Preparing Thermal States on a Digital Quantum Computer

by

Matthew Hagan

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Physics
University of Toronto

Preparing Thermal States on a Digital Quantum Computer

Matthew Hagan
Doctor of Philosophy

Department of Physics
University of Toronto
2025

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

This thesis is dedicated to
my brother JT and my sisters
Veronica and Brittany.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The goal of this thesis is to serve as a blueprint for creating quantum channels that can prepare thermal states of arbitrary systems. This framework was primarily created as an algorithmic process to prepare input states of the form $\rho(\beta) = \frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}$, known as Gibbs, Boltzmann, or thermal states, for simulations on a digital, fault-tolerant quantum computer. As $\rho(\beta)$ approaches the ground state as the inverse temperature β diverges, meaning Gibbs states serve as useful proxies for problems in which ground states dictate features of dynamics, such as the electron wavefunction in banded condensed matter systems or in chemical scenarios. One of the key features of our algorithm is the addition of only one single extra qubit outside of those needed to store the state of quantum system being simulated. This may seem like a minor technical achievement, given the existence of Gibbs samplers that also have only one extra qubit or at worst a constant number of qubits overhead, but the way in which this single qubit is utilized in our algorithm highlights the connections between our channel as an algorithmic tool to prepare states of interest and the model of thermalization that we believe the physical world may actually follow. The rest of this introduction is to provide context for how the technical results contained in later chapters of this thesis contribute to the growing interplay between Physics and Theoretical Computer Science within the realm of Quantum Computing.

The easiest way to grasp the context of this thesis is to first understand classical thermal states. In classical mechanics we typically have access to a Hamiltonian H that is a function on phase space (x, p) to the reals $\mathbb{R}$. We can turn this function into a probability distribution via the canonical ensemble $p_\beta(x, p) = \frac{e^{-\beta H(x,p)}}{\int dx dp e^{-\beta H(x,p)}}$, which can be thought of as the "density"

of particles near a particular (x, p) in phase space for a thermodynamically large collection of non-interacting particles under the same Hamiltonian H. The inverse temperature β, typically taken to be $\frac{1}{k_\beta T}$ where k_β is Boltzmann's constant and T the temperature, plays an important role in shaping the distribution $p_{\beta(x,p)}$. For example, in the $\beta \to \infty$ the distribution becomes concentrated at the minimum energy points of $H(x, p)$, which correspond to points with zero momentum $p = 0$ and are minimums of the potential energy $V(x)$. This shows that the problem of preparing classical thermal states somehow contains'' the problem of function optimization, and in fact being able to sample from classical thermal states for arbitrarily large β allows us to approximate the partition function $\mathcal{Z} = \int e^{-\beta H(x,p)} dx dp$, which is known to be a #P-Hard computational problem.

In order to understand when classical systems can reach thermal equilibrium we need to understand where does the concept of temperature come from? We can use the following thought experiment of two systems A and B that are completely isolated from their surroundings, each initially has some internal energy U_A and U_B. We could define entropies σ_A and σ_B from the distribution of the energy of each particle, averaged over some time window, and declare that an equilibrium has been reached between the two systems when the total entropy is constant

$$d\sigma = d\sigma_A + d\sigma_B = 0. \tag{1.1}$$

Then we know that the total energy of the two systems must remain constant, due to thermal isolation, so $dU = dU_A + dU_B = 0$. However, in this scenario the energy of the system is only a function of the entropy, so we have

$$dU = \frac{\partial U_A}{\partial \sigma_A} d\sigma_A + \frac{\partial U_B}{\partial \sigma_B} d\sigma_B = \left(\frac{\partial U_A}{\partial \sigma_A} - \frac{\partial U_B}{\partial \sigma_B} \right) d\sigma_A = 0. \tag{1.2}$$

This implies that when the two systems are in equilibrium $\frac{\partial U_A}{\partial \sigma_A} - \frac{\partial U_B}{\partial \sigma_B} = 0$ and the property $\frac{\partial U_A}{\partial \sigma_A}$ is equal to that of B. This is a good notion of temperature, so we define $T = 1/\frac{\partial U}{\partial \sigma}$.

The rationale for introducing this experiment is two-fold. On the first, it gives us a very rigorous definition of temperature by isolating exactly the condition meant behind "thermal equilibrium". We find that we can still define temperature even if we are not sure of how the two systems exchange energy. All we really need to know is that the systems are completely isolated and that energy is exchanged *somehow*, and that this energy exchange is captured by the entropy. The second main point of this experiment is that it gives a very concrete way to

thermalize a system to some temperature T: if you have another system at that temperature just put the two into contact! Eventually they will equilibrate to some intermediate temperature and the process can be repeated.

Although the above result tells us that two systems in thermal contact with each other will eventually reach thermal equilibrium it doesn't exactly tell us *how* it does so. Empirically, we have found a few different ways in which two systems can exchange heat:

- They can trade photons at various wavelengths, described by the Plank law, which then get "absorbed" by the various parties. This is how the earth reaches a thermal equilibrium with the sun and the vacuum of space.
- The nuclei of molecules can vibrate, which can then cause nearby nuclei and electrons to oscillate as well through Coulomb's law. This is the process of conduction, and can be extended to rigid crystals via phonon theory.
- Convection allows for hot parts of a liquid or gas to mix with colder parts. This brings the two extremal parts of a system in closer contact and allows for their temperatures to average out quicker.

These three mechanisms are typically sufficient for classical systems to equilibrate. Typically rates of heat exchange are measured empirically and the random microscopic effects are essentially "averaged out" at the macroscopic into a single coefficient of heat transfer.

If one has complete control over some system, such as a refrigerator, and can prepare it in a thermal state of temperature T_{cold}, we can then utilize this to cool down other systems via equilibration. Given our understanding that preparing thermal states is an incredibly challenging problem in the worst case, this raises the question: can we mimick this cooling process algorithmically on digital computers to sample from thermal distributions? If nature is able to cool systems down efficiently, we should be able to do so as well on computer simulations.

When trying to simulate a classical system we can typically compute the energy of the system $H(x, p)$ given a configuration of position and momenta. For example if we are trying to prepare the thermal state for a gas of non-interacting molecules with no background potential then the energy is just $\sum_i p_i^2/2m$. If we encode the positions into standardized 64-bit floating point numbers, computing the energy can be done in order $O(N)$ time, where N is the number of particles simulated. This is referred to as having oracle access to H, once we have a collection

(x, p) we can then compute H as a function call. The challenging problem is to then take this oracle H and output a list of samples $\{s_1, s_2, ..., s_S\}$, where s_i is a pair of position and momenta $s_i = (x_i, p_i)$, that replicate the thermal statistics. We want to use these samples to compute difficult thermal quantities, given some observable $O(x, p)$ (e.g. the velocity of a particle $v = |\vec{p}|$) we would like the following to be as close to each other as possible

$$\frac{1}{S}\sum_{i=1}^{S} O(x_i, p_i)\frac{e^{-\beta H(x_i,p_i)}}{\sum_{j=1}^{S} e^{-\beta H(x_j,p_j)}} \approx \int dxdp\ O(x, p)\frac{e^{-\beta H(x,p)}}{\int e^{-\beta H(x',p')}dx'dp'}. \tag{1.3}$$

This is known as Monte Carlo integration, as the integral on the right is replaced by the random process on the left.

The earliest technique for attacking such problems computationally was the Metropolis-Hastings algorithm. This generated samples $\{s_i\}$ using a two-step process that was repeated over and over. Starting with some random initial sample s_0 one then generates a proposed sample $\tilde{s_1}$ using a random transition $T(x' \mid x)$ over the state space. For example, if the system of interest is a collection of spins $\vec{\mu}_j$ then we can pick one spin uniformly and then randomly rotate it. Once we have the proposed new sample $\tilde{s_1}$ we decide to accept it, in which case $s_1 = \tilde{s_1}$ or reject it, in which case we go back to the previous point and set $s_1 = s_0$, with the randomized filter

$$\Pr[s_i = \tilde{s_i}] = \min\left(1, \frac{e^{-\beta H(\tilde{s_i})}/\mathcal{Z}}{e^{-\beta H(s_{i-1})}/\mathcal{Z}}\right) = \min\left(1, e^{-\beta(H(\tilde{s_i})-H(s_{i-1}))}\right). \tag{1.4}$$

This is known as the Metropolis filter and one key insight is that it only depends on the difference in energy between the previous state and the proposed state, the dependence on the partition function vanishes when taking the ratio. Since the initial sample can be random and the transition step is also randomized, this leads to an efficiently computable algorithm for generating new samples.

The last remaining question is if these samples are actually representative of the thermal distribution $p_\beta(x, p)$. So long as the transition matrix is ergodic, meaning any starting state (x', p') has a finite expected hitting time to reach any other arbitrary state (x'', p''), then the thermal distribution $p_\beta(x, p)$ satisfies a condition known as Detailed Balance which then guarantees convergence to $p_\beta(x, p)$. As the above process is a Markov chain, since the distribution over the next sample only depends on the state of the current sample, we can define the Markov transition probability (which is different from the state transition probabilities) as $\Pr[(x, p) \rightarrow$

(x', p')] and we say that a state probability distribution π satisfies Detailed Balance if the following equation holds

$$\Pr[(x_1, p_1) \to (x_2, p_2)]\pi(x_1, p_1) = \Pr[(x_2, p_2) \to (x_1, p_1)]\pi(x_2, p_2). \tag{1.5}$$

This condition, along with ergodicity, is sufficient to prove that the probability s_i, for $i \gg 1$, will be distributed according to π. The amount of samples needed to converge to π is an incredibly difficult question to answer and the subject of much study on Markov chains.

One of the most recent improvements to this algorithm came about in the late 1980′s and was formalized in Radford Neal's thesis in the Computer Science Department here at the University of Toronto [1]. This new algorithm, now called Hamiltonian Monte Carlo (HMC) but previously known as Hybrid Monte Carlo, modified the existing Metropolis-Hastings algorithm by changing the state transition function. Instead of choosing the next proposed sample randomly, in HMC the momentum variable is chosen from a Gaussian distribution with variance proportional to $1/\beta$ and then the time dynamics generated by $H(x, p)$ are used to propose a new sample. Since time evolution does not change the energy, the Metropolis filter then accepts every sample in the limit of perfect numerical integration of Hamilton's equations of motion. This technique allows for much higher dimensional state spaces to be explored and empirically leads to less correlated samples.

We introduce HMC for the conceptual changes it makes to the original Metropolis-Hastings method. In Metropolis-Hastings the filter step is really what fixes the distribution to mimic the Gibbs distribution and the transition step is simply to guarantee ergodicity. The filtration step in this sense is rather artificial and more computational in spirit. In HMC the filtration step is virtually eliminated by instead utilizing the dynamics provided by nature. We are only able to take advantage of this because in classical mechanics the position and momentum variables "commute", meaning that the thermal distribution is really a product of distributions over position and momenta as

$$e^{-\beta H(x,p)} = e^{-\beta \frac{p^2}{2m} - \beta V(x)} = e^{-\beta \frac{p^2}{2m}} \cdot e^{-\beta V(x)}, \tag{1.6}$$

and the partition function also splits as a product $\mathcal{Z} = \mathcal{Z}_x \cdot \mathcal{Z}_p$. In a sense, HMC is using time dynamics to provide equilibration not between multiple particles but instead between momenta and position, and due to the simple nature of momentum thermal states as Gaussian random variables we can prepare these states at arbitrary temperatures relatively easy.

