 1, 1, (g_S, g_B)) ; \quad I' = (1, (g_S, k_B), (k_S, g_B), 1)
\]

\[
J = (j, m, l, h) ; \quad J' = (j, m, l, h) \tag{4.23}
\]

This means \(\sigma\) only takes split indices (or index 1) in \(I\) to themselves in \(I'\), \(\tau\) only takes full indices in \(J\) to other full indices in \(J'\). The sums over these sets indices can be then performed independently.
\[
\int_{\mathcal{U}(d)} dV |\rho_S(t)|^{q^2} = \\
\sum_{\sigma, \tau \in S_4} \left( \sum_{k_B, g_B=1}^{d_B} \sum_{j, m, l, h=1}^d \delta_{I, \sigma(J')} \delta_{J, \tau(J')} e^{i(E_m + E_l - E_h)t} \right) Wg(d, \sigma \tau^{-1}) = \\
\sum_{\sigma, \tau \in S_4} \left( \sum_{k_B, g_B=1}^{d_B} \delta_{I, \sigma(J')} \left( \sum_{j, m, l, h=1}^d \delta_{J, \tau(J')} e^{i(E_m + E_l - E_h)t} \right) \right) Wg(d, \sigma \tau^{-1})
\]

(4.24)

In this expression, the obvious shorthand is \(\delta_{I, \sigma(J')} := \prod_{z=1}^4 \delta_{I_z, \sigma(I'_z)}\).

This allows the following simplification: We will create a table of contributions \(R_{\sigma}\) and \(Q_{\tau}\) due to \(\sigma, \tau \in S_4\) respectively.

\[
R_{\sigma} := \sum_{k_B, g_B=1}^{d_B} \delta_{I, \sigma(J')}, \quad Q_{\tau} := \sum_{j, m, l, h=1}^d \delta_{J, \tau(J')} e^{i(E_m + E_l - E_h)t}
\]

(4.25)

Each combination from each table must be multiplied by the corresponding Weingarten function, and forms one term of the final expression. From 4.24 and 4.25, we substitute:

\[
\int_{\mathcal{U}(d)} dV |\rho_S(t)|^{q^2} = \sum_{\sigma, \tau \in S_4} R_{\sigma} Q_{\tau} Wg(d, \sigma \tau^{-1})
\]

(4.26)

See the expressions for \(R_{\sigma}\) and \(R_{\tau}\) in Table 2.

<table>
<thead>
<tr>
<th>(\sigma, \tau \in S_4)</th>
<th>(R_{\sigma})</th>
<th>(Q_{\tau})</th>
<th>(\sigma, \tau \in S_4)</th>
<th>(R_{\sigma})</th>
<th>(Q_{\tau})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1\bar{d})</td>
<td>(</td>
<td>1_S\rangle \langle 1_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(1\bar{2})</td>
</tr>
<tr>
<td>(123)</td>
<td>(d_B</td>
<td>1_S\rangle \langle 1_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(13)</td>
</tr>
<tr>
<td>(132)</td>
<td>(</td>
<td>k_S\rangle \langle 1_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(14)</td>
</tr>
<tr>
<td>(124)</td>
<td>(d_B</td>
<td>1_S\rangle \langle 1_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(23)</td>
</tr>
<tr>
<td>(142)</td>
<td>(</td>
<td>1_S\rangle \langle g_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(24)</td>
</tr>
<tr>
<td>(134)</td>
<td>(</td>
<td>k_S\rangle \langle 1_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(34)</td>
</tr>
<tr>
<td>(143)</td>
<td>(d_B</td>
<td>1_S\rangle \langle 1_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(1234)</td>
</tr>
<tr>
<td>(234)</td>
<td>(</td>
<td>1_S\rangle \langle g_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(1243)</td>
</tr>
<tr>
<td>(243)</td>
<td>(d_B</td>
<td>1_S\rangle \langle 1_S</td>
<td>)</td>
<td>(\chi(t))</td>
<td>(1324)</td>
</tr>
<tr>
<td>((12)(34))</td>
<td>(d_B ^2</td>
<td>1_S\rangle \langle g_S</td>
<td>)</td>
<td>(d^2)</td>
<td>(1342)</td>
</tr>
<tr>
<td>((13)(24))</td>
<td>(d_B</td>
<td>k_S\rangle \langle g_S</td>
<td>)</td>
<td>(d^2)</td>
<td>(1423)</td>
</tr>
<tr>
<td>((14)(23))</td>
<td>(</td>
<td>1_S\rangle \langle 1_S</td>
<td>)</td>
<td>(\zeta(t))</td>
<td>(1432)</td>
</tr>
</tbody>
</table>

We will elaborate on the derivation of the various terms. In the following, we use the observation that \(\delta^{2}_{\bullet \bullet} = \delta^{\bullet \bullet}_{\bullet}\), and will move back and forth between index and bra-ket notation when needed. Let us start with the column of \(R_{\sigma}\).
For $\sigma = \text{Id}$, the identity, we have $I = I'$, or from 4.23:

$$(k_S, k_B) = 1, \quad 1 = (g_S, k_B), \quad 1 = (k_S, g_B), \quad (g_S, g_B) = 1$$

(4.27)

Which kills all the summands except for $k_B = 1 = g_B$, and elements except $g_S = 1 = k_S$.

$$R_{Id} = \sum_{k_B, g_B = 1}^{d_B} \delta_{(k_S, k_B), 1} \delta_{1, (g_S, k_B)} \delta_{1, (k_S, g_B)} \delta_{(g_S, g_B), 1} = |1_S \rangle \langle 1_S|$$

(4.28)

Permutations sending $(1, 4) \mapsto (1, 4)$ have the same assignment (as they permute $(2, 3) \mapsto (2, 3)$ as well). So for $\sigma \in \{(14), (23), (14)(23)\}$, $R_{\sigma} = R_{Id}$.

For $\sigma = (12)$, one finds:

$$(k_S, k_B) = (g_S, k_B), \quad 1 = 1, \quad 1 = (k_S, g_B), \quad (g_S, g_B) = 1$$

(4.29)

This is satisfied for all $k_B$ in the sum, contributing a factor $d_B$. All other indices are set to one and equation 4.25 simplifies to $R_{(12)} = d_B |1_S \rangle \langle 1_S|$. The same assignments hold for $\sigma \in \{(124), (123), (1234)\}$. When $\sigma \in \{(34), (143), (243), (1432)\}$, a factor $d_B$ comes from free summation over $g_B$, and $R_{\sigma} = d_B |1_S \rangle \langle 1_S|$ too.

For $\sigma = (13)$, see the assignment:

$$(k_S, k_B) = (k_S, g_B), \quad 1 = (g_S, k_B), \quad 1 = 1, \quad (g_S, g_B) = 1$$

(4.30)

Here, the condition on $k_S$ is trivial, the ket vector component in $|\rho_S(t)|^2$. All other indices are forced to 1: $R_{(13)} = |k_S \rangle \langle 1_S|$. Also $\sigma \in \{(132), (134), (1324)\}$ permute the 1st index to the 3rd, giving $R_{\sigma} = R_{(13)}$.

Conversely, $\sigma \in \{(24), (142), (234), (1423)\}$ all permute the 4th index to the 2nd and the rest of the indices are set to unity by equation 4.23. This means the $g_S$ index is free, or the ket vector component: $R_{\sigma} = |1_S \rangle \langle g_S|$

The next permutation not yet treated is $\sigma = (1243)$. Here we obtain the assignment:

$$(k_S, k_B) = (g_S, k_B), \quad 1 = 1, \quad 1 = (g_S, g_B), \quad (k_S, g_B) = (g_S, k_B)$$

(4.31)

This restricts to diagonal elements $k_S = g_S$, but both dummy bath indices $g_B, k_B$ are free, resulting in $R_{(1243)} = \delta_{k_B}^2 |1_S \rangle \langle 1_S| = R_{(12)(34)}$ by analogous reasoning. The final $R_{\sigma}$ comes from $\sigma \in \{(1342), (13)(24)\}$:

$$(k_S, k_B) = (k_S, g_B), \quad 1 = 1, \quad 1 = (g_S, g_B), \quad (g_S, k_B) = (g_S, k_B)$$

(4.32)

In this case, all bra and ket elements survive, only a $\delta_{g_B,k_B}$ kills one dummy sum, leaving a factor $d_B$: $R_{(1342)} = d_B \langle g_S| = R_{(13)(24)}$. 

44
We now move on to the column of $Q_\tau$ defined in 4.25. The first to consider is $\tau = \text{Id}$. From 4.23, we see immediately that there are only tautological conditions on the indices $j, m, l, h$. Then the sum factorizes:

$$Q_{\text{Id}} = \sum_{j, m, l, h=1}^d \delta_{j,m} e^{i(E_m + E_l - E_j - E_h)t}$$

$$= \left( \sum_{j, m=1}^d e^{i(E_m - E_j)t} \right) \left( \sum_{l, h=1}^d e^{i(E_l - E_h)t} \right) = \chi^2(t)$$

(4.33)

Where we are reminded of the definition of $\chi(t)$ in 4.12.

Continuing to $\tau = (12)$, we have the identification $j = m$, and $l, h$ are trivial. Then after annihilating the Kronecker delta against the summation over $m$, $E_m$ cancels $E_j$ in the exponent of 4.25:

$$Q_{(12)} = \sum_{j, m, l, h=1}^d \delta_{j,m} e^{i(E_m + E_l - E_j - E_h)t} = \sum_{j, l, h=1}^d e^{i(E_l - E_h)t} = d\chi(t)$$

(4.34)

We clarify that this is not a differential, but a multiplication by dimension $d$. The same results from $\tau \in \{(13), (24), (34)\}$, any permutation taking $m \lor l \leftrightarrow j \lor h$.

Turning to $\tau = (14)$, the assignment is $j = h$ after killing the Kronecker delta. This leaves us with a newly defined function:

$$Q_{(14)} = \sum_{j, m, l, h=1}^d \delta_{j,h} e^{i(E_m + E_l - E_j - E_h)t} = \sum_{j, l, m=1}^d e^{i(E_m + E_l - 2E_j)t}$$

(4.35)

We similarly define, for $\tau = (23)$, $m = l$:

$$Q_{(23)} = \sum_{j, m, l, h=1}^d \delta_{m,l} e^{i(E_m + E_l - E_j - E_h)t} = \sum_{j, l, m=1}^d e^{i(2E_m - E_h - E_j)t}$$

(4.36)

We note that, up until now, all elements have been real, and by construction the variance is real. However, in general $\xi \pm (t)$ are complex, and only combinations such as $\xi_- + \xi_+$ are real, where we can pair each term $e^\cdot$ with its conjugate $e^{-\cdot}$. A good consistency check will be that the final answer does not feature asymmetric combinations of these functions. On to the first three-cycle, $\tau = (123)$. Here, as with all three-cycles, 3 out of 4 indices are equated, such as $j = m = l$, and the 4th, $h$ is left free. In all these cases, two of the three equated indices kill their energies in the exponent, i.e. $E_m + E_l - E_j = E_m$, and the 4th has opposite sign. All three-cycles contribute a factor:
\[ Q_{(123)} = \sum_{j,m,l,h=1}^{d} \delta_{j,m} \delta_{m,l} \delta_{l,h} \delta_{h,j} e^{i(E_m + E_l - E_j - E_h)t} = \sum_{j,h=1}^{d} e^{i(E_j - E_h)t} \equiv \chi(t) \]

(4.37)

Analogously, all four-cycles equate all indices \( j = m = l = h \). Three summations kill four Kronecker-deltas, as the last is redundant (cyclic).

\[ Q_{(1234)} = \sum_{j,m,l,h=1}^{d} \delta_{j,m} \delta_{m,l} \delta_{l,h} \delta_{h,j} e^{i(E_m + E_l - E_j - E_h)t} = \sum_{j,l=1}^{d} e^{0} = d \]

(4.38)

They all contribute \( Q_{\tau} = d \). Continuing to the double two-cycle \((12)(34)\), \( j = m, l = h \), only 2 summations suffice to cancel the exponent:

\[ Q_{(12)(34)} = \sum_{j,m,l,h=1}^{d} \delta_{j,m} \delta_{l,h} e^{i(E_m + E_l - E_j - E_h)t} = \sum_{j,l=1}^{d} e^{0} = d^2 \]

(4.39)

The same is true for the cross-assignment, \( \tau = (13)(24) \). The final term is \( Q_{(14)(23)} \), where \( j = h, m = l \). This doubles the frequency of the exponent, defining:

\[ Q_{(14)(23)} = \sum_{j,m,l,h=1}^{d} \delta_{j,h} \delta_{m,l} e^{i(E_m + E_l - E_j - E_h)t} = \sum_{j,m=1}^{d} e^{i(2E_m - 2E_j)t} \\
\quad = \zeta(t) := d + \sum_{j < m} 2 \cos(2(E_m - E_j)t) \]

(4.40)

Finally, see table 3 for the numerators of the appropriate Weingarten functions, from [41]. All \( Wg(d, \sigma) \) for \( \sigma \) in a particular \( S_4 \) have the same denominator.

\[ Wg(d, \sigma) \bigg|_{\sigma \in S_4} = \frac{A_4(d, \sigma)}{B_4(d)}, \]

(4.41)

\[ B_4(d) = d^2(d^2 - 1)(d^2 - 4)(d^2 - 9) = \prod_{z=0}^{3} (d + z)(d - z) \]

<table>
<thead>
<tr>
<th>( \sigma ) conjugacy class</th>
<th>( A_4(d, \sigma) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 1, 1, 1</td>
<td>( d^4 - 8d^2 + 6 )</td>
</tr>
<tr>
<td>2, 1, 1</td>
<td>( -d^3 + 4d )</td>
</tr>
<tr>
<td>2, 2</td>
<td>( d^2 + 6 )</td>
</tr>
<tr>
<td>3, 1</td>
<td>( 2d^2 - 3 )</td>
</tr>
<tr>
<td>4</td>
<td>( -5d )</td>
</tr>
</tbody>
</table>
We are now equipped to perform the sum in 4.26, using tables 2 and 3. Simplifying the expression is arduous, but straightforward. From equation 4.41, it is clear that all terms of the sum will have the same denominator. We first collect the numerator, and state it here without the preceding simplification.

\[
\sum_{\sigma, \tau \in S_4} R_{\sigma} Q_{\tau} A_4(d, \sigma \tau^{-1}) = |1_S\rangle \langle 1_S| \left [ (\xi_-(t) + \xi_+(t) + \zeta(t) + \chi^2(t) - 4\chi(t)) \right ]
\]

\[
(d^4 - 2d^3 - 7d^2 + 8d + 12) - 2d_B \left \{ d^5 - 2d^4 - 9d^3 + 18d^2 - \\
\chi(t)(d^5 - 2d^4 - 5d^3 + 2d^2 + 4d + 24) + (\xi_-(t) + \xi_+(t) + \zeta(t) + \chi^2(t)) (d^3 - 4d^2 + d + 6) \right \}
\]

\[
- \left ( |k_S\rangle \langle 1_S| + |1_S\rangle \langle g_S| \right ) \left [ d^5 - 2d^4 - 9d^3 + 18d^2 - \chi(t)(d^3 - 2d^2 - 5d^3 + 2d^2 + 4d + 24) \right ] + \left ( d_B^2 1_S + d_B |k_S\rangle \langle g_S| \right ) \cdot
\]

\[
\left [ d^6 - d^5 - 11d^4 + 9d^3 + 18d^2 - 2\chi(t)(d^4 - 2d^3 - 7d^2 + 8d + 12) + \\
(\xi_-(t) + \xi_+(t) + \zeta(t) + \chi^2(t)) (d^2 - 5d + 6) \right ]
\]

(4.42)

The polynomials in \( d \) here can be factorized neatly, almost all are products of terms of the form \((d \pm z), z \in \{0, 1, 2, 3\}, \) as is denominator \( B_4(d) \).

\[
\sum_{\sigma, \tau \in S_4} R_{\sigma} Q_{\tau} A_4(d, \sigma \tau^{-1}) = |1_S\rangle \langle 1_S| \left [ (\Xi(t) - 4\chi(t))(d + 1)(d - 2)(d + 2)(d - 3) \right ]
\]

\[
- \left ( |k_S\rangle \langle 1_S| + |1_S\rangle \langle g_S| + 2d_B |1_S\rangle \langle 1_S| \right ) \left [ d^2(d - 2)(d - 3)(d + 3) \right ]
\]

\[
- \chi(t)(d - 2)(d + 2)(d - 3)(d^2 + d + 2) + \Xi(t)(d + 1)(d - 2)(d - 3) \right ]
\]

\[
+ \left ( d_B^2 1_S + d_B |k_S\rangle \langle g_S| \right ) \left [ d^2(d + 1)(d - 2)(d - 3)(d + 3) \right ]
\]

\[
- 2\chi(t)(d + 1)(d - 2)(d + 2)(d - 3) + \Xi(t)(d - 2)(d - 3) \right ]
\]

(4.43)

Here we have also substituted the placeholder:
\[ \Xi(t) := \xi_-(t) + \xi_+(t) + \zeta(t) + \chi^2(t) \] 

(4.44)

We can divide by \( B_A(d) \) in 4.41 to find the average of \( \rho_S \) squared.

\[
\int_{U(d)} dV |\rho_S(t)|^2 = |1_S\rangle \langle 1_S| \left[ \frac{\Xi(t) - 4\chi(t)}{d^2(d-1)(d+3)} \right] \\
- \left( |k_S\rangle \langle 1_S| + |1_S\rangle \langle g_S| + 2d_B |1_S\rangle \langle 1_S| \right) \left[ \frac{1}{(d+1)(d-1)(d+2)} \right] \\
- \frac{\chi(t)(d^2 + d + 2)}{d^2(d+1)(d-1)(d+3)} + \frac{\Xi(t)}{d^2(d-1)(d+2)(d+3)} \\
+ \left( d_B^2 |1_S\rangle \langle k_S| + d_B |1_S\rangle \langle g_S| \right) \left[ \frac{1}{(d-1)(d+2)} \right] \\
- \frac{2\chi(t)}{d^2(d-1)(d+3)} + \frac{\Xi(t)}{d^2(d-1)(d+1)(d+2)(d+3)} 
\] 

(4.45)

From this expression, we must subtract the squared average to find the variance, 4.21. From 4.16,

\[
\left| \int_{U(d)} dV \rho_S(t) \right|^2 = \left| \left( \chi(t) - \frac{1}{d^2 - 1} \right) |1_S\rangle \langle 1_S| + \left( \frac{d_B(d^2 - \chi(t))}{d(d^2 - 1)} \right) |1_S\rangle \langle 1_S| \right|^2 = \\
\left[ \left( \chi(t) - \frac{1}{d^2 - 1} \right)^2 + 2 \left( \chi(t) - \frac{1}{d^2 - 1} \right) \left( \frac{d_B(d^2 - \chi(t))}{d(d^2 - 1)} \right) \right] |1_S\rangle \langle 1_S| + \left( \frac{d_B(d^2 - \chi(t))}{d(d^2 - 1)} \right)^2 |1_S\rangle \langle 1_S| \\
\left[ \frac{d(\chi(t) - 1)^2 + 2d_B(\chi(t) - 1)(d^2 - \chi(t))}{d(d^2 - 1)^2} \right] |1_S\rangle \langle 1_S| + \left( \frac{d_B(d^2 - \chi(t))}{d(d^2 - 1)} \right)^2 |1_S\rangle \langle 1_S| 
\] 

(4.46)

Symbolically, the structure of this variance is:

\[
\text{Var}(\rho_S) = \begin{pmatrix}
|1_S\rangle & |g_S\rangle \\
|1_S\rangle & |k_S\rangle
\end{pmatrix}
\]

(4.47)

Composition of symbols simply means addition. The legend is the following:

On the 

(1, 1) position:
\[ \bigspadesuit \cong \frac{\Xi(t) - 4\chi(t)}{d^2(d-1)(d+3)} - \frac{\chi^2(t) - 2\chi(t) + 1}{(d^2 - 1)^2} - 2d_B \left[ \frac{d^2 + \chi^2(t) - d^2\chi(t) - \chi(t)}{d^2 - 1} \right] \]

The first row and column feature:

\[ \bigspadesuit \cong - \frac{1}{(d+1)(d-1)(d+2)} + \frac{\chi(t)(d^2 + d + 2)}{d^2(d+1)(d-1)(d+3)} - \frac{\Xi(t)}{d^2(d-1)(d+2)(d+3)} \]

On the diagonal, we have:

\[ \lozenge \cong d_B^2 \left[ \frac{1}{(d-1)(d+2)} - \frac{2\chi(t)}{d^2(d-1)(d+3)} + \frac{\Xi(t)}{d^2(d-1)(d+1)(d+2)(d+3)} \right. \]

\[ \left. - \frac{d^4 - 2d^2\chi(t) + \chi^2(t)}{d^2(d^2 - 1)^2} \right] \]

Finally, every element of the variance has a contribution:

\[ \heartsuit \cong d_B \left[ \frac{1}{(d-1)(d+2)} - \frac{2\chi(t)}{d^2(d-1)(d+3)} + \frac{\Xi(t)}{d^2(d-1)(d+1)(d+2)(d+3)} \right] \]

It is a straightforward exercise to check that the variance among the initial states vanishes, by plugging in \( t = 0 \) into all dependent functions. For this we use \( \Xi(0) = d^4 + 2d^3 + d^2 \) and \( \chi(0) = d^2 \). After this, it is interesting to consider the time average of the terms. First we will average of the time dependent terms. As before \( \langle \chi(t) \rangle_t = d \). Its average square is not \( d^2 \), though.

\[ \langle \chi^2(t) \rangle_t = \left( \left( d + \sum_{j < m} 2\cos((E_m - E_j)t) \right)^2 \right)_t = d^2 + 4d \sum_{j < m} \cos((E_m - E_j)t) + 4 \sum_{j < m, l < h} \cos((E_m - E_j)t) \cos((E_l - E_h)t) \]

\[ = d^2 + 4 \cdot 0 + 4 \sum_{j < m} \langle \cos^2 ((E_m - E_j)t) \rangle_t = d^2 + 4 \cdot \frac{d^2 - d}{2} \cdot \frac{1}{2} = 2d^2 - d \]
In the last line, we used the fact that only the terms with \((j, l) = (m, h)\) don’t average to zero, instead giving the average of the cosine squared, which is half the cosine squared plus sine squared. Note there are \((d^2 - d) / 2\) terms \(j < m\).

The following are straightforward, for only zero-frequency terms contribute:

\[
\langle \xi_+ (t) \rangle_t = \langle \xi_- (t) \rangle_t = \langle \zeta (t) \rangle_t = d
\]

Then:

\[
\langle \Xi (t) \rangle_t = \langle \chi^2 (t) + \xi_+ (t) + \xi_- (t) + \zeta (t) \rangle_t = 2d(d + 1)
\]

Seen by 4.52, 4.44, and linearity of expectations. Substituting these into 4.48 - 4.51, we find, the time average elements and their scaling limits:

\[
\langle \spadesuit \rangle_t = \frac{-3d^2 + d - 2 + 2d_B(2d^3 - 7d + 1)}{d(d - 1)(d + 3)(d + 1)^2} \rightarrow d \rightarrow \infty \frac{4d_B}{d^2}
\]

\[
\langle \clubsuit \rangle_t = \frac{d - 1}{d(d + 3)(d + 1)} \rightarrow d \rightarrow \infty \frac{1}{d^2}
\]

\[
\langle \heartsuit \rangle_t = \frac{d^2 + d^2 + 6d + 12}{d(d - 1)(d + 2)(d + 3)(d + 1)^2} \rightarrow d \rightarrow \infty \frac{-d_B^2}{d^3}
\]

\[
\langle \diamondsuit \rangle_t = \frac{d_B}{d(d + 3)} \rightarrow d \rightarrow \infty \frac{d_B}{d^2}
\]

One might be troubled by the negative expectation for \(\diamondsuit\) on the diagonal, but it will be compensated by the larger, positive value on all elements \(\diamondsuit\). The variance on the diagonal is thus expected to be lower than elsewhere. Moreover, the variance tends to the zero matrix as we increase the bath size, confirming intuition. The same is true when we increase the subsystem dimension. Although, enlarging the subsystem simply spreads occupation over a larger number of elements. The resulting lowered variance should not come as a surprise. The variance tends to zero faster than the average itself. The standard deviation, as the square root, has comparable scaling behavior with the average.

Tentative numerical Monte-Carlo estimates of the average 4.16 and variance 4.47 appear to corroborate these analytic results, though convergence is slow, and checking for many different times is laborious. For that reason, we suggest rigorous numerical validation as a good point of future study.