This crucial step of commutativity of x and p proves to be a difficult obstactle to overcome when extending HMC to a quantum mechanical setting. If one takes a naïve adoption to a continuous, single variable quantum Hamiltonian $H(\hat{x}, \hat{p})$ and utilizes Gaussian distributed momentum *kicks* $e^{ip_{\text{kick}}\hat{x}}|\psi\rangle$, where p_{kick} is Gaussian and $\hat{x}$ is the position operator that generates momentum translations, one can then show that the maximally mixed state is the unique fixed point of the dynamics. This says that our model of a thermalizing environment as providing random momentum shifts is fundamentally **wrong**. Instead, one needs to come up with a simpler system that can be prepared in the thermal state at the desired temperature and use that to cool (or heat) the system to the target temperature.

The smaller environment that we ended up landing on is a single additional two level system, or qubit. The next difficulty is then to choose an interaction between the qubit and the system of interest that leads to thermalization. When looking to physics for inspiration, the situation for thermalization is much less clear in the quantum setting compared to the clear picture we have classically. Even determining a model for equilibrium is not as clear cut. If we look at a closed setting with two Hilbert spaces $\mathcal{H}_A \otimes \mathcal{H}_B$, similar to our classical scenario, one needs to determine the interaction model between the two halves. Further, the energy of each half is no longer a single function but rather a random variable dependent on the density matrix ρ. Even computing derivatives of the mean energy is difficult as one needs an expression for ρ to compute the von Neumann entropy $-\operatorname{Tr}[\rho \log \rho]$.

There are two main approaches physicists have taken to avoid these complications: one involves a conjecture on large, closed quantum systems and the other involves modelling open quantum systems. The first involves a conjecture known as the Eigenstate Thermalization Hypothesis (ETH) [2] and outlines conditions in which a very large closed quantum system may replicate thermal expectations for *local* observables on single particles. This framework has many connections to quantum chaos and has been proven to hold in such chaotic systems, but remains unproven for generic quantum systems, hence the "hypothesis" in the name. The second connection involves studying the effects of a quantum environment on a quantum system and essentially ignoring the state of the environment. By making several assumptions, such as an infinitely large environment, weak coupling, and an assumption known as Markovianity (where the system and environment remain in product states after each interaction), one can show that the thermal state is the unique fixed point for generic open systems. This is done

using the Davies' Generators [3] that give rise to a Linbladian evolution on the system with the desired fixed point. However, there exist many scenarios in nature where these assumptions are not met, such as finite sized environments that retain some correlations with the system and strong coupling situations. Finding Linblad evolutions that allow for generic thermalization for arbitrary system and environment pairs remains an open question. However, we do note that a flurry of quantum algorithms have been developed in recent years that are modeled off of these Linblad dynamics [4–7].

The model for thermalization that we propose is instead rooted in the concepts developed by Hamiltonian Monte Carlo: prepare an additional register in the desired thermal state, let time evolution mix the two constituents, and then refresh the controllable register. As mentioned before, we will make do with just a single qubit ancilla. In order for this process to work there are two difficulties that must be overcome in the quantum mechanical setting: the first is that a Hamiltonian must be chosen for the single ancilla, which amounts to choosing an energy gap we denote γ, and the second is that an interaction term must be chosen between the ancilla and the system. One of the main contributions of this thesis is a rigorous proof that choosing from a suitably random ensemble of interactions is sufficient for thermalization.

This model turns out to have many beneficial properties that existing methods do not have. The first is that this model is very straightforward to implement on a quantum computer using existing algorithms. Evolution by a time independent Hamiltonian is the most complicated subroutine necessary. We will develop these primitives from scratch, showing how simple product formulas can be implemented at the gate level by utilizing a Pauli decomposition of the Hamiltonian. Chapter 2 provides a new framework for developing new product formulas from existing ones. These ideas were developed in [8] and allow for one to combine deterministic and randomized product formulas into a single channel. This makes it particularly well suited to simulating environmental effects on a system.

One other property that our model for thermalization has is that it explicitly models the environment, keeping track of it's non-equilibrium state throughout the evolution. This allows for us to turn the problem of preparing the system in a thermal state to instead allow for our ancilla qubit to probe and extract information about the system. For example, if we know that the system is already in a thermal state we can allow it the thermalize our ancilla qubit. By measuring the temperature of the ancilla we can infer the temperature of the system,

allowing us to do thermometry without creating an interaction model. This is advantageous for complicated quantum systems, as there currently only exists thermometry techniques for harmonic oscillators.

The remainder of this thesis is organized as follows. In Chapter 2 we introduce techniques for implementing existing product formulas on quantum computers, along with introducing new techniques for composing multiple simulation techniques for partitioned Hamiltonians. In Chapter 3 we demonstrate how these techniques can be used to prepare thermal quantum states on fault-tolerant quantum computers. These two chapters represent the main technical contributions of this thesis and draw on the papers [8] and [9] for the analytic ideas and [10] for numerical studies.

Chapter 2

Composite Simulations

The simulation of time-independent Hamiltonian dynamics is a fundamental primitive in quantum computing. To start, the computational problem of approximating the time dynamics of even k-local Hamiltonians (where k is a small constant) is BQP-Complete. This means that any computational problem that can be solved efficiently on a quantum computer can be efficiently reduced to a simulation problem.

The simulation of quantum systems remains one of the most compelling applications for future digital quantum computers [11–16]. As such, there are a plethora of algorithm options for compiling a unitary evolution operator $U(t) = e^{\{-iHt\}}$ to circuit gates [17–24]. Some of the simplest such algorithms are product formulas in which each term in a Hamiltonian $H = \sum_i h_i H_i$ is implemented as $e^{iH_i t}$. A product formula is then a particular sequence of these gates that approximates the overall operator $U(t)$. Two of the most well known product formula include Trotter-Suzuki Formulas [18,20,25,26] and the QDrift protocol in which terms are sampled randomly [24,27]. These two approaches are perhaps the most popular ancilla-free simulation methods yet discovered.

One of the main drawbacks of Trotter-Suzuki formulas is that each term in the Hamiltonian has to be included in the product formula regardless of the magnitude of the term. This leads to a circuit with a depth that scales at least linearly with the number of terms in H, typically denoted L. QDrift avoids this by randomly choosing which term to implement next in the product formula according to an importance sampling scheme in which higher weight terms have larger probabilities. The downside to QDrift is that it has the same asymptotic scaling with $\frac{t}{\varepsilon}$

as a first-order Trotter formula, meaning it is outperformed at large $\frac{t}{\varepsilon}$ by even a second-order Trotter formula.

In this paper we present a framework for combining simulation channels in a way that allows one to flexibly interpolate the gate cost tradeoffs between the individual channels. The primary example we study is the composition of Trotter-Suzuki and QDrift channels. This is motivated in some part as an effort to extend randomized compilers to include conditional probabilities and in some part to encapsulate progress in chemistry simulations of dropping small weight terms or shuffling terms around different time steps [28]. This latter concept was first developed with the idea of "coalescing" terms into "buckets" by Wecker et al. [28] and further explored by Poulin et al. [29]. They showed that grouping terms of similar sizes together to be skipped during certain Trotter steps led to negligible increases in error and reduced gate counts by about a factor of 10. Similar improvements are also seen in the randomized setting of [30]. In this work we extend on these ideas by placing a specific set of terms into a Trotter partition and the rest in a QDrift partition. This simple division can then be studied analytically and we are able to provide sufficient conditions on asymptotic improvements over completely Trotter or completely QDrift channels. Although we are not able to develop the idea of conditional samples in QDrift protocols, our procedure can be viewed as a specific subset of what a generic Markovian QDrift would look like. We briefly mention these generalizations in Section .

2.1 Related Work

Recent approaches have sought to use the advantages of randomized compilation as a subset of an overall simulation, such as the hybridized scheme for interaction picture simulations [31]. What separates these two works is that our approach offers a more flexible approach for generic time-independent simulation problems whereas the hybridized schemes are specifically tailored to taking advantage of the time dependence introduced by moving to an interaction picture. As such, the hybridized approach achieves asymptotic advantages when the size of the interaction picture term dominates the overall Hamiltonian. This typically occurs in instances in which the size of an operator is unbounded, which can occur in lattice field theory simulations or constrained systems. The way the hybridized scheme in [31] works is via a "vertical" stacking of simulation channels, for example one channel to handle the Interaction Picture rotations and then other channels on top of this to simulate the time-dependence it generates on the remaining

Hamiltonian terms. Our work instead remains in the Schrodinger time evolution picture and we perform a "horizontal" stacking of simulation techniques. By horizontal we mean for a given simulation time we split the Hamiltonian up into (potentially) disjoint partitions and simulate each partition for the full simulation time but with different techniques, such as Trotter or QDrift. These techniques allow us to achieve asymptotic improvements over either method for a loose set of assumptions.

There are two other simulation techniques that have been proposed recently that have a similar interpolation behavior between QDrift and Trotter channels. The first of these methods is the SparSto, or Stochastic Sparsification, technique by Ouyang, White, and Campbell [32]. The SparSto procedure randomly sparsifies the Hamiltonian and performs a randomly ordered first-order Trotter formula on the sampled Hamiltonian. They construct these probabilities such that the expected Hamiltonian is equal to the Hamiltonian being simulated. They then fix the expected number of oracle queries of the form $e^{iH_i t'}$ and give diamond distance bounds on the resulting channel error. The claim for interpolation between Trotter and QDrift is that one can fix the expected number of gates to be 1 for each time step, in which case the sparsification mimics QDrift, whereas if no sparsification is performed then the channel is simply implementing Trotter. They show that this allows for one to have reduced simulation error up to an order of magnitude on numerically studied systems as compared to Trotter or QDrift. One downside to these techniques is that the number of gates applied is a random variable, so making gate cost comparisons is rather difficult especially considering that no tail bounds on high gate cost sampled channels are provided. In [32] they prefer to fix the expected gate cost and analyze the resulting diamond norm error. In contrast, our procedures directly implement both QDrift and Trotter channels and have a fixed, deterministic gate cost.

The second method of note with both QDrift and Trotter behavior is that of Jin and Li [33]. They develop an analysis of the variance of a unitary consisting of a first-order Trotter sequence followed by a QDrift channel. They focus on bounding the Mean Squared Error (MSE) of the resulting channel and use a simple partition of the Hamiltonian terms based on spectral norm. Their partitioning scheme places all terms below some cutoff into the first-order Trotter sequence and all terms above the cutoff into the QDrift channel. Their main results show an interpolation of the MSE between 0 when the partitioning matches a solely Trotter channel and matching upper bounds for QDrift when all terms are randomly sampled. This work goes beyond

the results from Jin and Li by providing an analysis of the diamond distance between an ideal evolution and our implemented channel, which is more useful analytically than the MSE, as well as providing upper bounds on the number of gates needed in an implementation to meet this diamond distance. In addition our work remains independent of specific partitioning schemes as much as possible and instead places restrictions on which partitions achieve improvements. In the interest of practicality we do show methods for partitioning that can be useful in both the first-order and higher-order Trotter cases. Specifically for higher-order Trotter formulas we give a probabilistic partitioning scheme that is easily computable and matches gate cost upper bounds in the extreme limits as our probabilities saturate the QDrift and Trotter limits.