### 4.2 Mean Subsystem Dynamical Purity

The last quantity we will consider in this chapter, though a pleasantly insightful one, is the mean subsystem purity, defined in 2.27. Like the entropy, it quantifies disorder, but unlike entropy, it is not calculated through diagonalization of the density matrix. The derivation is reminiscent of the standard deviation: it is
also fourth order in elements of $V$. From 4.5 we contract the inner and outer subsystem indices in order to find $\text{Tr}(\rho_S^2(t))$

$$
\bar{\gamma}(t) := \int_{U(d)} dV \text{tr}(\rho_S^2(t)) = \int_{U(d)} dV \sum_{k_B,gB=1}^{d_B} \sum_{k_S,gS=1}^{d_S} \sum_{j,l,m,h=1}^d \left(V_{(k_S,kB)} V_{l(k_S,kB)} V_{l(gS,gB)} V_{m(gS,gB)} \right)^{i(E_l-E_j)t} V_{1h} V_{m1} V_{h1}^{\dagger} \epsilon^{i(E_h-E_m)t} 
$$

$$
\int_{U(d)} dV \sum_{k_B,gB=1}^{d_B} \sum_{k_S,gS=1}^{d_S} \sum_{j,l,m,h=1}^d \left(V_{(k_S,kB)} V_{l(k_S,kB)} V_{l(gS,gB)} V_{m(gS,gB)} \right)^{i(E_l-E_j) t} \epsilon^{i(E_h-E_m) t} 
$$

(4.56)

From this expression, we again compose the indices needed for Collins' integral, 4.6. Analogous to 4.23 observe:

$$
I = ((k_S,k_B),1,(g_S,g_B),1); \quad I' = (1,(g_S,k_B),1,(k_S,g_B)) \quad J = (j,m,l,h); \quad J' = (j,m,l,h) 
$$

(4.57)

Again the multi-indices decouple from the full indices. We can construct new $R_\sigma, Q_\tau$, which are similar to those used in the variance, although for scalar purity, these are all scalar expressions.

$$
R_\sigma := \sum_{k_B,gB=1}^{d_B} \sum_{k_S,gS=1}^{d_S} \delta_{I,\sigma(^{I'})}, \quad (4.58)
$$

$$
Q_\tau := \sum_{j,m,l,h=1}^{d} \delta_{J,\tau(^{J'})} \epsilon^{i(E_h-E_j-E_m) t} \quad (4.59)
$$

For the results, see table 4:

<table>
<thead>
<tr>
<th>$\sigma, \tau \in S_4$</th>
<th>$R_\sigma$</th>
<th>$Q_\tau$</th>
<th>$\sigma, \tau \in S_4$</th>
<th>$R_\sigma$</th>
<th>$Q_\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id</td>
<td>1</td>
<td>$\chi(t)$</td>
<td>(12)</td>
<td>$d_B$</td>
<td>$d\chi(t)$</td>
</tr>
<tr>
<td>(123)</td>
<td>$d_B$</td>
<td>$\chi(t)$</td>
<td>(13)</td>
<td>1</td>
<td>$\xi_+(t)$</td>
</tr>
<tr>
<td>(132)</td>
<td>$d_S$</td>
<td>$\chi(t)$</td>
<td>(14)</td>
<td>$d_S$</td>
<td>$d\chi(t)$</td>
</tr>
<tr>
<td>(124)</td>
<td>$d_B$</td>
<td>$\chi(t)$</td>
<td>(23)</td>
<td>$d_S$</td>
<td>$d\chi(t)$</td>
</tr>
<tr>
<td>(142)</td>
<td>$d_S$</td>
<td>$\chi(t)$</td>
<td>(24)</td>
<td>1</td>
<td>$\xi_-(t)$</td>
</tr>
<tr>
<td>(134)</td>
<td>$d_B$</td>
<td>$\chi(t)$</td>
<td>(34)</td>
<td>$d_B$</td>
<td>$d\chi(t)$</td>
</tr>
<tr>
<td>(143)</td>
<td>$d_S$</td>
<td>$\chi(t)$</td>
<td>(1234)</td>
<td>$d\cdot d_B$</td>
<td>$d$</td>
</tr>
<tr>
<td>(234)</td>
<td>$d_B$</td>
<td>$\chi(t)$</td>
<td>(1243)</td>
<td>$d_B$</td>
<td>$d$</td>
</tr>
<tr>
<td>(243)</td>
<td>$d_S$</td>
<td>$\chi(t)$</td>
<td>(1324)</td>
<td>$d_S$</td>
<td>$d$</td>
</tr>
<tr>
<td>(12)(34)</td>
<td>$d\cdot d_B$</td>
<td>$d^2$</td>
<td>(1342)</td>
<td>$d_B$</td>
<td>$d$</td>
</tr>
<tr>
<td>(13)(24)</td>
<td>1</td>
<td>$\zeta(t)$</td>
<td>(1432)</td>
<td>$d_S$</td>
<td>$d$</td>
</tr>
<tr>
<td>(14)(23)</td>
<td>$d\cdot d_S$</td>
<td>$d^2$</td>
<td>(1432)</td>
<td>$d\cdot d_S$</td>
<td>$d$</td>
</tr>
</tbody>
</table>
Due to the difference in signs in the exponent (there is no complex conjugate in 4.56 as there is in 4.22), the $Q_\tau$ column has a slightly different order than in table 2. Besides this, the derivation is exactly the same as in the previous section.

We now treat the case of $R_\sigma$. A look at 4.57 tells us any permutation taking $(1,3) \rightarrow (1,3)$ forces all indices to 1, and the sum in 4.58 is trivially unit, for $\sigma \in \{\text{Id}, (13), (24), (13)(24)\}$. Any that maps $1 \rightarrow 2 \lor 3 \rightarrow 4$, will leave a single free bath index, producing $R_\sigma = d_B$ after summation, for $\sigma \in \{(12), (34), (123), (124), (134), (234), (1243), (1342)\}$. Conversely, for a single free subsystem index, $1 \rightarrow 4 \lor 3 \rightarrow 1$: $R_\sigma = d_S$ when $\sigma \in \{(12), (34), (132), (134), (234), (1234), (1342)\}$. Lastly, mapping $1 \rightarrow 4 \land 3 \rightarrow 2$ leaves both subsystem indices undetermined, and equates the bath indices to each other, killing one summation. Then $R_\sigma = d_S d_B = d \cdot d_B$ when the bath indices are both free, and subsystem indices are equated, in $\sigma \in \{(12)(34), (1234)\}$. Remembering 4.41 and 4.56, and table 3, find the numerator:

$$\sum_{\sigma,\tau \in S_4} R_\sigma Q_\tau A_4(d, \sigma^{-1}) = (\Xi(t) - 4\chi(t)) \left[ d^4 - 2d^3 - 7d^2 + 8d + 12 \right] + (d_S + d_B)$$

$$\left[ d^2 - d^6 - 13d^5 + 13d^4 + 36d^3 - 36d^2 \right] - (\Xi(t) - 4\chi(t)) \left( d^3 - 3d^2 - 4d + 12 \right)$$

$$= \Gamma(t)(d+1)(d-2)(d+2)(d-3) + (d_S + d_B)$$

$$\cdot \left[ d^2(d-1)(d-2)(d+2)(d-3)(d+3) - \Gamma(t)(d-3)(d+2)(d-2) \right]$$

(4.60)

In the last line, we have defined another function $\Gamma(t) := \Xi(t) - 4\chi(t)$ to hold the time and energy dependence.

$$\Gamma(t) := \left( d + 2 \sum_{j<k} \cos ((E_j - E_k)t) \right)^2 + 2 \sum_{j<k} \cos ((2E_j - E_k - E_l)t)$$

$$+ 2 \sum_{j<k} \left( \cos (2(E_j - E_k)t) - 4 \cos ((E_j - E_k)t) \right) - 3d$$

(4.61)

See figure 4.3 and that beside it for typical forms of this function.

Though larger in absolute sense, $\Gamma(t)$ drops like $\chi(t)$. Commencing at its maximal value, when all phases of the cosines coincide at zero, it equilibrates around a value approximately the maximal value’s square root, when the cosines decohere. This transition again takes place after a duration empirically close to $\sigma^{-1}$ and can be thought of as moving from a ‘known’ to ‘unknown’ phase. For the function to be ‘well behaved’, i.e. have a damped character, considerably fewer dimensions are needed than with $\chi(t)$. 

52
At present the origin of the terms \((d \pm z)\) in expressions such as 4.60 is unclear to the author, however their appearance is fortuitous: they cancel neatly with the denominator. Dividing 4.60 by \(B_4(d)\) from 4.41, we find the mean (uniformly over systems) time dependent purity, for a system of \(d\) dimensions, subsystem and bath of \(d_S\) and \(d_B\) dimensions:

\[
\bar{\gamma}(t) = \frac{\Gamma(t)}{d^2(d-1)(d+3)} \left( 1 - \frac{d_S + d_B}{d+1} \right) + \frac{d_S + d_B}{d+1} \quad (4.62)
\]

This function is symmetric in \(S\) and \(B\). This makes sense. We are reminded that the nonzero eigenvalues of the reduced density matrices \(\rho_S\) and \(\rho_B\) are identical, a fact derived from the full system pure state’s Schmidt decomposition. The purity is only a function of the eigenvalues: \(\gamma(W\rho S W^\dagger) = \gamma(\rho_S)\) for any basis transformation \(W \in U(d_S)\). Also the entanglement entropy of both subsystem and bath are equal \[14\].

At \(t = 0\), \(\Gamma(t) = d^2(d-1)(d+3)\), so \(\gamma(0) = 1\), as all systems start out pure by assumption. Furthermore, for all times, when \(d_B = d\) or \(d_S = d\), the other being unity (so \(d_Bd_S = d\)), there is no tracing, and no loss of information. Indeed, also then \(\bar{\gamma}(t) = 1\). Finally, the time average value of \(\Gamma(t)\) is \(\langle \Gamma(t) \rangle_t = 2d(d-1)\), substituting this linearly gives us the time average, mean subsystem purity:

\[
\langle \gamma(t) \rangle_t = \frac{2}{d(d+3)} \left( 1 - \frac{d_S + d_B}{d+1} \right) + \frac{d_S + d_B}{d+1} \quad (4.63)
\]

Compare this to the average purity obtained by taking a random vector from a \(d = d_B \cdot d_S\) dimensional Hilbert space, and tracing out the \(d_B\) dimensions of the bath \[42\]:

\[
\langle \bar{\gamma} \rangle_{d_S,d_B} = \frac{d_S + d_B}{d+1} \quad (4.64)
\]

More interesting statistics of systems distributed according to these so called trace measures can be found in e.g. \[14\].
The minimum purity is found by minimizing $\frac{d_S + d_B}{d+1}$ over $d_S$, achieved for $d_S = d_B = \sqrt{d}$. So the most uncertainty occurs for an equal partition. This may be interpreted as a compromise between a larger bath taking more information, and a larger subsystem allowing for more possible states over which to spread probability. Finally, substituting $d_B = d_S$ into 4.63, we find the time average minimal purity and its limit:

$$\langle \bar{\gamma}(t) \rangle_t \bigg|_{d_B = d_S} = 2 \left( \frac{d_S^5 + 3d_S^3 + d_S^2 - 2d_S + 1}{d_S^2 (d_S^2 + 1)(d_S^2 + 3)} \right) \xrightarrow{d_S \to \infty} \frac{2}{d_S} \quad (4.65)$$

This scaling behaviour in the large dimension limit is twice minimum purity overall, $1/d_S$, see 2.28. The results in 4.19 would suggest that the mixing tends to the the minimum purity, however this is for a fixed $d_S$. Indeed in that case, taking $d_B \to \infty$ in 4.63 reduces it to $\langle \bar{\gamma}(t) \rangle_t \to 1/d_S$. There is no contradiction here: for a larger subsystem, $2/d_S$ is soon smaller than $1/d_S$ of a small subsystem.

Using the same techniques, other polynomial functions of the reduced density matrix can be averaged over all systems. We have simply treated the most obvious candidates.
Chapter 5

Quantum Differential Geometry

One promising concept, and a natural one to the field of quantum geometry, is to treat Hilbert space as a *Riemannian manifold*. This entails imbuing the space with a metric, and defining differential separation between states. Polkovnikov and Kolodrubetz [25] use a metric approach to thermalization. Understanding the behavior of the metric—and its shortcomings—in the thermodynamic limit is important when taking the geometric line of attack to this problem. Lightly motivated by that theme, we find that the transformation properties of these quantum geometric objects are interesting in their own right. The focus of this chapter will be predominantly the transformation properties of geometric objects under gauge symmetries.

This chapter is somewhat disconnected from the narrative of the thesis. For instance, it is not dependent on the bath/subsystem division ubiquitous above. Nonetheless, the author spent considerable time studying quantum metrics, so it is at least instructive to investigate.

5.1 Manifold of Quantum States

Consider, in all generality, a family of Hamiltonians $H(\lambda)$, acting on a $d$-dimensional Hilbert space $\mathcal{H}$ spanned by the energy spectral basis $\{|\psi_j(\lambda)\rangle\}$. Here $\lambda^\mu \in \mathcal{M}$ is a vector of parameters controlled by the experimenter, influencing the instantaneous spectrum. The dependence on $\lambda$ will only be made explicit when it serves the clarity. What is $\lambda$ physically? Think of actual dials such as magnetic field or position of some barrier.

5.1.1 Nondegenerate Case

For now, we take the system in an energy eigenstate $|\psi_n\rangle$ such as the ground state. The case of the $n^{th}$ excited state is analogous, as additionally, we assume
there are no level crossings anywhere on $M$, i.e. for:

$$H |\psi_j\rangle = E_j |\psi_j\rangle, \quad E_j = E_k \iff |\psi_j\rangle = |\psi_k\rangle \quad (5.1)$$

So this state can unambiguously be translated around $M$ to the same level. Later, we will generalize to degenerate states and then density matrices.

We wish to define the Quantum Geometric Tensor (QGT) from [15] to navigate $M$ in a gauge invariant way, on the state $|\psi_n\rangle \in \mathcal{H}$. It relates an inner product on $M$, the ‘overlap’ between the rate of change of the state for an infinitesimal step in parameter $\lambda^\mu$, and that in parameter $\lambda^\nu$. Following Provost and Vallee, we begin with the instinctive ansatz of the differential:

$$||\psi_n(\lambda + d\lambda) - \psi_n(\lambda)|| = \left(\partial_\mu \langle \psi_n | d\lambda^\mu \right) \left(\partial_\nu |\psi_n\rangle d\lambda^\nu\right) \quad (5.2)$$

Summation convention implied, and using the shorthand operators $\partial_\mu := \frac{\partial}{\partial \lambda^\mu}$. This leads us to investigate the metric-like expression below. Taking the differential operators inside the bras and kets:

$$Q_{\mu \nu} := \langle \partial_\mu \psi_n | \partial_\nu \psi_n \rangle \quad (5.3)$$

We also suppress of the explicit dependence of $|\psi_n\rangle$ on $\lambda$. In the nondegenerate case, the gauge symmetry of the Hilbert space is the $U(1)$ invariance of global phase, as all states technically live in a projective Hilbert space $\mathbb{C}P^{d-1}$, and physical observables are unaffected by which point of the ray $\{c |\psi_n\rangle \} \subset \mathbb{C}$ the state occupies [14]. On our manifold, this symmetry amounts to invariance of the system under multiplication of a phase, which need not be the same at all points $\lambda$. For some $\alpha : M \rightarrow \mathbb{R}$:

$$|\psi_n\rangle \mapsto |\psi'_n\rangle := e^{i\alpha(\lambda)} |\psi_n\rangle \quad (5.4)$$

i.e. these two kets define the same point on the manifold of rays. For any observable $\hat{O}$:

$$\langle \psi'_n | \hat{O} |\psi'_n\rangle = \langle \psi_n | e^{-i\alpha} \hat{O} e^{i\alpha} |\psi_n\rangle = \langle \psi_n | \hat{O} |\psi_n\rangle \quad (5.5)$$

The local expectation value is the same. However, the tensor is not fully local but involves adjacent points. It transforms under this transformation as:

$$Q_{\mu \nu} \rightarrow Q'_{\mu \nu} = \langle \partial_\mu \psi_n | \partial_\nu \psi_n \rangle - i \partial_\mu \alpha \langle \psi_n | \partial_\nu \psi_n \rangle + i \partial_\nu \alpha \langle \partial_\mu \psi_n |\psi_n\rangle + \partial_\mu \alpha \partial_\nu \alpha \langle \psi_n |\psi_n\rangle$$

$$= Q_{\mu \nu} + \partial_\mu \alpha \beta_\nu + \partial_\nu \alpha \beta_\nu^* + \partial_\mu \alpha \partial_\nu \alpha \quad (5.6)$$

Using Leibniz repeatedly in the first step, and invoking the normalization of $|\psi_n\rangle$ in the second. We have defined $\beta_\mu := -i \langle \psi_n | \partial_\mu \psi_n \rangle$. We gather that the exponentials cancel, but the exponents don’t.

Another consequence of the normalization is the following identity:

$$1 \equiv \langle \psi_n |\psi_n\rangle \iff \partial_\mu \langle \psi_n |\psi_n \rangle = 0 = \langle \partial_\mu \psi_n |\psi_n\rangle + \langle \psi_n | \partial_\mu \psi_n \rangle \quad (5.7)$$
But also:

\[ (\partial_\mu \psi_n | \psi_n)^* = (\psi_n | \partial_\mu \psi_n) \]  

(5.8)

So it follows that

\[ (\psi_n | \partial_\mu \psi_n) = - (\psi_n | \partial_\mu \psi_n)^* \Leftrightarrow \beta_\mu \in \mathbb{R} \]  

(5.9)

Note that we don’t assume differential expressions such as \( |\partial_\mu \psi_n \rangle \) normalized.

In order to make the tensor gauge invariant, we observe that under 5.4, \( \beta_\mu \) transforms as:

\[ \beta_\mu \mapsto \beta_\mu' = \beta_\mu + \partial_\mu \alpha \]  

(5.10)

So a gauge invariant expression denoted by curly \( Q_{\mu \nu} \), that will not be sensitive to the particular choice of phase, is given by

\[ Q_{\mu \nu} := Q_{\mu \nu} - \beta_\mu^* \beta_\nu = (\partial_\mu \psi_n | \partial_\nu \psi_n) - (-i^2) (\partial_\mu \psi_n | \psi_n) (\psi_n | \partial_\nu \psi_n) \]  

(5.11)

Where the star \( (\ast) \) serves to remind that we conjugate the bracket in the definition of \( \beta_\mu \). A way to look at this gauge invariant result, is taking the inner product in \( Q_{\mu \nu} \) in the subspace in which we have projected out the instantaneous state \( |\psi_n \rangle \):

\[ Q_{\mu \nu} = (\partial_\mu \psi_n | (1 - |\psi_n \rangle \langle \psi_n |) | \partial_\nu \psi_n) \]  

(5.12)

Now we will find an expression for this tensor in terms of \( H \). In order to do so, we first invoke the spectral theorem, writing the terms \( |\partial_\mu \psi_n \rangle \) in the basis \( \{ |\psi_j \rangle \} \).

\[ |\partial_\mu \psi_n \rangle = \sum_j c_{\mu j} |\psi_j \rangle \]  

(5.13)

Inserting this into 5.12, we find:

\[ Q_{\mu \nu} = \sum_{j,k} (c_{\mu j}^*)^* \langle \psi_j | (1 - |\psi_n \rangle \langle \psi_n |) c_{\nu k} |\psi_k \rangle = \sum_{j,k \neq n} (c_{\mu j}^*)^* c_{\nu k} \delta_{jk} = \sum_{j \neq n} (c_{\mu j}^*)^* c_{\nu j} \]  

(5.14)

Armed with the expression 5.14, we calculate the coefficients \( c_{\mu j}^* \) in terms of energy. Our starting point is to differentiate the time-independent Schrödinger equation with respect to \( \lambda^\mu \), and take the inner product with a \( \langle \psi_j |, j \neq n \). Essentially we are copying the Feynman-Hellman theorem [43].

\[ \partial_\mu (H | \psi_n \rangle) = \partial_\mu (E_n | \psi_n \rangle) \]  

\[ \partial_\mu H | \psi_n \rangle + H | \partial_\mu \psi_n \rangle = \partial_\mu E_n | \psi_n \rangle + E_n | \partial_\mu \psi_n \rangle \Rightarrow \]  

(5.15)

\[ \langle \psi_j | \partial_\mu H | \psi_n \rangle + \langle \psi_j | H | \partial_\mu \psi_n \rangle = \langle \psi_j | \partial_\mu E_n | \psi_n \rangle + \langle \psi_j | E_n | \partial_\mu \psi_n \rangle \]  

\[ \langle \psi_j | \partial_\mu H | \psi_n \rangle + \langle \psi_j | E_j | \partial_\mu \psi_n \rangle = \langle \psi_j | E_n | \partial_\mu \psi_n \rangle \]
In the last line we used the hermicity of \( H \) and the fact that the scalar \( \partial_\mu E_n \) is effectively diagonal between orthonormal eigenstates, killing one term. The result is the following identity:

\[
c_j^\mu \equiv \langle \psi_j | \partial_\mu \psi_n \rangle = \frac{\langle \psi_j | \partial_\mu H | \psi_n \rangle}{E_n - E_j} \quad (5.16)
\]

From 5.14:

\[
Q_{\mu\nu} = \sum_{j \neq n} \frac{\langle \psi_n | \partial_\mu H | \psi_j \rangle \langle \psi_j | \partial_\nu H | \psi_n \rangle}{(E_n - E_j)^2} \quad (5.17)
\]

In 5.12 we observe that exchanging \( \mu \leftrightarrow \nu \) amounts to swapping the bra and ket, or \( Q_{\mu\nu} = Q^*_{\nu\mu} \) is Hermitian in its indices. Because we will use this expression as a metric in the contraction \( Q_{\mu\nu} \lambda^\mu \lambda^\nu \) with the symmetric \( \lambda^\mu \lambda^\nu \), we may also use the symmetrized part of the metric:

\[
g_{\mu\nu} = \text{Re}(Q_{\mu\nu}) = \frac{1}{2} (Q_{\mu\nu} + Q_{\nu\mu}) = \frac{1}{2} (Q_{\mu\nu} + Q^*_{\mu\nu}) \quad (5.18)
\]

This is as far as the nondegenerate case takes us. However, as alluded to at the start of the chapter, this object is employed in the study of thermalization. In the thermodynamic limit, applicable to theories such as the Eigenstate Thermalization Hypothesis, the energy difference between closeby states can become arbitrarily small, from a macroscopic point of view. Roughly, \((E_n - E_j) \approx e^{-S} \approx 1/d\), so nearby energy differences scale inversely with the Hilbert space dimension. In contrast to subsection 2.4.1, here with \( S \) we mean the extensive thermodynamic entropy [25]. This poses a problem in equation the denominator of 5.17 (combined with the fact that the numerator scales inversely at a lower pace with \( d \)). In short: the tensor blows up in the large-dimensional limit. We regard this as a consequence of the effective degeneracy of the energy states on an energy shell \( E_n \in [E, E + \delta E] \), for instance defined by the microcanonical ensemble. This motivates us to investigate the behavior of the QGT in the case of a degenerate, but still finite, system.

### 5.1.2 Degenerate Case

In the following, we will consider the case of calculating \( g_{\mu\nu} \) on a state \( |\psi_n\rangle \in \mathcal{H}_{E_n} \), the subspace of eigenvectors corresponding to energy \( E_n \). For the sake of definiteness, assume a \( k \)-fold degenerate level, with \( m - 1 \) linearly independent energy eigenvectors existing with energies \( < E_n \). Thus with any partial ordering, basis vectors \( m \) through \( (m + k - 1) \) have energy \( E_n \). There are no crossings of levels on \( \mathcal{M} \). I.e. the somewhat strong assumption is that this degeneracy is independent of \( \lambda \).

The Hamiltonian offers us no preferred basis with which to span \( \mathcal{H}_{E_n} \). Thus the choice of basis, which can be made locally in \( \mathcal{M} \), is in essence a gauge choice, and each choice is related to each other by a \( U(k) \) rotation. We have seen above when the global phase of the state was a gauge choice. The subtlety
is, contrary to the nondegenerate case, there can be two different global states of the system \(|\phi_1(\lambda)\rangle, |\phi_2(\lambda)\rangle \in \mathcal{H}_{E_n}\) at the same point \(\lambda\), such that some other observable \(\hat{O} \neq \hat{H}\) could distinguish between the two: \(\langle\phi_1|\hat{O}|\phi_1\rangle \neq \langle\phi_2|\hat{O}|\phi_2\rangle\). Just take orthogonal basis states \(\langle\phi_1|\phi_2\rangle = 0\) and define \(\hat{O} := |\phi_1\rangle \langle \phi_1|\). \(\hat{O}\) is Hermitian so it is an observable, and both states are eigenstates of \(\hat{O}\), but \(\langle\phi_1|\hat{O}|\phi_1\rangle = 1, \langle\phi_2|\hat{O}|\phi_2\rangle = 0\). This is by construction not the case with the \(U(1)\) phase, for within one energy eigenspace there can be no two normalized orthogonal states.

Having mentioned this caveat to calling \(U(k)\) a gauge symmetry, we make an observation. Define \(\{|\phi_l(\lambda)\rangle\}, l \in \{1, \ldots, k\}\) a local orthonormal basis of \(\mathcal{H}_{E_n}\). This allows us to write an arbitrary eigenstate with energy \(E_n\) as:

\[
|\psi_n\rangle = \sum_l c_l |\phi_l\rangle, \quad \sum_l |c_l|^2 = 1
\]  

(5.19)

Let us start from the time-independent Schrödinger equation again:

\[
H |\psi_n\rangle = E_n |\psi_n\rangle
\]

\[
\partial_\mu \left( H \sum_l c_l |\phi_l\rangle \right) = \partial_\mu \left( E_n \sum_l c_l |\phi_l\rangle \right)
\]

\[
\sum_l \left( \partial_\mu E_n c_l |\phi_l\rangle + H \partial_\mu c_l |\phi_l\rangle + H c_l |\partial_\mu \phi_l\rangle \right) =
\]

\[
\sum_l \left( \partial_\mu E_n c_l |\phi_l\rangle + E_n \partial_\mu c_l |\phi_l\rangle + E_n c_l |\partial_\mu \phi_l\rangle \right)
\]  

(5.20)

Now we take the inner product with a single basis state \(\langle \phi_a |\), and move terms around to facilitate notation.

\[
\langle \phi_a | H \partial_\mu c_a |\phi_a\rangle + \sum_l \left( \langle \phi_a | \partial_\mu H c_l |\phi_l\rangle + \langle \phi_a | H c_l |\partial_\mu \phi_l\rangle \right) =
\]

\[
\langle \phi_a | \left( E_n \partial_\mu c_a + \partial_\mu E_n c_a \right) |\phi_a\rangle + \sum_l \langle \phi_a | E_n c_l |\partial_\mu \phi_l\rangle
\]  

(5.21)

Using hermicity of \(H\), we cancel the first term on the LHS of 5.21 with the first term on the RHS, as well as the second part of the LHS sum against the sum on the RHS.

\[
\sum_l \langle \phi_a | \partial_\mu H c_l |\phi_l\rangle = \langle \phi_a | \partial_\mu E_n c_a |\phi_a\rangle
\]

\[
c_a = \sum_l c_l \frac{\langle \phi_a | \partial_\mu H |\phi_l\rangle}{\partial_\mu E_n}
\]  

(5.22)

\[
\delta_{la} = \frac{\langle \phi_a | \partial_\mu H |\phi_l\rangle}{\partial_\mu E_n}
\]
Where the last step is inferred as it must hold for arbitrary \( \{c_j\} \). In other words, the upshot is that the derivative of \( H \) does not appear to mix the degenerate states. This is a consequence of the assumption that the degeneracy is constant on \( \mathcal{M} \).

We wish to build up the machinery as we did in the previous section. In the degenerate case, however, the gauge freedom isn’t generated by an abelian phase \( \alpha(\lambda) \) introduced in 5.4, but a non-abelian matrix valued phase \( A \). The degenerate case cannot be completed in the manner shown in 5.11, there is no way to define a ‘matrix \( \beta_\mu \)’ that can be added to the ansatz, to make it gauge invariant. Observe that \( \langle \psi_n | \partial_\mu A \partial_\nu A | \psi_n \rangle \neq \langle \psi_n | \partial_\mu A | \psi_n \rangle \langle \psi_n | \partial_\nu A | \psi_n \rangle \) if \( A \) is a noncommuting operator. The framework of projectors in equation 5.12 is more useful. We will transition to a different language.