The rest of the paper is organized as follows. We first provide a brief summary of the main results in Section 2.2. After reviewing known results and notation in Section 2.3, we explore methods for creating Composite channels using First-Order Trotter Formulas with QDrift in Section 2.4 as a warmup. This is broken down into three parts in which we find the gate cost for an arbitrary partition, we then give a method for producing a good partitioning, and then we analyze conditions in which a Composite channel can beat either first-order Trotter or QDrift channels. In Section 2.5 we then extend this framework to more general higher-order Trotter Formulas. This section mirrors the organization of the first-order Trotter section, namely we find the cost of an arbitrary partition, we give a method for producing a partition efficiently, and then we analyze when one could see improvements over the constituent channels. Finally, in Section 2.7 we discuss extensions to this model that allow a flexible interpolation between various types of product formulas that could be leveraged numerically.

2.2 Main Results

2.3 Preliminaries

In this section we will first introduce the necessary notation we will use and then state known results about Trotter-Suzuki formulas and QDrift channels. We work exclusively with time-independent Hamiltonians H in a 2^n dimensional Hilbert space $\mathcal{H}$. We also assume that H consists of L terms $H = \sum_{i=1}^{L} h_i H_i$ where h_i represents the spectral norm of the term, H_i is a Hermitian operator on $\mathcal{H}$, and $\|H_i\| = 1$. Note without loss of generality we can always assume $h_i \geq 0$, as we can always absorb the phase into the operator H_i itself. We use $\|M\|$ to refer to

the spectral norm, or the magnitude of the largest singular value of M. We use λ to refer to the sum of h_i, namely $\lambda = \sum_i h_i$. We will also use subscripts on λ, such as λ_A to refer to sums of subsets of the terms of H. For example, if $H = 1H_1 + 2H_2 + 3H_3$ and $G = 1H_1 + 2H_2$, then $\lambda = 6$ and $\lambda_G = 3$.

We use $U(t)$ to refer to the unitary operator e^{iHt} and $\mathcal{U}(t)$ to refer to the channel $\rho \mapsto U(t)\rho U(t)^\dagger$. We will be particularly concerned with simulations of subsets of the terms of H, which we denote as follows. We typically work with a partition of H into two matrices $H = A + B$, and we let $A = \sum_i a_i A_i$ and $B = \sum_j b_j B_j$, where we have simply relabeled the relevant h_i and H_i into a's, b's, A's, and B's. This allows us to define the exact unitary time evolution operators $U_{A(t)} = e^{iAt}$ and channels $\mathcal{U}_{A(t)} = U_{A(t)}\rho U_{A(t)}^\dagger$, similarly defined for B.

Although much of the literature for Trotter-Suzuki formulas is written in terms of unitary operators $U = e^{iHt}$ acting on state vectors $|\psi\rangle$ for our purposes it will prove most natural to consider a product formula as a channel $\mathcal{U} = e^{iHt}\rho e^{-iHt}$ acting on a density matrix ρ. After reviewing known results on unitary constructions of Trotter-Suzuki formulas we give a straightforward extension of these bounds to channels.

2.3.1 Product Formulas

We now show how to implement basic product formulas, namely Trotter-Suzuki or just Trotter formulas as well as QDrift, assuming access to arbitrary single qubit unitaries and controlled NOT gates. We will first show how to implement an arbitrary Pauli rotation e^{iPt} for some Pauli string P using no additional ancilla qubits. Then we will define the Trotter-Suzuki construction and give heuristic evidence for the first order scaling. We avoid giving a rigorous proof and instead refer the reader to the canonical paper by Childs et. al [26]. Lastly, we will present the construction of QDrift by Campbell [24], providing a heuristic proof of correctness.

Definition 2.1 (Trotter-Suzuki Formulae): *Given a Hamiltonian H, let $S^{(1)}(t)$ denote the first-order Trotter-Suzuki time evolution operator and channel as*

$$S^{(1)}(t) := e^{ih_L H_L t}...e^{ih_1 H_1 t} = \prod_{i=1}^{L} e^{ih_i H_i t},$$
$$\mathcal{S}^{(1)}(\rho; t) := S^{(1)}(t) \cdot \rho \cdot S^{(1)}(t)^\dagger. \tag{2.7}$$

This first order formula serves as the base case for the recursively defined higher-order formulas

$$S^{(2)}(t) := e^{ih_1 H_1 (\frac{t}{2})}...e^{ih_L H_L (\frac{t}{2})} e^{ih_L H_L (\frac{t}{2})}...e^{ih_1 H_1 (\frac{t}{2})} = S^{(1)}(t/2) \cdot S^{(1)}(-t/2)^\dagger$$
$$S^{(2k)}(t) := S^{(2k-2)}(u_k t) \cdot S^{(2k-2)}(u_k t) \cdot S^{(2k-2)}((1-4u_k)t) \cdot S^{(2k-2)}(u_k t) \cdot S^{(2k-2)}(u_k t)$$
$$\mathcal{S}^{(2k)}(\rho; t) := S^{(2k)}(t) \cdot \rho \cdot S^{(2k)}(t)^\dagger, \tag{2.8}$$

where $u_k := \frac{1}{4-4^{\frac{1}{2k-1}}}$. In addition we define $\Upsilon_k := 2 \cdot 5^{k-1}$ as the number of "stages" in the higher-order product formula, although we will typically just write Υ when the order is apparent.

Despite their simplicity, Trotter-Suzuki formulas are fiendishly difficult to analyze. For decades the only error analysis that existed was worst-case analysis that often drastically overestimated the actual error. It was known that the first order expression depended on the commutator structure among the terms, but this was not generalized until 2021 in [26], 25 years after Lloyd's original work [34]. We will follow [26] and denote the expression that captures this commutator scaling as α_{comm}, and sometimes when space is needed this may be abbreviated to α_C when the context is clear, which we define as

$$\alpha_{\text{comm}}(H, 2k) := \sum_{\gamma_i \in \{1,...,L\}} \left(\prod h_{\gamma_i}\right) \left\| \left[H_{\gamma_{2k+1}}, \left[H_{\gamma_{2k}}, \left[..., \left[H_{\gamma_2}, H_{\gamma_1}\right]...\right]\right]\right] \right\|_\infty. \tag{2.9}$$

We will also make heavy use of the restriction of α_{comm} to subsets of the total Hamiltonian H. For example, if $H = A + B$ then we define the commutator structure over A as

$$\alpha_{\text{comm}}(A, 2k) := \sum_{\gamma_i \in \{1,...,L_A\}} \left(\prod a_{\gamma_i}\right) \left\| \left[A_{\gamma_{2k+1}}, \left[A_{\gamma_{2k}}, \left[..., \left[A_{\gamma_2}, A_{\gamma_1}\right]...\right]\right]\right] \right\|_\infty. \tag{2.10}$$

This then allows us to decompose the total commutator structure into 3 pieces: commutators that contain only terms from A, commutators that contain only terms from B, and commutators that contain *at least* one term from both A and B

$$\alpha_{\text{comm}}(H, 2k) = \alpha_{\text{comm}}(A, 2k) + \alpha_{\text{comm}}(B, 2k) + \alpha_{\text{comm}}(\{A, B\}, 2k). \tag{2.11}$$

We also note the following bounds that will be used later. We can ignore the commutator structure and use the triangle inequality to get

$$\begin{aligned}
\alpha_C(H, 2k) &= \sum_{\gamma_i \in \{1,\ldots,L\}} \left(\prod h_{\gamma_i}\right) \left\| \left[H_{\gamma_{2k+1}}, \left[H_{\gamma_{2k}}, \left[\ldots, \left[H_{\gamma_2}, H_{\gamma_1}\right]\ldots\right]\right]\right] \right\| \\
&\leq \sum_{\gamma_i \in \{1,\ldots,L\}} \left(\prod h_{\gamma_i}\right) 2^{2k} \left\| H_{\gamma_{2k+1}} \right\| \left\| H_{\gamma_{2k}} \right\| \ldots \left\| H_{\gamma_1} \right\| \\
&= 2^{2k} \prod_{i=1}^{2k+1} \sum_{\gamma_i \in \{1,\ldots,L\}} h_{\gamma_i} \\
&= 2^{2k} \|h\|^{2k+1}.
\end{aligned} \tag{2.12}$$

Similar arguments show the following

$$\alpha_C(A, 2k) \leq 2^{2k} \|h_A\|^{2k+1} \tag{2.13}$$

$$\alpha_C(\{A, B\}, 2k) \leq 2^{2k} \sum_{i=1}^{2k} \|h_A\|^l \|h_B\|^{2k+1-l} \leq 2^{2k} \|h_A\|^{2k+1}. \tag{2.14}$$

This allows us to give the error associated with a a Trotter-Suzuki formula in the following theorem.

Theorem 2.2 (Trotter-Suzuki [26]): *Let $S^{(2k)}$ be the Trotter-Suzuki unitary as given in Definition 2.1 for the Hamiltonian $H = \sum_{i=1}^{L} h_i H_i$. Then the spectral norm of the difference between the Trotter-Suzuki formulas $S^{(1)}(t/r)$ and $S^{(2k)}(t/r)$ and the ideal evolution $U(t/r)$ is given by*

$$\begin{aligned} \left\| U(t/r) - S^{(1)}(t/r) \right\|_\infty &\leq \frac{t^2}{2r^2}\, \alpha_{\text{comm}}(H, 1), \\ \left\| U(t/r) - S^{(2k)}(t/r) \right\|_\infty &\leq \frac{(\Upsilon t)^{2k+1}}{r^{2k+1}(k+1/2)} \alpha_{\text{comm}}(H, 2k). \end{aligned} \tag{2.15}$$

The associated operator exponential cost can be computed via standard time-slicing arguments as

$$\begin{aligned} C^{(1)}_{\text{Trot}}(H, t, \varepsilon) &= L \left\lceil \frac{t^2}{2\varepsilon} \sum_{i,j} \left\| \left[H_i, H_j \right] \right\| \right\rceil \\ C^{(2k)}_{\text{Trot}}(H, t, \varepsilon) &= \Upsilon L \left\lceil \frac{(\Upsilon t)^{1+1/2k}}{\varepsilon^{1/2k}} \left(4\alpha_{\text{comm}} \frac{(H, 2k)^{1/2k}}{2k+1} \right) \right\rceil \end{aligned} \tag{2.16}$$

The complete proof of the above theorem is very nontrivial and beyond the scope of this thesis. See [26] for complete details, the proof of the higher order bounds can be found in Appendix E and the first order expression is found in Proposition 9 in Section V. Instead, we provide a heuristic proof for the first order error for completeness.

TODO: Probably should include a proof that spectral norm bounds on a unitary channel imply a diamond distance bound that is different only by a factor of 2.

TODO: Also probably should do a bit better proof below, can use a heuristic that all Taylor series of the exponential have the same special time t? or maybe just produce a bound?