### 5.2 Linear Algebraic Formulation

At this point we elaborate on an alternative formulation and derivation. In the spirit of Quantum Mechanics as a form of Linear Algebra, define a ket-vector as a normalized column vector in \( \mathbb{C}^d \):

\[
|\psi\rangle \sim \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_d \end{pmatrix}, \quad \sum_j |c_j|^2 = 1 \tag{5.23}
\]

These hold the coefficients of each linearly independent basis vector of some basis of \( \mathcal{H} \). In the nondegenerate case, we naturally take the ordered basis of energy eigenstates. The degenerate case admits a partial ordering. Operators \( \hat{O} \) are again \( d \times d \) normal matrices (satisfying \( \hat{O}^\dagger \hat{O} = \hat{O} \hat{O}^\dagger \) \[30\]), and a bra vector is a complex conjugate row vector.

\[
\langle \psi | = |\psi\rangle^\dagger = (c_1^* \quad c_2^* \quad \ldots \quad c_N^*), \quad \sum_j |c_j^*|^2 = 1 \tag{5.24}
\]

Which is evidently still normalized. The derivative \( |\partial_\mu \phi\rangle \) is the element-wise derivative of the entries of the vectors. Now we consider the construction of the QGT.

#### 5.2.1 Nondegenerate case

Let us briefly repeat the nondegenerate case. Here, all energy levels are distinct, and our state initially occupies one of them, the \( n \)th one, with energy \( E_n \). This state is equivalent to a vector \( |\psi_n\rangle \) with \( c_j = c \cdot \delta_{jn} \), for some \( c \in \mathbb{C} \) satisfying \( |c|^2 = 1 \). Though the elements outside the \( n \)th place are zero by construction, their derivatives \( \partial_\mu \) need not be.
\[ |\psi_n\rangle = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ c \\ 0 \\ \vdots \\ 0 \end{pmatrix} \] \text{ }^{n\text{th}}\text{ position} \quad (5.25)

The description is not unique. An additional global (in \( \mathcal{H} \)) phase may be applied locally (in \( \mathcal{M} \)):

\[ |\psi_n\rangle \mapsto (u(\lambda) \cdot 1) |\psi_n\rangle = |\psi'_n\rangle \quad (5.26) \]

For \( e^{i\alpha(\lambda)} = u(\lambda) \in U(1) \) an arbitrary element of the symmetry group of the wave vector. In this formalism, it is multiplied by the \( d \times d \) identity: multiplying all components by the same phase has no effect on the physics. Then, following [44] we wish for this \( u \) to also not feature in our expression for \( Q_{\mu\nu} \).

\[ Q_{\mu\nu} := \langle \partial_\mu \psi'_n | \partial_\nu \psi'_n \rangle = \partial_\mu (\langle \psi_n | u^\dagger \rangle \partial_\nu (u | \psi_n \rangle)) \quad (5.27) \]

I.e. for the tensor to only depend on the coefficients \( c_j \), and their derivatives, in combinations such as \( \partial^* c_j \partial^* c_j \) so that also the particular \( U(1) \) phase chosen in the individual basis vectors cancels out. In order to find a suitable definition, we expand the expression in 5.27 and substitute the generator: \( u = e^{i\alpha}, u^\dagger = e^{-i\alpha} \).

\[ \langle \partial_\mu \psi'_n | \partial_\nu \psi'_n \rangle = (\langle \partial_\mu \psi_n | u^\dagger + \langle \psi_n | \partial_\mu u \rangle \partial_\nu (u | \psi_n \rangle) + u | \partial_\nu \psi_n \rangle) \\
= (\langle \partial_\mu \psi_n | e^{-i\alpha} - (\langle \psi_n | i \partial_\mu \alpha e^{-i\alpha} | \psi_n \rangle + e^{i\alpha} | \partial_\nu \psi_n \rangle) \\
= (\langle \partial_\mu \psi_n | - \langle \psi_n | i \partial_\mu \alpha | \psi_n \rangle + | \partial_\nu \psi_n \rangle) \quad (5.28) \]

From this expression, it is clear that the effect of the phase \( u \), or its generator \( \alpha \), lives only in the subspace spanned by the \( n \)th basis vector. If we modify our expression of the QGT, to project out this subspace inside the inner product, the phase will no longer feature. Define:

\[ P^1_n := 1 - |\psi_n\rangle \langle \psi_n | = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \quad (5.29) \]
The nomenclature references one (1) dimension being projected out, namely the \( n^{th} \). Then if we define our new QGT as follows:

\[
Q_{\mu \nu} := (\partial_\mu \Psi_n | P_1 | \partial_\nu \Psi_n) = \sum_{j \neq n} \partial_\mu c_j^* \partial_\nu c_j
\]  

(5.30)

Where the \( \{c_j\} \) are specifically the elements in \(|\psi_n\rangle\), we have recovered the gauge invariant form.

### 5.2.2 Degenerate Case

In this language, the generalization to the degenerate case is straightforward. The occupied wave vector, which we will still dub \(|\psi_n\rangle\), is now subtle. The positions corresponding to the dimensions of the degenerate subspace \( H_{E_n} \) are the only ones that are occupied, the rest vanishing. Formally:

\[
c_j = 0 \quad \forall j \notin \{m, m+1, \ldots, m+k-1\},
\]

and the remaining \( c_l, l \in \{m, m+1, \ldots, m+k-1\} \) satisfy the normalization condition \( \sum_l |c_l|^2 = 1 \). Conventionally, \( l \) will label the degenerate dimensions. Then our state has the form:

\[
|\psi_n\rangle = \begin{pmatrix}
0 \\
\vdots \\
0 \\
c_m \\
c_{m+1} \\
\vdots \\
c_{m+k-1} \\
0 \\
\vdots \\
0
\end{pmatrix}
\]  

(5.31)

The particular distribution of the \( c_l \) depends on the choice of basis used to span \( H_{E_n} \), and a priori the system (defined by \( H \)) has no preferred basis. We wish for also \( Q_{\mu \nu} \) to reflect this, and not depend on the basis. We must however start from some arbitrary basis of all of \( \mathcal{H} \), defined locally at each \( \lambda \). A subset of this, denoted \( \{|\phi_l\rangle\} \), are the basis vectors that span the subspace \( H_{E_n} \), so that \(|\psi_n\rangle = \sum_l c_l |\phi_l\rangle\).

Then any other \( k \)-dimensional basis \( \{|\phi'_l\rangle\} \) is obtained from this one by means of a unitary basis transformation \( W \in \mathcal{U}(k) \). With intuitive abuse of summed index \( h: |\phi'_l\rangle = \sum_h W_{jh} |\phi_h\rangle \). To have this work on the entire basis, we must augment our transformation, in a manner that leaves all the other dimensions invariant. To this end, \( W \) is nestled in a larger matrix \( U \in \mathcal{U}(d) \), so that the 1\(st\) row of \( W \) is in the \( m^{th} \) row of \( U \), and likewise for columns. Outside the dimensions \( \{m, \ldots, m+k-1\} \), \( U \) is the identity matrix. This way, \( U \) acts nontrivially only upon the degenerate subspace. The resulting block diagonal structure is the following:
This is a generic element of the gauge symmetry group of our system. We may express the state equally well as:

$$|\psi'_n\rangle = U |\psi_n\rangle$$ (5.33)

So long as $U$, which may vary from point to point in $M$, is always of the form in 5.32, it is simply a matter of representation, and in particular the choice should not effect the value of $Q_{\mu\nu}$.

The attentive reader may point out that we must also consider a scalar global phase $u$ as in the nondegenerate 5.26 multiplying the entire expression 5.32. However, it will appear, fixing this $U$ symmetry results in something manifestly invariant under the global phase $u$ as well.

Now we again consider the ansatz

$$Q_{\mu\nu} := \langle \partial_\mu \psi'_n | \partial_\nu \psi'_n \rangle = (\langle \partial_\mu \psi_n | U^\dagger + \langle \psi_n | \partial_\mu U^\dagger \rangle (\partial_\nu U | \psi_n) + U | \partial_\nu \psi_n \rangle) \quad (5.34)$$

And mimic the derivation above, although we must be more careful due to the noncommuting nature of $U$, not a multiple of the identity. We may still substitute $U$ in exponential form, generated by a $d \times d$ Hermitian matrix $A$ thusly: $U = e^{iA}$. In our particular case, we make use of the $Y$ generating $W = e^{iY}$ in $k$ dimensions, to find the form:

$$A = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \quad (5.35)$$

From the definition, $(e^{iA})^\dagger = e^{-iA} = e^{-iA}, \partial_\mu e^{iA} = i\partial_\mu A \cdot e^{iA}$, and $\partial_\mu (U^\dagger) = (\partial_\mu U)^\dagger$, because $\lambda$ is real. Substituting these identities:

$$Q_{\mu\nu} = (\langle \partial_\mu \psi_n | e^{-iA} - \langle \psi_n | ie^{-iA} \partial_\mu A^\dagger \rangle (i\partial_\nu A e^{iA} | \psi_n) + e^{iA} | \partial_\nu \psi_n \rangle) \quad (5.36)$$
The crucial observation is that the elements of $A$ outside the $k$ columns and rows of $H_{E_0}$ are always zero, or constant, for any $\lambda$: at any point in $\mathcal{M}$, the degeneracy only allows us to mix these $k$ basis vectors. This means that the same is true of the derivative $\partial_\mu A$. This matrix has the same form as $A$, nonvanishing elements only in the diagonal block between rows and columns $m$ and $m + k - 1$. For a $Q_{\mu\nu}$ independent of $A$, we must find a projector in the bracket that will exactly kill the terms involving $A$. A little thought assures us that the smallest amount of information we can project out, is exactly these columns and rows. We posit the projector $P^k_n$:

$$P^k_n := 1 - \sum_l |\phi_l\rangle \langle \phi_l| = \begin{pmatrix} 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{pmatrix}$$

(5.37)

In 5.37 the $d \times d$ identity matrix is modified with $k$ zeroes ($\emptyset$) on the diagonal elements $m$ through ($m + k - 1$). The following are immediate:

$$P^k_n U = P^k_n = UP^k_n = (P^k_n)^2, \quad P^k_n \partial_\mu A = \emptyset = \partial_\mu AP^k_n$$

(5.38)

Then, modifying 5.34, we propose:

$$Q_{\mu\nu} := \langle \partial_\mu \psi_n' | P_k | \partial_\nu \psi_n' \rangle = (\langle \partial_\mu \psi_n | e^{-iA} - (\psi_n | i e^{-iA} \partial_\mu e^{iA} | \psi_n) + e^{iA} | \partial_\nu \psi_n \rangle)
= (\langle \partial_\mu \psi_n | P_k - (\psi_n | i e^{-iA} \emptyset)(i e^{iA} | \psi_n) + P_k | \partial_\nu \psi_n \rangle)
= \langle \partial_\mu \psi_n | P_k | \partial_\nu \psi_n \rangle = \sum_{j < m, j \geq m + k} \partial_\mu c_j^* \partial_\nu c_j$$

(5.39)

Which is again manifestly gauge invariant, also for global QM phase $u$. In conclusion we gather that for the QGT evaluated at some energy $E_n$, the entire subspace corresponding to that energy must be projected out, be it one or $k$-dimensional.

### 5.3 Curvature Operator

In this section, we take a detour to calculate the general form of the Berry curvature, after Michael Berry [16], valid for both the degenerate and nondegenerate cases. Following [45], we define this curvature as:
\[ F_{\mu\nu} := 2 \cdot \text{Im}(Q_{\mu\nu}) = Q_{\mu\nu} - Q_{\mu\nu} \]  
(5.40)

We wish to show how this curvature is formed by the relevant connection, abelian or not. To this end, we assume we are in a state with energy \( E_n \) defined by the initial condition \( \vec{b} \in \mathbb{C}^N \) with |\( \vec{b} \)| = 1. In the nondegenerate case, the elements of \( \vec{b} \) satisfy \( b_j = \delta_{n,j} \). In the degenerate case, the nonzero components of \( \vec{b} \) are spread over the \( k_n \) basis vectors that span \( \mathcal{H}_{E_n} \). Then, initially, our state (dropping the index \( n \)) is given by

\[ |\psi\rangle = \sum_j b_j |\phi_j\rangle \]

and trivially \( b_j = \langle \phi_j | \psi \rangle \).

It is important to note that in this section the \( \{b_j\}\) give the initial distribution among basis vectors, but they do not vary with \( \lambda \).

The abelian Berry connection operator, or non-abelian Wilczek-Zee [17] connection living in Hilbert-Schmidt space, are jointly defined as follows:

\[ A_\mu := \sum_{j,k} |\phi_j\rangle \langle \phi_j | \partial_\mu |\phi_k\rangle \langle \phi_k|, \quad A^{jk}_\mu = \langle \phi_j | \partial_\mu \phi_k \rangle \]

Here, summation is only over the occupied (degenerate) subspace, other elements are zero. This construction is sensible, for the relation:

\[ \langle \psi | \partial_\mu \psi \rangle = \left( \sum_j b_j^* \langle \phi_j | \right) \partial_\mu \left( \sum_k b_k |\phi_k\rangle \right) = \sum_{jk} \langle \psi | \phi_j \rangle \langle \phi_j | \partial_\mu \phi_k \rangle \langle \phi_k | \psi \rangle = \langle \psi | A_\mu | \psi \rangle \]  
(5.42)

In this spirit, the full \( F_{\mu\nu} \) can be expressed as an observable in terms of \( A_\mu \).