Proof Heuristic First Order: Compute a Taylor Series for the Trotter formula and the ideal evolution. First the ideal evolution:

$$U(t/r) = e^{iH\frac{t}{r}} = \mathbb{1} + \frac{it}{r} H + O((t/r)^2). \tag{2.17}$$

Then the Trotter terms:

$$\prod_{i=1}^{L} e^{ih_i H_i t'} = \prod_{i=1}^{L} (\mathbb{1} + ih_i t' H_i + O(t'^2)) = \mathbb{1} + it' \sum_{i=1}^{L} h_i H_i + O(t'^2). \tag{2.18}$$

Pretty clear to see that in the difference $U(t/r) - S^{(1)}(t/r)$ the zeroth and first order terms vanish, leaving only the second order. □

2.3.2 Randomized Product Formulas

We now introduce QDrift [24], one of the first randomized compilers for quantum simulation. The main idea of QDrift is that instead of iterating through each term in the Hamiltonian to construct a product formula, or even a random ordering of terms as in [20], each exponential is chosen randomly from the list of terms in H. Each term is selected with probability proportional to it's spectral weight, the probability of choosing H_i is $\frac{h_i}{\sum_j h_j} =: \frac{h_i}{\|h\|}$, and then simulated for a time $\tau = \|h\|t$. This is the protocol for a single sample. As we will denote the portion of the Hamiltonian that we simulate with QDrift in later sections as B we let N_B denote the number of samples used.

Definition 2.3 (QDrift Channel): *Let N_B denote the number of samples, $\|h\| = \sum_{i=1}^{L} h_i$, and $\tau := \frac{\|h\|t}{N_B}$. The QDrift channel for a single sample is given as*

$$\mathcal{Q}(\rho; t, 1) := \sum_{i=1}^{L} \frac{h_i}{\|h\|} e^{-iH_i\|h\|t} \cdot \rho \cdot e^{+iH_i\|h\|t}, \tag{2.19}$$

and the QDrift channel for N_B samples is

$$\mathcal{Q}(t, N_B) := \mathcal{Q}(t/N_B, 1)^{\circ N_B}. \tag{2.20}$$

Once we have the channel defined we can then state the main results of [24].

Theorem 2.4 (QDrift Cost): *Given a Hamiltonian H, time t, and error bound ε, the ideal time evolution channel $\mathcal{U}(t)$ can be approximated using $N_B = \frac{4t^2\|h\|^2}{\varepsilon}$ samples of a QDrift channel. This approximation is given by the diamond distance*

$$\|\mathcal{U}(t) - \mathcal{Q}(t, N_B)\|_\diamond \leq \frac{4t^2\|h\|^2}{N_B}. \tag{2.21}$$

The number of operator exponentials N_B is then given as

$$C_{\text{QD}}(H, t, \varepsilon) = N_B = \left\lceil \frac{4t^2\|h\|^2}{\varepsilon} \right\rceil. \tag{2.22}$$

Proof Theorem 2.4: This one is also a Taylor's Series

$$\mathcal{U}_{\text{QD}}(t) \tag{2.23}$$

□

2.4 First Order Composite Channels

We now turn towards combining the two product formulas given in Section 2.3 in a Composite channel. We first will assume that the Hamiltonian has already been partitioned into two pieces $H = A + B$, where A will be simulated with a first order Trotter formula and B with QDrift. Given a fixed partitioning allows for us to compute the diamond distance error in the resulting channel, which then allows us to bound the number of operator exponentials needed to implement the channel. The resulting cost function will then be parametrized by the partitioning, which we can then use to determine an optimal partitioning algorithm. Finally, we give a specific instance in which a Composite channel can offer asymptotic improvements in query complexity over either a purely Trotter or QDrift channel.

2.4.1 Query Complexity

To analyze the error of our Composite channel we need to first reduce the overall time evolution channel $\rho \mapsto e^{-iHt}\rho e^{+iHt}$ into the simpler pieces that we can analyze with our Trotter and QDrift results. Assuming a partitioning $H = A + B$, where A consists of terms that we would like to simulate with Trotter and B has the terms we would like to sample from with QDrift. We now introduce the "outer-loop" error $E_{\{A,B\}}$ induced by this partitioning, which is as follows

$$E_{\{A,B\}}(t) := e^{-iHt}\rho e^{+iHt} - e^{-iBt}e^{-iAt}\rho e^{+iAt}e^{+iBt}. \tag{2.24}$$

We use the phrase "outer-loop" as this decomposition is done before any simulation channels are implemented. This gives the first order Composite channel $\mathcal{C}$ as

$$\mathcal{C}^{(1)}(t) := \mathcal{Q}_B(t, N_B) \circ \mathcal{S}_A^{(1)}(t). \tag{2.25}$$

We will first bound the error of this approximation to the ideal evolution. This error bound will then allow us to bound the number of exponentials needed to approximate the ideal dynamics.

Lemma 2.5 (First-Order Composite Error): *Given a Hamiltonian H partitioned into a first order Trotter term A and QDrift term B such that $H = A + B$, the first order Composite Channel $\mathcal{C}^{(1)}$ has an error of at most*

$$\left\|\mathcal{U}(t) - \mathcal{C}^{(1)}(t)\right\|_\diamond \leq t^2 \left(\sum_{i,j} a_i a_j \left\| \left[A_i, A_j\right] \right\| + \sum_{i,j} \left\| \left[A_i, B_j\right] \right\| + \frac{4\|h_B\|^2}{N_B} \right). \tag{2.26}$$

Proof: We will first use an outer-loop decomposition to get the error associated by our partitioning, note we temporarily supress arguments of (t) for clarity,

$$\begin{aligned} \left\|\mathcal{U} - \mathcal{C}^{(1)}\right\|_\diamond &= \left\|\mathcal{U} - \mathcal{U}_B \circ \mathcal{U}_A + \mathcal{U}_B \circ \mathcal{U}_A - \mathcal{C}^{(1)}\right\| \\ &\leq \left\|\mathcal{U} - \mathcal{U}_B \circ \mathcal{U}_A\right\|_\diamond + \left\|\mathcal{U}_B \circ \mathcal{U}_A - \mathcal{C}^{(1)}\right\|_\diamond. \end{aligned} \tag{2.27}$$

We then can bound the leftmost term using the error decomposition

$$\left\|\mathcal{U} - \mathcal{U}_B \circ \mathcal{U}_A\right\|_\diamond \leq 2\|U - U_B \cdot U_A\|_\infty \leq t^2 \sum_{i,j} a_i b_j \left\| \left[A_i, B_j\right] \right\|_\infty. \tag{2.28}$$

And the rightmost term can be bounded using the subadditivity of the diamond distance

$$\begin{aligned} \left\|\mathcal{U}_B \circ \mathcal{U}_A - \mathcal{C}^{(1)}\right\|_\diamond &= \left\|\mathcal{U}_B \circ \mathcal{U}_A - \mathcal{Q}_B \circ \mathcal{S}_A^{(1)}\right\|_\diamond \\ &\leq \left\|\mathcal{U}_B - \mathcal{Q}_B\right\|_\diamond + \left\|\mathcal{U}_A - \mathcal{S}_A^{(1)}\right\|_\diamond \\ &\leq \frac{4\|h_B\|t^2}{N_B} + t^2 \sum_{i,j} a_i a_j \left\| \left[A_i, A_j\right] \right\|_\infty. \end{aligned} \tag{2.29}$$

Substituting Equation (2.28) and Equation (2.29) into Equation (2.27) yields the statement. □

Now that we have an upper bound on first order error for an arbitrary t we can leverage this into a bound on the number of operator exponentials to reach arbitrary error ε using standard

time-slicing arguments. By letting $t \to t/r$ and then repeating our Composite channel r times we can control the accumulated error from each step. One of the beautiful features of product formulas is that this time-slicing leads to an overall reduction in the error, or in other words

$$\lim_{r\to\infty} \left(e^{iA\frac{t}{r}} e^{iB\frac{t}{r}}\right)^r = e^{i(A+B)t}. \tag{2.30}$$

The following theorem utilizes a quantitative variant of the above, along with the error bounds we just proved, to provide the first order Composite cost Theorem.

Theorem 2.6 (First-Order Composite Cost): *Given a time t, error bound ε, and a partitioned Hamiltonian $H = A + B$, the first order Composite Channel $\mathcal{C}^{(1)}$ approximates the ideal time evolution operator $\left\|\mathcal{U}(t) - \mathcal{C}^{(1)}(t/r)^{\circ r}\right\|_\diamond \leq \varepsilon$ using no more than*

$$C^{(1)}_{\text{Comp}} = (L_A + N_B)\left\lceil \frac{t^2}{\varepsilon}\left(\sum_{i,j} a_i a_j \left\|\left[A_i, A_j\right]\right\| + \sum_{i,j}\left\|\left[A_i, B_j\right]\right\| + \frac{4\|h_B\|^2}{N_B}\right)\right\rceil \tag{2.31}$$

operator exponential queries.

Proof: We first will use the fact that since H commutes with itself the time evolution operator can be decomposed into r steps as $\mathcal{U}(t) = \mathcal{U}(t/r)^{\circ r}$. Then we can use the sub-additivity of the diamond norm with respect to channel composition to get the bound

$$\left\|\mathcal{U}(t) - \mathcal{C}^{(1)}(t/r)^{\circ r}\right\|_\diamond = \left\|\mathcal{U}(t/r)^{\circ r} - \mathcal{C}^{(1)}(t/r)^{\circ r}\right\|_\diamond \leq r\left\|\mathcal{U}(t/r) - \mathcal{C}^{(1)}(t/r)\right\|_\diamond. \tag{2.32}$$

Now we utilize Lemma 2.5 to upper bound the single step error

$$\left\|\mathcal{U}(t) - \mathcal{C}^{(1)}(t/r)^{\circ r}\right\|_\diamond \leq r\left(\frac{t^2}{r^2}\right)\left(\sum_{i,j} a_i a_j \left\|\left[A_i, A_j\right]\right\| + \sum_{i,j}\left\|\left[A_i, B_j\right]\right\| + \frac{4\|h_B\|^2}{N_B}\right). \tag{2.33}$$

In order for the above to be upper bounded by ε we require

$$r \geq \left(\frac{t^2}{\varepsilon}\right)\left(\sum_{i,j} a_i a_j \left\|\left[A_i, A_j\right]\right\| + \sum_{i,j}\left\|\left[A_i, B_j\right]\right\| + \frac{4\|h_B\|^2}{N_B}\right), \tag{2.34}$$

and since increasing r only increases the number of operator exponentials used we simply set r to be the ceiling of the right hand side. This then yields the theorem as we have r applications of $\mathcal{C}^{(1)}$ and each application uses L_A operator exponentials for the Trotter channel and N_B samples of the QDrift channel. □

2.4.2 First-Order Parameter Settings

Our next task will be to determine "parameter settings" that optimize this gate cost, namely the partitioning $A + B$ and setting the number of QDrift samples N_B. To do this it will prove useful to have a continuous, non-integer variant of the gate cost expression which we define as

$$\tilde{C}^{(1)}_{\text{Comp}} := (L_A + N_B)\frac{t^2}{\varepsilon}\left(\sum_{i,j} a_i a_j \| [A_i, A_j] \| + \sum_{i,j} \| [A_i, B_j] \| + \frac{4\|h_B\|^2}{N_B}\right). \tag{2.35}$$

This is the same as $C^{(1)}_{\text{Comp}}$ but without the ceiling operation $\lceil \cdot \rceil$. Although a user could use any value of N_B they want, such as always setting $N_B = 1$, we provide the following setting for N_B that is optimal with respect to the continuous variant of the gate cost.