From 5.40:

\[ F_{\mu\nu} = \langle \partial_\mu \psi | \left( 1 - \sum_l |\phi_l\rangle \langle \phi_l| \right) |\partial_\nu \psi \rangle - \langle \partial_\nu \psi | \left( 1 - \sum_l |\phi_l\rangle \langle \phi_l| \right) |\partial_\mu \psi \rangle = \langle \partial_\mu \psi | \partial_\nu \psi \rangle - \langle \partial_\nu \psi | \partial_\mu \psi \rangle - \sum_l \left( \langle \partial_\mu \psi | \phi_l \rangle \langle \phi_l | \partial_\nu \psi \rangle - \langle \partial_\nu \psi | \phi_l \rangle \langle \phi_l | \partial_\mu \psi \rangle \right) \]

(5.43)

Where again the summation is over the (degenerate) occupied subspace only. Evaluating each antisymmetric half of this equation separately:

65
\[ \langle \partial_\mu \psi | \partial_\nu \psi \rangle - \langle \partial_\mu \psi | \partial_\nu \psi \rangle = \sum_{jk} (b_j^* \langle \partial_\mu \phi_j | b_k | \partial_\nu \phi_k \rangle - b_j^* \langle \partial_\nu \phi_j | b_k | \partial_\mu \phi_k \rangle) \]

\[ = \langle \psi | \sum_{jk} \langle \partial_\mu \phi_j | \partial_\nu \phi_k \rangle \langle \partial_\nu \phi_j | \partial_\mu \phi_k \rangle \langle \phi_k | \phi_j \rangle \psi \rangle \]

\[ = \langle \psi | \sum_{jk} \langle \partial_\mu \phi_j | \partial_\nu \phi_k \rangle + \langle \phi_j | \partial_\mu \partial_\nu \phi_k \rangle - \langle \partial_\nu \phi_j | \partial_\mu \phi_k \rangle \rangle \psi \rangle \]

\[ = \langle \psi | (\partial_\mu A_\nu - \partial_\nu A_\mu) | \psi \rangle \]

\[(5.44)\]

We anticipated the product rule in the 3rd equality, and used the commuting attribute of derivatives. Next,

\[ - \sum_l \left( \langle \partial_\mu \psi | \phi_l \rangle \langle \phi_l | \partial_\nu \psi \rangle - \langle \partial_\nu \psi | \phi_l \rangle \langle \phi_l | \partial_\mu \psi \rangle \right) \]

\[ = \langle \psi | \left( \sum_{jk} \langle \partial_\mu \phi_j | \partial_\nu \phi_k \rangle \langle \partial_\nu \phi_j | \partial_\mu \phi_k \rangle \langle \phi_k | \phi_j \rangle \psi \rangle \right) \]

\[ = \langle \psi | \left( \sum_{jlmk} \langle \partial_\mu \phi_j | \phi_l \rangle \langle \phi_l | \phi_m \rangle \langle \phi_m | \partial_\nu \phi_k \rangle \langle \partial_\nu \phi_k | \phi_j \rangle \psi \rangle - \langle \phi_j | \partial_\mu \phi_k | \psi \rangle \langle \phi_k | \partial_\nu \phi_m | \psi \rangle \langle \phi_m | \psi \rangle \right) \]

\[ = \langle \psi | (A_\mu^\dagger A_\nu - A_\nu^\dagger A_\mu) | \psi \rangle \]

\[(5.45)\]

Where we insert a summed over delta function in the second equality. Expanding on this theme for the final step, we note that the local bases are always chosen such that

\[ \langle \phi_j | \phi_k \rangle = \delta_{jk} \Leftrightarrow \partial_\mu \langle \phi_j | \phi_k \rangle = \langle \partial_\mu | \phi_j \rangle \langle \phi_j | \phi_k \rangle + \langle \phi_j | \partial_\mu \phi_k \rangle = 0 \]

\[(5.46)\]

In terms of the connection, \( A_{\mu}^{jk} = -A_{kj}^{\mu} \), or \( A_{\mu} = -A_{\mu}^\dagger \), is skew-Hermitian. Combining 5.43, 5.44 & 5.45, we find:

\[ F_{\mu\nu} = \langle \psi | (\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]) | \psi \rangle \]

\[(5.47)\]

This is the correct, gauge covariant expression [46]. In the nondegenerate case the commutator vanishes. The occupied subspace is then one-dimensional, and the connection is effectively scalar (\( A_{\mu}^{jk} \) is a matrix with a single nonzero entry, on the diagonal).

There are other scenarios to map out, such as pure states that are superpositions of energy eigenstates, which in turn may or may not be degenerate. We
will skip these, and move on to mixed states and their treatment as a Riemannian manifold.

### 5.4 Mixed States

Recall our discussion of density matrices $\rho$ from subsection 2.3.1. In the following, for convenience, we assume them diagonal in the energy eigenbasis, at least where this is nondegenerate. This is not a strong assumption. Off diagonal terms resulting from superpositions of energy eigenstates with distinct energies $E_j, E_k$, will oscillate with a frequency $(E_j - E_k)/2\pi$, setting $\hbar = 1$. Over time, it is impossible in practice to keep track of the exact phase, and the instantaneous value will be lost: the states decohere. This situation, with random, unknown relative phases between the energy states, is physically indistinguishable from the diagonal ensemble. In a sense, $\rho(t)$ precesses periodically around its time average $\langle \rho \rangle_t$, which is diagonal.

Inside degenerate subspaces, $k_p \times k_p$ matrix sub-blocks may appear on the diagonal, with $k_p$ the degeneracy of the $p^{th}$ energy level. On mixed states, there is a natural generalization of the QGT, or specifically, its real part $g_{\mu\nu}$.

#### 5.4.1 Bures Metric

In this subsection, we will introduce a metric on the manifold of mixed state density matrices. It is called the Bures metric, after Donald Bures [20].

First we introduce the Bures distance $d_B$, directly related to the fidelity [21], introduced in 2.30.

$$d_B^2(\rho, \eta) := 2 - 2F(\rho, \eta) \quad (5.48)$$

A metric results by taking the Bures distance between infinitesimally separated states $\rho, \eta = \rho + d\rho$. Then, the infinitesimal version becomes:

$$d_B^2(\rho, \rho + d\rho) = \frac{1}{2} \Tr(d\rho G), \quad G\rho + \rho G = d\rho \quad (5.49)$$

Involving the Hermitian one-form only defined implicitly through the second equation.

Mirroring the sections above, we assume there is some mapping from $\lambda \mapsto \rho(\lambda)$ through the Hamiltonian. Then expand the Taylor series:

$$\rho + d\rho = \rho + \partial_\tau \rho d\lambda^\tau + \frac{1}{2} \partial_\mu \partial_\nu \rho d\lambda^\mu d\lambda^\nu + \ldots \quad (5.50)$$

We will see that the first order term cancels due to the normalization condition on the basis states. Then the first nonvanishing dependence on $d\lambda$ is quadratic, and we curtail the higher order terms. What is left is of the form of a metric. We cite from [14]:

67
\begin{align}
  d^2_G(\rho, \rho + d\rho) &= \frac{1}{2} \text{Tr}(\partial_\mu \rho L_\nu) d\lambda^\mu d\lambda^\nu, \quad \frac{\rho L_{\mu\nu} + L_{\nu\mu}}{2} = \partial_{\nu\mu} \rho \tag{5.51}
\end{align}

Here, generalizing \( G \), \( L_\nu \) is the symmetric logarithmic derivative.

In order to recover the link to the old metric \( g_{\mu\nu} \), we must take \( \rho = \langle \psi | \psi \rangle \) pure. We first use expression 2.29 for the fidelity on pure states, and notationally take the differential \( d |\psi\rangle \) inside the ket:

\begin{align}
  F(\langle \psi | \psi + d\psi \rangle) &= |\langle \psi | \psi + d\psi \rangle| = \sqrt{\langle \psi | \psi + d\psi \rangle \langle \psi + d\psi | \psi \rangle} \tag{5.52}
\end{align}

With \( \langle \psi | \psi \rangle = 1 \), we find:

\begin{align}
  F(\langle \psi | \psi + d\psi \rangle) &= \sqrt{1 + \langle \psi | d\psi \rangle + \langle d\psi | \psi \rangle} \tag{5.53}
\end{align}

We now specify:

\begin{align}
  d |\psi\rangle &= |\partial_\tau \psi \rangle d\lambda^\tau + \frac{1}{2} |\partial_\mu \partial_\nu \psi \rangle d\lambda^\mu d\lambda^\nu + \ldots \tag{5.54}
\end{align}

Inserting 5.54 into 5.53, and keeping terms up to quadratic in \( d\lambda \):

\begin{align}
  F(\langle \psi | \psi + d\psi \rangle) &= \left[ 1 + \left( \langle \psi | \partial_\tau \psi \rangle + \langle \partial_\tau \psi | \psi \rangle \right) d\lambda^\tau + \right.
\end{align}

\begin{align}
  &\left. \left( \frac{1}{2} \langle \psi | \partial_\mu \partial_\nu \psi \rangle + \frac{1}{2} \langle \partial_\mu \partial_\nu | \psi \rangle + \langle \partial_\mu | \psi \rangle \langle \psi | \partial_\nu \psi \rangle + \langle \partial_\nu | \psi \rangle \langle \psi | \partial_\mu \psi \rangle \right) d\lambda^\mu d\lambda^\nu \right]^{\frac{1}{2}} = \tag{5.55}
\end{align}

\begin{align}
  \left[ 1 + \left( \langle \partial_\mu | \psi \rangle \langle \psi | \partial_\nu \psi \rangle - \frac{1}{2} \langle \partial_\mu | \partial_\nu \psi \rangle - \frac{1}{2} \langle \partial_\nu | \partial_\mu \psi \rangle \right) d\lambda^\mu d\lambda^\nu \right]^{\frac{1}{2}}
\end{align}

In the last line, we used the following identities:

\begin{align}
  \langle \psi | \partial_\tau \psi \rangle + \langle \partial_\tau \psi | \psi \rangle = \partial_\tau \langle \psi | \psi \rangle = \partial_\tau 1 = 0 \tag{5.56}
\end{align}

and:

\begin{align}
  \langle \psi | \partial_\mu \partial_\nu \psi \rangle + \langle \partial_\mu \partial_\nu | \psi \rangle + \langle \partial_\mu | \psi \rangle \langle \psi | \partial_\nu \psi \rangle + \langle \partial_\nu | \psi \rangle \langle \psi | \partial_\mu \psi \rangle = \partial_\mu \partial_\nu \langle \psi | \psi \rangle = 0 \tag{5.57}
\end{align}

From 5.55, we may un-symmetrize the negative terms using the symmetric \( d\lambda^\mu d\lambda^\nu \), to obtain:
$F(\psi, \psi + d\psi) = \left[ 1 + \left( \langle \partial_\mu \psi | \psi \rangle \langle \psi | \partial_\nu \psi \rangle - \langle \partial_\mu \psi | \partial_\nu \psi \rangle \right) d\lambda^\mu d\lambda^\nu \right]^{\frac{1}{2}}$

$= \left[ 1 + \langle \partial_\mu | \psi \rangle \left( | \psi \rangle \langle \psi | -1 \right) | \partial_\nu \psi \rangle d\lambda^\mu d\lambda^\nu \right]^{\frac{1}{2}} = \left[ 1 - g_{\mu\nu} d\lambda^\mu d\lambda^\nu \right]^{\frac{1}{2}} \approx 1 - \frac{1}{2} g_{\mu\nu} d\lambda^\mu d\lambda^\nu \quad (5.58)$

Using 5.12 and 5.18, and the Maclaurin series of $\sqrt{1-x} = 1 - \frac{1}{2} x + \ldots$ Then from 5.48, the relation is immediate:

$\approx 1 - \frac{1}{2} g_{\mu\nu} d\lambda^\mu d\lambda^\nu \quad (5.59)$

This is one of the attractive qualities of the Bures metric. It reduces nicely to the quantum geometric tensor in the special case of pure states, i.e. it is Fubini-Study adjusted. Also, it is Fisher adjusted, coinciding with the Fisher-Rao metric of classic probability distributions when restricted to diagonal ensembles. On non-diagonal density matrices, it additionally takes into account the purely quantum phenomenon of noncommutativity between $\rho$ and $\rho + d\rho$ [19]. Furthermore, as long as both arguments are in the same one, it is independent of the basis, through the that property of the fidelity.

### 5.4.2 Gauge Invariance of the Bures Metric

When constructing density matrices, all elements are the product of a bra and a ket vector. Any abelian global phase is annihilated this way, so any function of density matrices is automatically invariant under quantum $U(1)$ global phase. The same can not be said for larger gauge symmetries.

Let us assume the spectrum is multiply degenerate, consisting of $L$ energy levels $E_n$ for $n \in \{1, 2, \ldots, L\}$. These levels have degeneracies $k_n$, forcing $\sum_n k_n = d = \text{dim}(\mathcal{H})$. Construct $\rho$ in an energy eigenbasis, which of course is not unique. An element of the total gauge symmetry is given by a product of non-abelian unitary group elements:

$U = \prod_n U_n, \quad U_n \in U(k_n) \quad (5.60)$

In matrix notation, in this basis, it is clear that each $U_n$ only works on the basis vectors $\{\phi_l\}$ spanning $\mathcal{H}_{E_n}$, so the factors commute.
The reader reminded that the elements \((U_n)\) above are \(k_n \times k_n\) block submatrices.