Lemma 2.7: *Let $\tilde{C}^{(1)}_{\text{Comp}}$ denote the continuous relaxtion to the cost of a first-order Composite channel with a given partitioning $H = A + B$. The optimal assignment of the number of QDrift samples N_B with respect to $\tilde{C}^{(1)}_{\text{Comp}}$ is given by*

$$N_B = \frac{2\|h_B\|\sqrt{L_A}}{\sqrt{\left(\sum_{i,j} a_i a_j \| [A_i, A_j] \| + \sum_{i,j} a_i b_j \| [A_i, B_j] \|\right)}}. \tag{2.36}$$

This assignment is not valid if both $[A_i, A_j] = 0$ for all A_i, A_j and $[A_i, B_j] = 0$ for all A_i and B_j.

Proof: We first compute the derivative of $\tilde{C}^{(1)}_{\text{Comp}}$ with respect to N_B as

$$\frac{\partial \tilde{C}_{\text{Comp}}}{\partial N_B} = \frac{t^2}{\varepsilon}\left[\sum_{i,j} a_i a_j \| [A_i, A_j] \| + \sum_{i,j} a_i b_j \| [A_i, B_j] \| - \frac{4\|h_B\|^2 L_A}{N_B^2}\right]. \tag{2.37}$$

Setting this equal to 0 and solving for N_B yields Equation (2.36). We then compute the second derivative as

$$\frac{\partial^2 \tilde{C}^{(1)}_{\text{Comp}}}{\partial N_B^2} = \frac{4t^2\|h_B\|^2 L_A}{\varepsilon N_B^3}, \tag{2.38}$$

which is always positive and therefore indicates that the optima found is a minimal cost with respect to N_B. □

There is still one remaining problem with the first-order Composite channel that we must address before we can compare it to existing product formulas: partitioning. Up until now we have assumed that a partitioning was given, but this is not a realistic assumption to make. There

are many heuristics that one could, and most likely should, take advantage of when implementing an actual channel. For example, in a chemistry simulation one can put the nuclei-electron interactions, which are typically stronger than the electron-electron interactions, into Trotter and use a spectral norm cutoff to determine the remainder. On could construct a Hubbard like model on a grid but with long-range interactions, treating the "tunneling" kinetic term and nearest neighbor interactions with Trotter and then sampling the long-range interactions with QDrift. These kinds of heuristics will most likely be important in reducing simulation costs but will ultimately be application dependent.

We would like to provide a general purpose algorithm, one that works for any Hamiltonian and matches our intuition that the above heuristics capture. The algorithm we propose is a gradient based scheme that is based on a weighting of the Hamiltonian terms H_i. This weighting is based on the following trick, for every term H_i we introduce a parameter w_i that accounts for the weight of H_i in the Trotter partition. Then our Hamiltonian can be written as

$$H = \sum_i h_i H_i = \sum_i (w_i h_i H_i + (1 - w_i) h_i H_I) = \sum_i w_i h_i H_i + \sum_i (1 - w_i) h_i H_i. \tag{2.39}$$

Then our partitioning is just $A = \sum_i w_i h_i H_i$ and $B = \sum_i (1 - w_i) h_i H_i$. We also would like to keep our interpretation of $w_i h_i$ and $(1 - w_i) h_i$ as spectral norms of the respective term in A and B, so we restrict $w_i \in [0, 1]$. In this sense the weights could be thought of as probabilities, but we will not make use of any expectations or other probabilistic notions here so they should just be thought of as weight parameters. We will make use of a probabilistic variant of this scheme in Section 2.5.

Once an initial weighting of each term is chosen, say $w_i = 1$ for the term with the largest spectral norm $h_{\max}$ and $w_i = 0$ for every other term or maybe $w_i = \frac{1}{2}$ for all i, we propose a greedy algorithm for computing a new set of weights $w_i{}'$. This greedy algorithm is based on the following gradient calculation of the weighted first order Composite channel cost

$$\frac{\partial \tilde{C}^{(1)}_{\text{Comp}}}{\partial w_m} = (L_A + N_B) \frac{t^2}{\varepsilon} \left(h_m \sum_j h_j \left\| \left[H_j, H_m \right] \right\| - \frac{8 h_m \sum_i (1 - w_i) h_i}{N_B} \right). \tag{2.40}$$

This gradient could be used in gradient descent, once the derivatives for all w_m are computed and put into a vector one could update the parameters with $w_i{}' = w_i - \eta \nabla_w \tilde{C}^{(1)}_{\text{Comp}}$, where η is some learning rate.

Although this gradient descent based algorithm would be relatively easy to compute and implement in practice, it does not take advantage of the fact that our gradients are not only analytic, but linear with respect to a *single* parameter w_m. This means that if we only update a single weight parameter at a time we can find an optimal assignment of w_m with respect to the partial derivative given above. Setting the derivative equal to 0 and solving for w_m yields

$$\frac{\partial \tilde{C}^{(1)}_{\text{Comp}}}{\partial w_m} = 0 \implies {w_m}' = 1 - \sum_{i \neq m} \frac{h_i}{h_m} \left(\frac{\|[H_i, H_m]\|}{8} - (1 - w_i) \right). \tag{2.41}$$

Once ${w_m}'$ is determined then we can move on to ${w_{m+1}}'$ until all parameters have been updated, at which point we can repeat until the parameters do not change. We unfortunately cannot provide more analysis, such as how many iterations this process will take, due to the coupling of the parameters.

Although our expression does not give exact partitionings that can be calculated in one go, our greedy algorithm based on Equation (2.41) does capture two key pieces of intuition that we suspect good partitions will have. The first is that if a term H_m commutes with every other term ($\forall i \neq m \ [H_i, H_m] = 0$) then we have that ${w_m}' = 1 + \sum_{i \neq m} \frac{h_i(1-w_i)}{h_m} \geq 1$, so we then restrict the update to ${w_m}' = 1$. This implies that H_m is completely placed into the Trotter partition, as we would expect. The other piece of intuition is that smaller terms are pushed more towards the QDrift side of the partitioning. This can be seen from Equation (2.41) while considering the limit as $h_m \to 0$. If we assume that $\|[H_i, H_m]\| \geq (1 - w_i)$ on average, then the expression becomes $w_m \to -\infty$ in this limit, which we stop at 0. In total, this indicates that large terms that do not commute with small terms with most of their weight in the Trotter partition tend to push the weight of small terms more towards QDrift.

2.4.3 Comparison with Trotter and QDrift

Now that we have analyzed the cost and given a partitioning scheme for the first order Composite channel, we would like to know under what conditions this Composite channel can lead to comparable errors with lower gate cost. Instead of aiming to show that a Composite channel will outperform either first-order Trotter or QDrift for arbitrary Hamiltonians, we instead illustrate a concrete setting in which we achieve guaranteed asymptotic improvements. In later sections we are able to show more generic conditions in which asymptotic improvements can be obtained for higher-order formulas.

To start, let H be a Hamiltonian that has a partitioning into A and B such that the following conditions hold.

1. The number of non-zero commutators between terms in A scales with the square root of L_A. Mathematically,
$$|\{(i,j) : \| [A_i, A_j] \| \neq 0\}| =: N_{\text{nz}}^2 \in o(L_A). \tag{2.42}$$
2. The strength of the B terms, $\|h_B\| = \sum_i b_i$, is asymtotically less than the maximum commutator norm divided by the number of terms in A
$$\|h_B\| \leq \frac{a_{\max} N_{\text{nz}}^2}{L_A}, \tag{2.43}$$
where $a_{\max} = \max_i a_i$.
3. The number of terms in the A partition is vanishingly small compared to the total number of terms: $L_A \in o(L)$.

Next, we can use the optimal N_B value from Lemma 2.7 and Equation (2.43) to show that
$$N_B^{-1} \in O\left(\frac{1}{\|h_B\|}\sqrt{N_{\text{nz}}^2 a_{\max}^2 + L_A a_{\max}\|h_B\|}\right) = O\left(\frac{N_{\text{nz}} a_{\max}}{\|h_B\|}\right). \tag{2.44}$$
Similarly we have $N_B \in O\left(\frac{\|h_B\|\sqrt{L_A}}{a_{\max} N_{\text{nz}}}\right)$. Thus, Theorem 2.6 gives us the asymptotic number of operator exponentials as
$$\begin{aligned} C_{\text{Comp}}^{(1)} &\in O\left(\frac{t^2}{\varepsilon}\left(L_A + \frac{\|h_B\|\sqrt{L_A}}{a_{\max} N_{\text{nz}}}\right)\left(a_{\max}^2 N_{\text{nz}}^2 + L_A a_{\max}\|h_B\| + \frac{\|h_B\| N_{\text{nz}} a_{\max}}{\sqrt{L_A}}\right)\right) \\ &\in O\left(\frac{t^2 L_A}{\varepsilon}(a_{\max}^2 N_{\text{nz}}^2)\right) \\ &\in o\left(\frac{t^2}{\varepsilon} L_A^2 a_{\max}^2\right). \end{aligned} \tag{2.45}$$
The lowest order Trotter formula for this simulation has the following asymptotic operator exponential cost, as given in
$$C_{\text{Trot}}^{(1)} \in O\left(\frac{t^2}{\varepsilon}(L N_{\text{nz}}^2 a_{\max}^2)\right) \in o\left(\frac{t^2}{\varepsilon} L L_A a_{\max}^2\right) \subseteq \omega\left(C_{\text{Comp}}^{(1)}\right). \tag{2.46}$$
For QDrift we can use Theorem 2.4 to compute

$$
\begin{aligned}
C_{\mathrm{QD}} &\in O\left(\frac{t^2}{\varepsilon}(L_A a_{\max} + \|h_B\|)^2\right) \\
&\in O\left(\frac{t^2 L_A^2 a_{\max}^2}{\varepsilon}\left(1 + \left(\frac{N_{\mathrm{nz}}}{L_A}\right)^2\right)\right) \\
&\in O\left(\frac{t^2 L_A^2 a_{\max}^2}{\varepsilon}\right) \\
&\in \omega\Big(C^{(1)}_{\mathrm{Comp}}\Big).
\end{aligned} \tag{2.47}
$$

This shows that the Composite channel has asymptotically better cost over the methods it composes, i.e. $C^{(1)}_{\mathrm{Comp}} \in o\Big(\min\Big(C^{(1)}_{\mathrm{Trot}}, C_{\mathrm{QD}}\Big)\Big)$.

Although this example may be a little contrived, it does show in a completely rigorous manner that there do exist scenarios in which even first order Composite techniques could provide significant constant factor improvements. This provides strong evidence that more detailed research is needed to understand when Composite techniques can provide advantages. We provide further numeric evidence comparing Composite techniques to Trotter and QDrift for some small systems in Section 2.6.

2.5 Higher Order Composite Channels

We now move on from first-order Trotter formulas to arbitrary higher-order Trotter formulas. To analyze this case there are a few distinct differences with the first-order channels. The first is that we now have a choice for what order formula we would like to use for the outer-loop decomposition. Previously, for the first-order decomposition we used $\mathcal{U}_B \circ \mathcal{U}_A$, but it will prove useful in our analysis to match the outer-loop order with the inner-loop Trotter order. For example, a second order outer-loop decomposition would look like $\mathcal{U}_A(t/2) \circ \mathcal{U}_B(t) \circ \mathcal{U}_A(t/2)$, where we combined the two innermost $\mathcal{U}_B(t/2)$ for compactness. The next difference is that the time scaling between QDrift, Trotter, and the outer-loop errors could all be of different orders in $\frac{t}{r}$ which leads to a non-analytically solvable polynomial in r. The last issue that we address is that the commutator structure is no longer quadratic with respect to the Hamiltonian spectral norms, so we cannot follow the term weighting partitioning scheme from the first-order case. We will follow the same organizational structure as the first-order case and first set up our definitions and bound the diamond distance error, then compute the number of $e^{iH_i t}$ queries,

followed by developing a partitioning scheme, and finally discuss the cost comparisons between our Composite channel and its constituents.

2.5.1 Query Complexity

In order to determine the number of queries needed for a Composite channel to approximate $\mathcal{U}$ we first need to bound the diamond distance error for a single iteration. We will then use time-slicing arguments similar to the proof of Theorem 2.6 to compute the number of operator exponentials required for an accurate approximation. First, we need to give a rigorous definition of the higher order Composite channel.

Definition 2.8 (Higher Order Composite Channel): *Given a Hamiltonian H partitioned into two terms A and B, let $\mathcal{C}^{(2k,2l)}$ denote the associated Composite channel that utilizes a $2k^{\text{th}}$ order inner-loop for the Trotter-Suzuki partition $\mathcal{S}_A^{(2k)}$ and has a $2l^{\text{th}}$ order outer-loop. The outer-loop construction for the Composite channel can be constructed recursively from the base case for $l = 1$, which is given by*

$$\mathcal{C}^{(2k,2)}(t) := \mathcal{Q}_B(t/2) \circ \mathcal{S}_A^{(2k)}(-t/2)^\dagger \circ \mathcal{S}_A^{(2k)}(t/2) \circ \mathcal{Q}_B(t/2), \tag{2.48}$$

and the higher-order outer-loop Composite channels are recursively defined as

$$\mathcal{C}^{(2k,2l)}(t) := \mathcal{C}^{(2k,2l-2)}(u_l t)^{\circ 2} \circ \mathcal{C}^{(2k,2l-2)}((1-4u_l)t) \circ \mathcal{C}^{(2k,2l-2)}(u_l t)^{\circ 2}, \tag{2.49}$$

where u_l and the number of stages Υ are the same as in Definition 2.1. We will typically ignore the distinction between inner and outer loops and use $\mathcal{C}^{(2k)} = \mathcal{C}^{(2k,2k)}$.

Lemma 2.9 (Higher-Order Composite Error): *The diamond distance of a single higher-order Composite channel to the ideal time evolution channel is upper bounded as*

$$\left\| \mathcal{U}(t) - \mathcal{C}^{(2k)}(t) \right\|_\diamond \leq 2\frac{(\Upsilon t)^{2k+1}}{k+1/2}(\alpha_C(\{A,B\},2k) + \Upsilon\alpha_C(A,2k)) + \Upsilon\frac{4\|h_B\|^2 t^2}{N_B}. \tag{2.50}$$

We will use the following definitions for brevity

$$P(t) := t^{2k+1}\frac{2\Upsilon^{2k+1}}{k+1/2}(\alpha_{\text{comm}}(\{A,B\},2k) + \alpha_{\text{comm}}(A,2k))$$

$$Q(t) := t^2\frac{4\Upsilon\|h_B\|^2}{N_B}. \tag{2.51}$$

Proof of Lemma 2.9:

$$\begin{aligned}\left\|\mathcal{U}(t)-\mathcal{C}^{(2k)}(t)\right\|_\diamond &\leq \left\|\mathcal{U}(t)-\mathcal{S}^{(2k)}(\{A,B\},t)\right\|_\diamond + \left\|\mathcal{S}^{(2k)}(\{A,B\},t)-\mathcal{C}^{(2k)}(t)\right\|_\diamond \\ &\leq 2\left\|e^{iHt}-S^{(2k)}(\{A,B\},t)\right\| + \left\|\mathcal{S}^{(2k)}(\{A,B\},t)-\mathcal{C}^{(2k)}(t)\right\|_\diamond. \end{aligned} \tag{2.52}$$

We can use Theorem 2.2 to bound the outer-loop error on the left as

$$\left\|e^{iHt}-S^{(2k)}(\{A,B\},t)\right\| \leq \frac{(\Upsilon t)^{2k+1}}{k+1/2}\alpha_{\text{comm}}(\{A,B\},2k). \tag{2.53}$$

We then use an inductive proof to argue that the inner-loop errors can be bounded as

$$\left\|\mathcal{S}^{(2l)}(\{A,B\},t)-\mathcal{C}^{(2k,2l)}(t)\right\|_\diamond \leq \Upsilon\left(\left\|\mathcal{U}_A(t)-\mathcal{S}_A^{(2k)}(t)\right\|_\diamond + \left\|\mathcal{U}_B(t)-\mathcal{Q}_B(t)\right\|_\diamond\right), \tag{2.54}$$

where the induction is over the outer-loop indexing of $2l$.

- **Base Case ($2l=1$):**

$$\begin{aligned}&\left\|\mathcal{C}^{(2k,2)}(t)-\mathcal{S}^{(2)}(\{A,B\})(t)\right\|_\diamond \\ &= \left\|\mathcal{Q}_B(t/2)\circ\mathcal{S}_A^{(2k)}(-t/2)^\dagger\circ\mathcal{S}_A^{(2k)}(t/2)\circ\mathcal{Q}_B(t/2)-\mathcal{U}_B(t/2)\circ\mathcal{U}_A(-t/2)^\dagger\circ\mathcal{U}_A(t/2)\circ\mathcal{U}_B(t/2)\right\|_\diamond \\ &\leq 2\left\|\mathcal{U}_A(t/2)-S_A^{(2)}(t/2)\right\|_\diamond + 2\left\|\mathcal{U}_B(t/2)-\mathcal{Q}_B(t/2)\right\|_\diamond. \end{aligned} \tag{2.55}$$

  Since $\Upsilon_1=2$ this matches the induction hypothesis.

- **Inductive Step:** In this scenario we assume that the hypothesis in Equation (2.54) holds for $2l-2$ and we would like to show it holds for $2l$. We do so via the recursive structure given in Equation (2.49) and Definition 2.1, which allows us to express the hypothesis as

$$
\begin{aligned}
&\left\|\mathcal{C}^{(2k,2l)}(t)-\mathcal{S}^{(2l)}(\{A,B\},t)\right\|_{\diamond}\\
&=\Big\|\mathcal{C}^{(2k,2l-2)}(u_l t)^{\circ 2}\circ\mathcal{C}^{(2k,2l-2)}((1-4u_l)t)\circ\mathcal{C}^{(2k,2l-2)}(u_l t)^{\circ 2}\\
&\quad-\mathcal{S}^{(2l-2)}(\{A,B\},u_l t)^{\circ 2}\circ\mathcal{S}^{(2l-2)}(\{A,B\},(1-4u_l)t)\circ\mathcal{S}^{(2l-2)}(\{A,B\},u_l t)^{\circ 2}\Big\|_{\diamond}\\
&\leq 4\left\|\mathcal{C}^{(2k,2l-2)}(u_l t)-\mathcal{S}^{(2l-2)}(\{A,B\},u_l t)\right\|_{\diamond}\\
&\quad+\left\|\mathcal{C}^{(2k,2l-2)}((1-4u_l)t)-\mathcal{S}^{(2l-2)}(\{A,B\},(1-4u_l)t)\right\|_{\diamond}\\
&\leq 4\Upsilon_{l-1}\left(\left\|\mathcal{U}_A(t)-\mathcal{S}_A^{(2k)}(t)\right\|_{\diamond}+\left\|\mathcal{U}_B(t)-\mathcal{Q}_B(t)\right\|_{\diamond}\right)\\
&\quad+\Upsilon_{l-1}\left(\left\|\mathcal{U}_A(t)-\mathcal{S}_A^{(2k)}(t)\right\|_{\diamond}+\left\|\mathcal{U}_B(t)-\mathcal{Q}_B(t)\right\|_{\diamond}\right)\\
&\quad=5\Upsilon_{l-1}\left(\left\|\mathcal{U}_A(t)-\mathcal{S}_A^{(2k)}(t)\right\|_{\diamond}+\left\|\mathcal{U}_B(t)-\mathcal{Q}_B(t)\right\|_{\diamond}\right)\\
&=\Upsilon_{l}\left(\left\|\mathcal{U}_A(t)-\mathcal{S}_A^{(2k)}(t)\right\|_{\diamond}+\left\|\mathcal{U}_B(t)-\mathcal{Q}_B(t)\right\|_{\diamond}\right).
\end{aligned}
\tag{2.56}
$$

Therefore the inductive step holds.

One point of emphasis we would like to make is that we are explicitly not utilizing the smaller times that come with the recursive outer-loop step. This is simply due to the difficulty of bookkeeping for each different time step used and it will be sufficient to use the upper bound of t. We can now continue with our proof of the lemma by substituting in the known Trotter-Suzuki and QDrift error terms, leading us to

$$
\begin{aligned}
\left\|\mathcal{U}(t)-\mathcal{C}^{(2k)}(t)\right\|_{\diamond}&\leq 2\frac{(\Upsilon t)^{2k+1}}{k+1/2}\alpha_{\text{comm}}(\{A,B\},2k)+\Upsilon\left(\left\|\mathcal{U}_A(t)-\mathcal{S}_A^{(2k)}(t)\right\|_{\diamond}+\left\|\mathcal{U}_B-\mathcal{Q}_B(t)\right\|_{\diamond}\right)\\
&\leq 2\frac{(\Upsilon t)^{2k+1}}{k+1/2}(\alpha_{\text{comm}}(\{A,B\},2k)+\Upsilon\alpha_{\text{comm}}(A,2k))+\Upsilon\frac{4\|h_B\|^2t^2}{N_B}.
\end{aligned}
\tag{2.57}
$$

Note that the extra factor of Υ in front of $\alpha_{\text{comm}}(A)$ comes from the fact that we have Υ copies of the A simulation channel as opposed to just one outer-loop decomposition. □

Now that we have bounded the diamond distance error for a single time step we can proceed with our time-slicing arguments to produce a controllable error bound. This will lead us to our final expression for the query cost of a higher order composite method.