As before, we wish for the Bures metric to be gauge invariant under this larger gauge symmetry, which is equivalent to demanding this of the fidelity between infinitesimally separated mixed states. A local (in \(\mathcal{M}\)) gauge transformation is given by \(|\phi_l(\lambda)\rangle \mapsto U(\lambda)|\phi_l(\lambda)\rangle\) (we will again suppress the argument).

In the language of density matrices, this says the gauge transformation is given by 
\[
\rho \mapsto U\rho U^\dagger.
\]

We now consider the ramifications for the fidelity. Inspired by [19], we take equation 2.30:

\[
\mathcal{F}(\rho, \rho + \delta\rho) = \text{Tr} \left( \sqrt{\rho} (\rho + \delta\rho) \sqrt{\rho} \right)
\]

\[
\mapsto \text{Tr} \left( \sqrt{U\rho U^\dagger} (U\rho U^\dagger + \delta(U\rho U^\dagger)) \sqrt{U\rho U^\dagger} \right)
\]

\[
= \text{Tr} \left( U \sqrt{\rho} U^\dagger (U\rho U^\dagger + \delta U\rho U^\dagger + U\delta U^\dagger + U\delta U^\dagger U^\dagger) U \sqrt{\rho} U^\dagger \right)
\]

\[
= \text{Tr} U \sqrt{\rho} (U + U^\dagger \delta U + \delta \rho - \rho U^\dagger \delta U) \sqrt{\rho} U^\dagger
\]

\[
= \text{Tr} \left( \sqrt{\rho} (\rho + \delta \rho + [U^\dagger \delta U, \rho]) \sqrt{\rho} \right)
\]

(5.62)

Using the fact that for any diagonalizable matrix \(A\), we have \((U^\dagger \sqrt{A} U)^2 = U^\dagger A U \leftrightarrow \sqrt{U^\dagger A U} = U^\dagger \sqrt{A} U\), thus we may bring the unitary conjugations outside the square roots. We are only interested in positive roots of positive operators. Then the cyclic property of the trace allows us to kill the outer unitary elements, and finally we use \(\delta(U^\dagger U) = \delta I = 0 \iff \delta U^\dagger U = -U^\dagger \delta U\). It is evident that the commutator term in 5.62 is gauge dependent, and we would like to remove it. It is also immediate, that in the case where \(U\) is a scalar global phase, then so is its variation, thus the fidelity is gauge invariant once more.

Intuitively, one might like to integrate out the gauge freedom of the choice of basis for each separate degenerate subspace. Unfortunately, integrating over the unitary group often results in vanishing elements, by symmetry, if it can be
performed at all. Alternatively then, we decide to coarse-grain the Hilbert space, in order to have every column or row of the density matrix represent an entire energy eigenspace. One proposal by the author is to pre-process the density matrix this way. In the same basis, we introduce the sub-block structure of $\rho$, each sub-block living in the tensor product of at most two energy eigenspaces:

$$
\rho = \begin{pmatrix}
R_{11} & R_{12} & \cdots & R_{1L} \\
R_{21} & R_{22} & \cdots & R_{2L} \\
\vdots & \vdots & \ddots & \vdots \\
R_{L1} & R_{L2} & \cdots & R_{LL}
\end{pmatrix}
$$

(5.63)

Where $R_{jk}$ is a $k_j \times k_k$ submatrix of $\rho$, and $R_{jk} = R_{kj}^\dagger$. Consider the action of $U$ in 5.61 on $\rho$ in this form:

$$
\rho \mapsto U \rho U^\dagger = \begin{pmatrix}
U_1 R_{11} U_1^\dagger & U_1 R_{12} U_2^\dagger & \cdots & U_1 R_{1L} U_L^\dagger \\
U_2 R_{21} U_1^\dagger & U_2 R_{22} U_2^\dagger & \cdots & U_2 R_{2L} U_L^\dagger \\
\vdots & \vdots & \ddots & \vdots \\
U_L R_{L1} U_1^\dagger & U_L R_{L2} U_2^\dagger & \cdots & U_L R_{LL} U_L^\dagger
\end{pmatrix}
$$

(5.64)

One of the defining qualities of a unitary transformation is that it preserves the norm of any vector it acts on. If it acts on a matrix from the left, it preserves the norm of the columns, and from the right of the rows. The off-diagonal blocks of 5.64, are conjugated by different, arbitrary unitaries, so the only obvious conserved quantity within this subspace is the total sub-block matrix norm.

$$
||R_{jk}|| := \sqrt{\text{Tr} \left( R_{jk} R_{kj}^\dagger \right)} = \sqrt{\text{Tr} \left( R_{jk} R_{kj} \right)}
$$

(5.65)

This is analogous to taking the norm of each column in a single row vector, and taking the norm of that row. The norm is manifestly invariant under the gauge symmetry:

$$
||U_j R_{jk} U_j^\dagger|| = \sqrt{\text{Tr} \left( U_j R_{jk} U_j^\dagger U_k R_{kj} U_k^\dagger \right)} = \sqrt{\text{Tr} \left( R_{jk} R_{kj}^\dagger \right)} \equiv ||R_{jk}||
$$

(5.66)

The blocks along the diagonal are square, and are conjugated by the same unitary matrix left and right. All their eigenvalues are conserved under the symmetry. A coarse-grained density matrix, with only one dimension for each energy eigenspace, is proposed as follows:

$$
\rho \mapsto \rho_C = \begin{pmatrix}
\text{Tr} R_{11} & ||R_{12}|| & \cdots & ||R_{1L}|| \\
||R_{21}|| & \text{Tr} R_{22} & \cdots & ||R_{2L}|| \\
\vdots & \vdots & \ddots & \vdots \\
||R_{L1}|| & ||R_{L2}|| & \cdots & \text{Tr} R_{LL}
\end{pmatrix}
$$

(5.67)

A number of remarks: This transformation is only well posed in the energy eigenbasis, or more generally in a basis in which the gauge symmetry is defined
like that in 5.61. Much information about the relative phases between energy eigenstates is lost. For instance, $\rho_C$ is always real and positive. This does afford it the desirable quality of uniqueness: there is only one coarse-grained version of any density matrix. Furthermore:

- $\rho_C \in \mathbb{R}^{L \times L}$.
- $\rho_C$ is symmetric, and with reality then Hermitian: $\rho_C = \rho_C^T = \rho_C^\dagger$.
- One quickly verifies that $\text{Tr} \, \rho_C = \text{Tr} \, \rho = 1$.
- In order to be a well defined density matrix it must additionally be positive semidefinite: $z^\dagger \rho_C z \geq 0 \forall z \in \mathbb{C}^L$. This is equivalent to it having nonnegative eigenvalues [14]. Although the author has not been able to prove this final attribute rigourously, it is conjectured and verified in numerical tests of 1000 of randomly generated density matrices $\rho$, reduced by any combination of degeneracies. None of the calculated coarse-grained matrices are found to have negative eigenvalues.
- Most importantly, it is manifestly invariant under the action of gauge group $U$ on $\rho$.

In general, the purity $\gamma(\rho_C) \neq \gamma(\rho)$, will not be preserved, but is expected to increase, as the (nonzero) eigenvalues change. This is obvious in the case of a diagonal $\rho$. Perhaps calculating the fidelity $F(\rho_C, \rho_C + d\rho_C)$ is more robust against this gauge group, although this brings other obvious complications to life. Unfortunately, we will not go so far as to use the QGT or Bures metric to tackle thermalization ourselves within the scope of this work. For that, we refer to Polkovnikov and Kolodrubetz [25].
Chapter 6

Conclusions and Outlook

6.1 Summary

In summary, we have expounded the theory needed to interpret information loss in quantum mechanics. We introduced a number of useful geometric notions and expressions for quantifying distance between states, notably the fidelity, Quantum Geometric Tensor (QGT), and Bures metric. Other formulas assess ignorance of the reduced subsystem state, these are the purity and entropy. We have considered a number of properties of the QGT, such as its behavior under abelian, and later non-abelian gauge symmetries. These can come about in the case of a degenerate spectrum. The mixed state generalization of the QGT is the Bures metric, and it was found not to be well behaved under non-abelian gauge transformation. A step towards resolving this may have been proposed with the coarse-grained density matrix. Moving to simulations, we shed light on the typical behavior of a quantum system with respect to its subsystems. Typical, in the sense of being drawn uniformly from the possible sets of eigenvectors, a 'uniform ensemble'. We saw that subsystems with less degrees of freedom than their baths very quickly are locally described best by a state approaching the completely mixed one, an effect increasing with bath size. This meant that for most times, no local experiment can tell the subsystem apart, one time from another. In order to reach this conclusion, we evaluated the average fidelity between the subsystem at one time and others. The fidelity is uniquely driven by distinguishability through measurements, and is a natural measure in this sense. This is a strong case for equilibration, a step towards thermalization: at later times, measurements on the subsystem will be constant. Conversely, a product initial state is typically distinguishable from a later time state under coupled evolution. This indicates that such an unentangled system will evolve away from its statistically exceptional initial product state, and seldom return. Moreover, we saw that the temporal variance from the mean was typically very low, indicating that the averages give a good estimate of the instantaneous situation. Finally, the time averaged entanglement entropy appeared to approach
the maximum, as bath size was increased: the typical system maximally mixes its subsystems.

Continuing to analytics, we found it possible to integrate functions of subsystem reduced density matrices over all possible Hamiltonian eigenbases in the unitary group. This yielded an insight into the ‘descent into disorder’ of the average system from the uniform ensemble. The resulting functions feature only Hilbert space dimensions, and a single, albeit intricate, function of time and the system’s spectrum. The first result is the average density matrix itself, at any time of its evolution, $\bar{\rho}(t)$. Starting in a product state, in this expression we saw the initial condition compete with the maximally mixed state. For all times, traces of the initial information persisted, until the bath was taken infinite, and statistically it was diluted so far the initial information becomes impossible to recover. After this, an expression was derived for the matrix-valued variance over the unitary group of the density matrix. Its elements decay at least at the same rate as the amplitude of the average density matrix elements themselves. Finally, the scalar purity was averaged, $\bar{\gamma}(t)$, showing the degree of mixing over time, and the typical transition from ‘well known’ to ‘poorly known’ phase of the subsystem. It was observed that even at infinite times, for the extremal ratio of subsystem to bath, the system tends to be twice as ‘pure’ as the maximally mixed state.

The underlying philosophy in this work is that the quantum system has no knowledge of- or investment in- the experimenter’s sense of locality. The partition into subsystem and bath is ours, and highly artificial. In a mathematical sense, system doesn’t have any preference a priori, which degrees of freedom to couple more or less. In fact, entanglement is only defined after we choose a basis in which to form superpositions. We proceed to judge locality by its likeness to our prescription of a good, spatial basis, reflecting our own spatially localized nature. Another legitimate basis might be that of Fourier modes, but there are infinitely (though compactly) many. One goal of this report was to make basis-independent statements about thermalization and equilibration. In this endeavor, one necessarily obtains results so general and symmetric they no longer describe our world closely, from which they aim to detach themselves.

### 6.2 Outlook

In order to reinstate reality, an important next step is to revisit the simulations and calculations with a nonuniform distribution of eigenvectors over the unitary group. Think of increasing correlation between adjacency of particles or spins, and their coupling strength. An integral weighing the systems by their Hamiltonian eigenbasis’s distance to some preferred one is straightforward to construct, unfortunately not to evaluate. The following is a potential not invariant under the choice of basis $V$:

$$P(V) = C \exp \left( -\lambda \cdot \text{Tr} \left( [V, \Omega][V, \Omega]^{\dagger} \right) \right), \quad \mathbb{1} \neq \Omega \in \mathcal{U}(d) \quad (6.1)$$

Here $\Omega$ is some preferred basis, $\lambda$ controls the strength of the weighting, and
$C$ is a constant insuring $\int_{\mathcal{U}(d)} d\mathcal{V}^2 = 1$. As the traced expression is of the
form $AA^\dagger$, it is positive semidefinite, giving nonnegative trace. It vanishes for $V$ being a basis commuting with $\Omega$, maximizing the probability. An obvious choice to promote the computational basis is a diagonal $\Omega$, though $\Omega$ must be nondegenerate so it was proposed to take elements $\Omega_{jj} := e^{2\pi ij}$. This can be further generalized by adding potential terms of the form $\text{Tr} \left( f(\Omega) \right)$ for some matrix function $f$. Numerically, this is accessible, but to repeat the analytic integrals of section 4, the exact integrands are powers of elements of $V$. The best that can be done is a perturbative expansion in the sum defining the exponent.

Concerning the end of chapter 5, this story is admittedly unfinished. One idea, coarse-graining the density matrix, was proposed to remedy the Bures metric’s transformation properties, but by no means have we made conclusive arguments that it is the best route. Our main result is pointing out the problem.