Theorem 2.10 (Higher-Order Composite Cost): *Given a time t, error bound ε, and a partitioned Hamiltonian $H = A + B$ the $2k^{\text{th}}$ order Composite channel $\mathcal{C}^{(2k)}$ utilizes at most*

$$C_{\text{Comp}}^{(2k)} \leq \Upsilon(\Upsilon L_A + N_B)\left\lceil \frac{4^{\frac{1}{2k}}(\Upsilon t)^{1+\frac{1}{2}k}}{((2k+1)\varepsilon)^{\frac{1}{2k}}}(\alpha_C(\{A,B\}) + \Upsilon\alpha_C(A)) + \frac{4\Upsilon\|h_B\|^2 t^2}{N_B\varepsilon}\right\rceil, \tag{2.58}$$

gates,where the α_C are both of order $2k$, to meet the error budget given by

$$\left\|\mathcal{U}(t) - \mathcal{C}^{(2k)}(t/r)^{\circ r}\right\|_\diamond \leq \varepsilon. \tag{2.59}$$

Using the upper bounds provided for Trotter-Suzuki and QDrift evolution channels and defining

$$q_B := \frac{\alpha_{\text{Comm}}(B, 2k)}{\alpha_{\text{Comm}}(H, 2k)} \tag{2.60}$$

to capture the amount of "commutator structure" of H that is contained in B, we can rewrite this upper bound as

$$C_{\text{Comp}}^{(2k)} \leq \Upsilon(\Upsilon L_A + N_B)\left\lceil \frac{C_{\text{Trot}}^{(2k)}(H,t,\varepsilon)}{\Upsilon^{1-1/2k}}\frac{(1-q_B)^{1/2k}}{L} + C_{\text{QD}}(H,t,\varepsilon)\frac{\Upsilon}{N_B}\left(\frac{\|h_B\|}{\|h\|}\right)^2\right\rceil. \tag{2.61}$$

Proof: We start off by utilizing standard time-slicing arguments to bound our single step distance

$$\left\|\mathcal{U}(t) - \mathcal{C}^{(2k)}(t/r)^{\circ r}\right\|_\diamond = \left\|\mathcal{U}(t/r)^{\circ r} - \mathcal{C}^{(2k)}(t/r)^{\circ r}\right\|_\diamond \leq r\left\|\mathcal{U}(t/r) - \mathcal{C}^{(2k)}(t/r)\right\|_\diamond. \tag{2.62}$$

Using our results in Lemma 2.9 we can then bound the single time step error as

$$\left\|\mathcal{U}(t/r) - \mathcal{C}^{(2k)}(t/r)\right\|_\diamond \leq \frac{P(t)}{r^{2k+1}} + \frac{Q(t)}{r^2}. \tag{2.63}$$

Our goal is to find a lower bound on r that will guarantee that the above is less than ε. In previous arguments we had monomials in r which allowed us to take roots to compute a bound, but the polynomial $ar^n + br^2 = c$ does not have closed form solutions for $n > 5$. We could try to provide closed solutions for second and fourth order Trotter-Suzuki formulas but we will instead provide a generic bound that will work for all higher order expressions.

Our route to constructing such abound comes from requiring the following inequalities be satisfied for $r_{\text{min}} < r$

$$\frac{P(t)}{r^{2k+1}} + \frac{Q(t)}{r^2} \leq \frac{P(t)}{r_{\text{min}}^{2k+1}} + \frac{Q(t)}{r_{\text{min}}^2} \leq \frac{\varepsilon}{r} \leq \frac{\varepsilon}{r_{\text{min}}}. \tag{2.64}$$

We can then create the intermediate inquality

$$\frac{P(t)}{r^{2k+1}} + \frac{Q(t)}{r^2} \leq \frac{P(t)}{r^2 r_{\min}^{2k-1}} + \frac{Q(t)}{r^2} \leq \frac{P(t)}{r_{\min}^{2k+1}} + \frac{Q(t)}{r_{\min}^2} \leq \frac{\varepsilon}{r}. \tag{2.65}$$

Pulling these powers of r out allow us to simplify the inequality to

$$\frac{1}{\varepsilon}\left(\frac{P(t)}{r_{\min}^{2k-1}} + Q(t)\right) \leq r. \tag{2.66}$$

Our final inequality then comes from using only powers of $r_{\min}$ and noting the fact that $Q(t) > 0$ for all t. We have

$$\begin{aligned} \frac{P(t)}{r_{\min}^{2k+1}} &< \frac{P(t)}{r_{\min}^{2k+1}} + \frac{Q(t)}{r_{\min}^2} \leq \frac{\varepsilon}{r_{\min}} \\ \frac{P(t)}{\varepsilon} &< r_{\min}^{2k} \\ \left(\frac{P(t)}{\varepsilon}\right)^{1/2k} &< r_{\min}, \end{aligned} \tag{2.67}$$

therefore achieving our bound on r, which can be thought of as simply taking r large enough to ensure the Trotterized error is sufficiently small.

By plugging Equation (2.67) into Equation (2.66) yields an explicit lower bound on r as

$$\left(\frac{P(t)}{\varepsilon}\right)^{1/2k} + \frac{Q(t)}{\varepsilon} < r. \tag{2.68}$$

This matches the intuition developed from Trotter-Suzuki formulas in which the error decreases rapidly with the order of the formula, but leads to overall higher gate counts due to exponentially increasing constant factors, namely Υ_k. We now can write down the number of operator exponentials explicitly. As we have Υ stages of interleaved product formulas and each stage has one application of a $2k$ order Trotter-Suzuki formula and one QDrift channel with N_B samples we have $\Upsilon(\Upsilon L_A + N_B)$ operator exponentials per time-slice. By taking the ceiling of the derived bound on r we arrive at

$$C_{\text{Comp}}^{(2k)} \leq \Upsilon(\Upsilon L_A + N_B)\left\lceil \left(\frac{P(t)}{\varepsilon}\right)^{1/2k} + \frac{Q(t)}{\varepsilon} \right\rceil, \tag{2.69}$$

plugging in the definitions of $P(t)$ and $Q(t)$ from Equation (2.51) yields Equation (2.58) in the theorem statement. Equation (2.61) is derived from the following inequalities

$$\frac{P(t)^{1/2k}}{\varepsilon^{1/2k}} \leq \frac{C_{\text{Trot}}^{(2k)}(H,t,\varepsilon)}{L} \frac{(1-q_B)^{1/2k}}{\Upsilon^{1-1/2k}}$$
$$\frac{Q(t)}{\varepsilon} = \Upsilon \frac{C_{\text{QD}}(H,t,\varepsilon)}{N_B} \left(\frac{\|h_B\|}{\|h\|}\right)^2. \tag{2.70}$$

These two inequalities are straightforward substitutions by plugging in results from the product formula costs in Theorem 2.2 and Theorem 2.4 into Equation (2.51), along with the definition of q_B. □

2.5.2 Conditions for Improvement

Now that we have bounded the Composite channel error and computed the query cost we ask the natural question: "When is a Composite channel better than just using Trotter?" Our first answer to this question will be analytic and can be found in Theorem 2.11 and are summarized in Table 1. We will be able to derive asymptotic conditions when a fixed partitioning can outperform either Trotter or QDrift. One issue that arises when making these comparisons is that we are comparing a Composite channel to *two* different simulation methods, each with their own query cost. The relative performance of Trotter to QDrift is dependent on the simulation time t and the error ε. It turns out that the ratio $\frac{C_{\text{QD}}}{C_{\text{Trot}}}$ depends on a power of the ratio $\frac{t}{\varepsilon}$. For very accurate and long simulations we observe that Trotter has superior cost, but whenever error requirements are not high or the simulation time is relatively short QDrift is the more efficient simulation method. One thought experiment to illuminate this is the limit as $t \ll 1$, in which case QDrift can replicate the exact time evolution statistics with a single sample whereas Trotter-Suzuki methods need to implement one operator exponential per term of the Hamiltonian.

Theorem 2.11 (Conditions for Composite Channel Improvements): *Let H be a family of Hamiltonians along with a partitioning scheme to generate a partition $H = A + B$ that varies with L. For a simulation time t and diamond distance error bound of ε, let ξ be the number such that $C_{\text{QD}}^{\xi} = C_{\text{Trott}}^{(2k)}$. There exists asymptotic regimes for the parameters L_A, $\|h_B\|$, and N_B such that*

$$C_{\text{Comp}}^{(2k)} \in o\Big(\min\Big\{C_{\text{Trot}}^{(2k)}, C_{\text{QD}}\Big\}\Big), \tag{2.71}$$

outlined below.

For the case when $C_{\text{Trott}}^{(2k)} < C_{\text{QD}}$, corresponding to $0 < \xi < 1$, if the following are satisfied

1. $L_A(1-q_B)^{1/2k} \in o(L)$,
2. $\|h_B\| \in o\left(\|h\|^{\xi}\left(\frac{\sqrt{\varepsilon}}{t}\right)^{1-\xi}\right)$,
3. $N_B \in \Omega(L_A)$ *and* $N_B \in o\left(\frac{L}{(1-q_B)^{1/2k}}\right)$,

then we have that $C_{\text{Comp}}^{(2k)} \in o\Big(C_{\text{Trot}}^{(2k)}\Big) = o\Big(\min\Big\{C_{\text{Trot}}^{(2k)}, C_{\text{QD}}\Big\}\Big)$.

For the case when $C_{\text{QD}} \leq C_{\text{Trot}}^{(2k)}$, corresponding to $\xi \geq 1$, if the following are satisfied

1. $$L_A \in o\left(L^{\xi}\left(\frac{t^{(2k+1)(\xi-1)}}{\varepsilon^{\xi-1}}\frac{\alpha_C(H)^{\xi}}{\alpha_C(A)+\alpha_C(\{A,B\})}\right)^{1/2k}\right), \tag{2.72}$$
2. $\|h_B\| \in o(\|h\|)$,
3. *and* $N_B \in \Theta(L_A)$,

then we have $C_{\text{Comp}}^{(2k)} \in o\big(C_{\text{QD}}\big) = o\Big(\min\Big\{C_{\text{Trot}}^{(2k)}, C_{\text{QD}}\Big\}\Big)$.

Note that for $\xi = 1$ the conditions on L_A and $\|h_B\|$ are the same in both cases: $L_A \in o(L)$ and $\|h_B\| \in o(\|h\|)$. The conditions on N_B are satisfied by $N_B \in \Theta(L_A)$ as the condition $N_B \in o\left(\frac{L}{(1-q_B)^{1/2k}}\right)$ is not valid when $\xi = 1$.

Proof: We start with the expression for the Composite channel cost from Theorem 2.10 repeated here for clarity

$$C_{\text{Comp}}^{(2k)} \leq \Upsilon(\Upsilon L_A + N_B)\left\lceil \frac{C_{\text{Trot}}^{(2k)}(H,t,\varepsilon)}{\Upsilon^{1-1/2k}}\frac{(1-q_B)^{1/2k}}{L} + C_{\text{QD}}(H,t,\varepsilon)\frac{\Upsilon}{N_B}\left(\frac{\|h_B\|}{\|h\|}\right)^2\right\rceil. \tag{2.73}$$

We will split our analysis up into the two cases outlined in the theorem statement.