Returning to Random Matrix Theory, the elegance of this framework allowed us to decouple the statistics of eigenvalues and eigenvectors in treating systems and their prevalence. We chose to focus on the eigenvectors, as they ultimately determine what can happen in a system, where the eigenvalues determine mostly how fast it happens. That is not to say all the interesting behaviour is in the vectors and there is still much to be said in closing this gap. Picking up after sections 3 and 4, one can also impose a global distribution $P(\lambda)$ on the energies, on top of the Vandermonde determinant resulting from a coordinate transformation. This makes statements on the expected equilibration times more rigorous. Besides this, one can think of restricting initial conditions to a certain energy shell or band, as is often done in work mimicking Von Neumann [10], and attempt to recover the microcanonical ensemble more closely. Finally, when taking infinite size, or thermodynamic limits, the concept of temperature can be added, allowing for Gibbs ensembles, involving other kinds of entropy, etc. A further wealth of rich and interesting physics may be discovered.
Appendix A

Fidelity Plots

Here we have collected the plots deemed superfluous to the narrative in chapter 3.

Figure A.1: Distribution of 4000 random Hamiltonians: Their time average fidelity to the initial pure state, $\langle F_0 \rangle$, for a system of $n = 3, 4, 5$ spins, and $n_S = 1$ spin. Surface under each plot is normalized to one.
Figure A.2: Distribution of 4000 random Hamiltonians: Their time average fidelity to the initial pure state, $\langle F_0 \rangle$, for a system of $n = 6, 7, 8$ spins, and $n_S = 1$ spin. Surface under each plot is normalized to one.

Figure A.3: Distribution of 4000 random Hamiltonians: Their time average fidelity between all times, $\langle F_\infty \rangle$, for a system of $n = 3, 4, 5$ spins, and $n_S = 2$ spins. The surface under each plot is normalized to one.
Figure A.4: Distribution of 4000 random Hamiltonians: Their time average fidelity between all times, $\langle F_\infty \rangle$, for a system of $n = 6, 7, 8$ spins, and $n_S = 2$ spins. The surface under each plot is normalized to one.

Figure A.5: Distribution of 4000 random Hamiltonians: Their temporal variance of the fidelity between different times, $\sigma_\infty^2$, for a system of $n = 3, 4, 5$ spins, and $n_S = 2$ spins. The surface under each plot is normalized to one.
Figure A.6: Distribution of 4000 random Hamiltonians: Their time averaged normalized entropy $\langle S_1 \rangle$, for $n = 3, 4, 5, 6, 7, 8$ and $n_S = 1$. The entropy is scaled by the subsystem dimension, such that the maximum entropy is 1, corresponding to $\rho_S \propto \mathbb{1}$. 
Appendix B

Python Code

Preamble: These commands are needed in most of the functions, and should be imported at the beginning of the script.

```python
from numpy import arange, imag as im, cumsum
from numpy import zeros, kron, dot, random as rand, mean
from numpy import exp, log, diag, linalg as lina, arccos
from numpy import array as arr, real as re, linspace, var
from numpy import median, trace as tr, ...
... count_nonzero as hamming
from numpy import cos, sin, eye, pi, reshape as resh
from scipy.linalg import logm, fractional_matrix_power as ...
... frapow
from scipy.linalg import expm, triu, det, norm, svd, sqrtm
from random import sample, seed
from matplotlib import pyplot as plt
from scipy.special import erf, digamma
from scipy.optimize import root
from copy import deepcopy
from scipy.stats.stats import pearsonr
from time import clock
from itertools import permutations
from math import factorial
```

The function ‘compactify’ coarse-grains the density matrix ‘rho’, according to a vector of degeneracies ‘D’. $\sum_{j=1}^{L} D_j = d$, the Hilbert space dimension. See subsection 5.4.2.
def compactify(rho,D):
    
    d=len(D)
    E = [0]
    E.extend(cumsum(D))
    
    rho2 = zeros((d,d))
    
    for n in range(d):
        for m in range(n+1,d):
            sub = rho[E[n]:E[n+1],E[m]:E[m+1]]
            rho2[n,m]=(re(tr(dot(sub,sub.T.conj()))))**0.5
            rho2[m,n]=deepcopy(rho2[n,m])
    
    for n in range(d):
        rho2[n,n]=re(tr(rho[E[n]:E[n+1],E[n]:E[n+1]]))
    
    return rho2

The function 'genrho' generates a random density matrix of dimension 'dim' [14]:

```python
def genrho(dim):
    A = 1j*rand.normal(0,1,(dim,dim))+...
    ...rand.normal(0,1,(dim,dim))
    rho = dot(A,A.T.conj())/tr(dot(A,A.T.conj()))

    rho -= 1j*im(diag(diag(rho)))
    
    return rho
```

The function 'rtr' takes a square matrix 'rho', and traces out the 'bath' rightmost qubits in the computational basis.
The function 'ltr' takes a square matrix 'rho', and traces out the 'subs' leftmost qubits in the computational basis.

```
def ltr(rho,subs):
    m = 2**subs
    n = int(len(rho)/m)
    rho2 = zeros((n,n),dtype=complex)
    for a in range(n):
        for b in range(n):
            for c in range(m):
                rho2[a][b] += rho[a+n*c][b+n*c]
    return rho2
```

The function 'unita' generates a uniformly random unitary matrix of size 'd', according to the method explained in 2.6.1.

```
def unita(dim):
    R = (rand.exponential(1,(dim,dim)))**0.5
    phi = (rand.random((dim,dim)))*2*pi
    Z = R*(cos(phi)+1j*sin(phi))
    q,r = lina.qr(Z)
    lam = eye(dim)*(Z/abs(Z))
    Q = dot(q, lam)
    return Q
```

The function 'puretrace' takes a Hilbert space vector 'psi', turns it into a

82
density matrix, and traces out 'trace' dimensions from the right in the computational basis.

```python
def puretrace(psi, trace):
    A = resh(psi, (int(len(psi)/trace), trace))
    rho = dot(A, A.T.conj())
    return rho
```

The function 'genham' generated a random Hermitian matrix of dimensions 'size' times 'size', with eigenvalues between -pi and pi. The eigenvalues are mutually repulsive, and the eigenvectors are from the Haar measure, as in section 2.6.

```python
def genham(size):
    Hh = 1j*logm(unita(size))
    Hu = triu(Hh, 1)
    H = re(diag(diag(Hh)))+Hu+Hu.T.conj()
    return H
```

The function 'evol' creates the data used in the histograms in chapter 3. It generates 'runs' Hamiltonians (4000 was used) for 'subs' subsystem spins (1 and 2 were used). Output are the mean values \( \langle F_0 \rangle \), \( \langle F_\infty \rangle \), \( \sigma_0 \), variances \( \sigma_\infty^2 \), and \( \sigma_0^2 \), normalized entanglement entropy \( S_1 \), and the minima over time \( \min_t F_0 \) and \( \min_t F_\infty \) in vectors. See section 2.2.
def evol(runs, subs):
    compareinitial = []
    compareall = []
    entro = []
    samplesize = 150

    for sites in range(3, 9):
        clo = clock()
        init = [[], [], []]
        ra = [[], [], []]
        ent = [[], [], []]
        for I in range(runs):
            B = [[], [], []]
            V = unita(2**sites)
            ei = 2*pi*rand.random(2**sites)-pi
            rho1 = puretrace(dot(V, exp(-1j*ei*rand.random()... 1000)*V[0].conj())), 2**(sites-subs))
            for t in range(samplesize):
                rho2 = deepcopy(rho1)
                rho1 = puretrace(dot(V, exp(-1j*ei*... rand.random()**1000)*V[0].conj())), 2**... (sites-subs))
                B[0].append(re(rho1[0][0]**0.5))
                B[1].append(re(tr(sqrtm(dot(sqrtm(rho2)... rho1), sqrtm(rho2)))))
                B[2].append(-re(tr(dot(rho1, logm(rho1))))/... (subs*log(2))))
                init[0].append(mean(B[0]))
                init[1].append(var(B[0]))
                init[2].append(float(min(B[0])))
                ent[0].append(mean(B[2]))
                ent[1].append(var(B[2]))
                ent[2].append(float(min(B[2])))
                ra[0].append(mean(B[1]))
                ra[1].append(var(B[1]))
                ra[2].append(float(min(B[1])))
        compareinitial.append(init)
        compareall.append(ra)
        entro.append(ent)
        print ('seconds elapsed: ', int(clock()-clo), ' for ... ... sites: ', sites)

    return compareinitial, compareall, entro
The function 'conjclass' takes a vector 'A', consisting of the numbers 0 to \( q-1 \), as the result of the mapping from \([0,1,\ldots,q-1]\) by the permutation \( \sigma \in S_q \), as a proxy for \( \sigma \). It returns the cycle lengths of the conjugacy class of \( \sigma \). See subsection 2.7.

```python
def conjclass(A):
    c = []
    a = []

    for n in range(len(A)):
        if n not in a:
            a.append(n)
            t = 1
            m = deepcopy(n)
            while A[m] != n:
                m = A[m]
                a.append(m)
                t += 1
            c.append(t)

    return sorted(c, reverse=True)
```

The function 'finvert' refers to the list 'permutations([0,1,2,3])' as a proxy for the elements of \( S_4 \), and returns the list entry number of the inverse of permutation 'c'.

```python
def finvert(c):
    L = list(permutations([0,1,2,3]))
    A1 = arr(L[c])

    for d in range(24):
        B1 = arr(L[d])
        C1 = A1[B1]
        if all(C1==arr([0,1,2,3])):
            f = deepcopy(d)

    return f
```

The function 'wintegrate2' takes a collection of data in 'data', collected in tables 2 or 4, and prints in the console, term by term, the result of the unitary integrals in chapter 4. If \( \text{pri} \neq 1 \), the output of printed terms is suppressed.
def wintegrate2(data,pri):
    siglabel,tauilabel,siglegend,tauilegend = data
    l = len([j for i in siglegend for j in i])
    h = deepcopy(l)
    dim=1
    while dim < h:
        h /=dim
        dim+=1
        if factorial(dim) != l:
            print('incorrect legend length')
    lens = len(siglabel)
    lent = len(tauilabel)
    lenw = dim + int(dim>3)+int(dim>4)
    L = list(permutations(list(range(dim))))
    final = zeros(lens*lent*lenw)
    for s in range(l):
        svec=[]
        for a in range(lens):
            svec.append(int(s in siglegend[a]))
        for t in range(l):
            tvec = []
            for a in range(lent):
                tvec.append(int(t in tauilegend[a]))
            wvec = zeros(lenw)
            SIG = arr(L[s])
            TAUI = arr(L[t])
            c = conjclass(SIG[TAUI])
            typ = sum(c)-len(c)+int(c[0]>2)+int(c[0]*sum(c)>16)
            wvec[typ]=1
            final += kron(kron(svec,tvec),wvec)
    if pri == 1:
        Wdenom = ['d','d(d^2-1)','d(d^2-1)(d^2-4)','d^2(d^2-1)(d^2-4)'
                  'd^2(d^2-1)(d^2-4)(d^2-9)','d^2(d^2-1)(d^2-4)(d^2-9)
                  (d^2-16)']
        Wenum = [['1'],['d','-1'],['d^2-2','-d','2'],
                  ['d^2-2','-d','2'],
                  'd^4-8d^2+6','-d^3+4d','d^2+6','-5d'],
                  ['d^5-20d^3+78d','-d^4+14d^2+24','d^3-2d','-2d^3-18d
                  '-2d^2-24','-5d^2+24','14d']
        print('denominator =',Wdenom[dim-1],...'nonzero terms coming up')
        for a in range(lens):
            for b in range(lent):
                for w in range(lenw):
                    n = lent*lenw*a+lenw*b+w
                    if final[n] != 0:
                        input('next?')
                        print(final[n],*',siglabel[a],','...,tauilabel[b],','
                        ',Wenum[dim-1][w])
    return final
The function 'storage4' can be called without arguments to serve as the data in 'wintegrate2' for the calculation of the mean dynamical subsystem purity $\bar{\gamma}(t)$.

```python
def storage4():
    siglabel = ['1', 'B', 'dB', 'A', 'dA']
    tauilabel = ['d', 'd^2', 'X', 'X^2', 'dX', 'e2', 'e3', 'e-3']
    siglegend = [[0, 5, 14, 16], [1, 3, 6, 8, 10, 11, 13, 15], [7, 9], ...
                  [2, 4, 12, 17, 19, 20, 21, 22], [18, 23]]
    tauilegend = [[9, 10, 13, 17, 18, 22], [7, 23], [3, 4, 8, 11, 12,...
                  15, 19, 20], [0], [1, 2, 6, 21], [16], [14], [5]]
    return siglabel, tauilabel, siglegend, tauilegend
```

The function 'storagevar' can be called without arguments to serve as the data in 'wintegrate2' for the calculation of the average $\rho^2$.

```python
def storagevar():
    siglabel = ['(1,1)', '(1,1)*B', '(i,1)', '(1,j)', '(i,j)*B', ...
                '(i,i)*B^2']
    tauilabel = ['d', 'd^2', 'X', 'X^2', 'dX', 'e2', 'e3', 'e-3']
    siglegend = [[0, 2, 21, 23], [1, 4, 6, 8, 9, 11, 18, 20],...
                 [12, 14, 17, 19], [3, 5, 15, 22], [13, 16], [7, 10]]
    tauilegend = [[9, 10, 13, 17, 18, 22], [7, 16],...
                   [3, 4, 8, 11, 12, 15, 19, 20], [0], [1, 5, 6, 14], [23], [21], [2]]
    return siglabel, tauilabel, siglegend, tauilegend
```
Bibliography