- $0 < \xi < 1$

In this scenario we have that $C_{\text{QD}} > C_{\text{Trot}}^{(2k)}$, so we can pull out the $C_{\text{Trot}}^{(2k)}$ cost and parametrize the ratio $\frac{C_{\text{QD}}}{C_{\text{Trot}}^{(2k)}} = C_{\text{QD}}^{1-\xi}$.

$$C_{\text{Comp}}^{(2k)} \leq C_{\text{Trot}}^{(2k)} \Upsilon(\Upsilon L_A + N_B)\left(\frac{(1-q_B)^{1/2k}}{\Upsilon^{1-1/2k} L} + C_{\text{QD}}(H,t,\varepsilon)^{1-\xi}\frac{\Upsilon}{N_B}\left(\frac{\|h_B\|}{\|h\|}\right)^2\right). \tag{2.74}$$

We can then show that $C_{\text{Comp}}^{(2k)} \in o\left(C_{\text{Trot}}^{(2k)}\right)$ if we are able to show that every term in the expansion of the above two factors are $o(1)$. We do so term by term.

$$\begin{aligned}
L_A \in o\left(\frac{L}{(1-q_B)^{1/2k}}\right) &\implies \Upsilon^{1+1/2k}\frac{(1-q_B)^{1/2k} L_A}{L} \in o(1), \\
\|h_B\| \in o\left(\|h\|^{\frac{\xi}{2}}\left(\frac{\sqrt{\varepsilon}}{t}\right)^{1-\xi}\right) &\implies \Upsilon^2\left(\frac{\|h_B\|}{\|h\|}\right)^2 C_{\text{QD}}^{1-\xi} = \left(\frac{t^2}{\varepsilon}\right)^{1-\xi}\frac{\|h_B\|^2}{\|h\|^\xi} \in o(1), \\
N_B \in o\left(\frac{L}{(1-q_B)^{1/2k}}\right) &\implies \Upsilon^{1/2k}(1-q_B)^{1/2k}\frac{N_B}{L} \in o(1).
\end{aligned} \tag{2.75}$$

The last term we have is

$$\Upsilon^3 C_{\text{QD}}^{1-\xi}\frac{\|h_B\|^2}{\|h\|^2}\frac{L_A}{N_B}. \tag{2.76}$$

Using the QDrift cost expression we have that $C_{\text{QD}}^{1-\xi} \leq 4^{1-\xi}\left(\frac{t^2}{\varepsilon}\right)^{1-\xi}\|h\|^{2(1-\xi)}$. This tells us that $C_{\text{QD}}^{1-\xi}\frac{\|h_B\|^2}{\|h\|^2} \in o(1)$ given the assumption $\|h_B\| \in o\left(\|h\|^{\frac{\xi}{2}}\left(\frac{\sqrt{\varepsilon}}{t}\right)^{1-\xi}\right)$. The total term above is then in $o(1)$ given the assumption that $N_B \in \Omega(L_A)$. As this is the last term in the expansion we have completed the $0 < \xi < 1$ case.

- $\xi \geq 1$

In this scenario we have $\frac{C_{\text{Trot}}^{(2k)}}{C_{\text{QD}}} = \left(C_{\text{Trot}}^{(2k)}\right)^{1-\xi}$ which allows us to write

$$C_{\text{Comp}}^{(2k)} \leq C_{\text{QD}} \Upsilon(\Upsilon L_A + N_B)\left(\frac{\left(C_{\text{Trot}}^{(2k)}\right)^{1-\xi}}{\Upsilon^{1-1/2k}}\frac{(1-q_B)^{1/2k}}{L} + \frac{\Upsilon}{N_B}\left(\frac{\|h_B\|}{\|h\|}\right)^2\right). \tag{2.77}$$

We will tackle the hardest term in this expansion first, which is the one involving $\left(C_{\text{Trot}}^{(2k)}\right)^{1-\xi}$. Using the cost expression for Trotter given in Theorem 2.2 we have

$$\begin{aligned}
&\Upsilon^{1+1/2k}(1-q_B)^{1/2k}\frac{L_A}{L}\left(C_{\text{Trot}}^{(2k)}\right)^{1-\xi} \\
&= \Upsilon^{2+1/2k}\left(\frac{t^{1+1/2k}}{\varepsilon^{1/2k}}\right)^{1-\xi}\frac{L_A}{L^\xi}\left(\frac{\alpha_C(A) + \alpha_C(\{A,B\})}{\alpha_C(H)^\xi}\right)^{1/2k}.
\end{aligned} \tag{2.78}$$

This expression is in $o(1)$ given Assumption 1 from the theorem statement. We can then reduce the term involving N_B and $C_{\text{Trot}}^{(2k)}$ to the previous term as $N_B \in \Theta(L_A)$

$$\Upsilon N_B \frac{\left(C_{\text{Trot}}^{(2k)}\right)^{1-\xi}}{\Upsilon^{1-1/2k}} \frac{(1-q_B)^{1/2k}}{L} \in \Theta\left((1-q_B)^{1/2k} \frac{L_A}{L} \left(C_{\text{Trot}}^{(2k)}\right)^{1-\xi}\right) \in o(1), \tag{2.79}$$

where the last inclusion was shown for the previous term. The last two terms in the expansion involving the spectral norms are as follows

$$\begin{aligned} \|h_B\| \in o(\|h\|) &\implies \Upsilon^2 \left(\frac{\|h_B\|}{\|h\|}\right)^2 \in o(1) \\ \|h_B\| \in o(\|h\|) \text{ and } N_B \in \Theta(L_A) &\implies \Upsilon \frac{L_A}{N_B} \left(\frac{\|h_B\|}{\|h\|}\right)^2 \in o(1). \end{aligned} \tag{2.80}$$

As we have shown all four terms in the expansion are $o(1)$ we have that $C_{\text{Comp}}^{(2k)} \in o\big(C_{\text{QD}}\big) = o\Big(\min\Big\{C_{\text{QD}}, C_{\text{Trot}}^{(2k)}\Big\}\Big)$ for $0 < \xi < 1$ which completes the proof. □

Now that we have concrete bounds we would like to build some intuition for the assumptions that go into the theorem. As the expressions become fairly unwieldy in the generic setting we can isolate the scenario where we expect the most benefit from using a Composite framework, which is when $C_{\text{QD}} = C_{\text{Trot}}^{(2k)}$ or $\xi = 1$. The rationale behind this intuition is that if $C_{\text{QD}} \ll C_{\text{Trot}}^{(2k)}$, we can imagine building a composite channel by starting with a solely QDrift partitioning scheme and then moving over the most advantageous terms to the Trotter partition. We have a lot less room until the costs of Trotter begin to add up. Similar logic holds for the $C_{\text{QD}} \gg C_{\text{Trot}}^{(2k)}$ regime. In the intermediate regime $C_{\text{QD}} \approx C_{\text{Trot}}^{(2k)}$ we have a bit more flexibility to move terms around without bumping in to these costly partitions.

Another benefit to analyzing the $\xi = 1$ scenario is that the resulting assumptions needed for cost improvements simplify significantly. The three requirements reduce to the following:

1. $L_A \in o(L)$, which we use the simplification that $\alpha_{\text{comm}}(H) \geq \alpha_{\text{comm}}(A) + \alpha_{\text{comm}}(\{A, B\})$ implies that $L_A \in o(L)$ is sufficient to meet the exact requirement in Theorem 2.10,
2. $\|h_B\| \in o(\|h\|)$,
3. and $N_B \in \Theta(L_A)$.

These convey much more intuition than the generic conditions we proved. Simply put these conditions say that if you can find a partitioning that contains most of the spectral weight of the Hamiltonian in a small number of terms then then the resulting Composite channel will

be asymptotically cheaper than using a single simulation method. This of course only holds rigorously at the ratio of $\frac{t}{\varepsilon}$ such that Trotter and QDrift costs are equal, but we will demonstrate numerically in Section 2.6 that these advantages hold in nearby values of t and ε. To summarize this section we provide the following table that contains the requirements in Theorem 2.11 but in a easier to read format.

	$C_{\text{QD}} > C_{\text{Trot}}^{(2k)} \iff 0 < \xi < 1$	$C_{\text{QD}} \leq C_{\text{Trot}}^{(2k)} \iff \xi \geq 1$
$L_A \in$	$o\left(\frac{L}{(1-q_B)^{1/2k}}\right)$	$o\left(L^{\xi}\left(\frac{t^{1+1/2k}}{\varepsilon^{1/2k}}\right)^{\xi-1}\frac{\alpha_C(H)^{\xi/2k}}{(\alpha_C(A)+\alpha_C(\{A,B\}))^{1/2k}}\right)$
$\|h_B\| \in$	$o\left(\|h\|^{\xi}\left(\frac{\sqrt{\varepsilon}}{t}\right)^{1-\xi}\right)$	$o(\|h\|)$
(Lower Bound) $N_B \in$	$\Omega(L_A)$	$\Omega(L_A)$
(Upper Bound) $N_B \in$	$o\left(\frac{L}{(1-q_B)^{1/2k}}\right)$	$O(L_A)$

Table 1: Summary of asymptotic requirements for parameters of interest when $C_{\text{QD}}^{\xi} = C_{\text{Trot}}^{(2k)}$ to yield $C_{\text{Comp}}^{(2k)} \in o\left(\min\left\{C_{\text{QD}}, C_{\text{Trot}}^{(2k)}\right\}\right)$.

2.5.3 `chop` Partitioning Scheme

As we have seen throughout, the partition used to create a Composite channel has a significant impact on the resulting number of operator exponentials needed. This makes partitioning an important problem, but one that is also fairly challenging as the solution space is 2^L. In this section we show how a simple partitioning scheme called `chop` can create partitions that work exceptionally well for systems with large separations between the largest spectral norm terms and the smallest spectral norms. `chop` creates a partition $A + B$ of a Hamiltonian given a norm cutoff h_{chop}, all terms with spectral norm above h_{chop} are placed into A and all those below are placed into QDrift:

$$A_{\texttt{chop}} := \sum_{i=1}^{L} \mathbb{I}\left[h_i \geq h_{\text{chop}}\right] h_i H_i, \qquad B_{\texttt{chop}} := \sum_{i=1}^{L} \mathbb{I}\left[h_i < h_{\text{chop}}\right] h_i H_i, \tag{2.81}$$

where we use $\mathbb{I}[\text{Proposition}]$ to denote the standard indicator function where $\mathbb{I}[\text{True}] = 1, \mathbb{I}[\text{False}] = 0$.

`chop` will prove to be a very useful partitioning scheme both analytically and numerically. Analytically we will be able to show that it satisfies the conditions outlined in Theorem 2.11 for specific Hamiltonians. Numerically, it is very simple to create a specified partition from a Hamiltonian and further it is straightforward to optimize as the partition can be adjusted with

a a single parameter h_{chop}. This still leaves open the problem of choosing the right number of QDrift samples N_B, but we did not find this parameter an issue to optimize analytically or numerically. In [35] we provided a proabablistic partitioning scheme that is tuned solely through N_B. This scheme was very flexble, we were able to show that it saturates to the Trotter and QDrift costs in the appropriate limits as well as asymptotic cost improvements for very specific scenarios with high probability, but it's complicated analysis makes it an unfit candidate for inclusion in this thesis. Instead, we will focus on showing how `chop` can outperform Trotter or QDrift with rapidly decaying Hamiltonians in the theorem below.

TODO: Make sure that the conditions below are not flipped.

Theorem 2.12 (Simulation Improvements for Exponentially Decaying Hamiltonians): *Let H be a Hamiltonian $H = \sum_i h_i H_i$ such that the spectral norms decay exponentially $h_i = 2^{-i}$. Then the* `chop` *partitioning scheme that places the largest* $\log L$ *terms into Trotter and the remaining terms into QDrift, which corresponds to a norm cutoff of $h_{\text{chop}} = \frac{1}{L}$, satisfy the conditions for asymptotic improvement outlined in Theorem 2.11 whenever the following hold.*

1. $N_B = L_A = \log(L)$.
2. *If $0 < \xi < 1$ ($C_{\text{QD}} > C^{(2k)}_{\text{Trot}}$), then the simulation time is bounded from above by $t \in o\left(L^{\frac{1}{1-\xi}}\sqrt{\varepsilon}\right)$.*
3. *If $\xi \geq 1$ ($C_{\text{QD}} \leq C^{(2k)}_{\text{Trot}}$), then $t^{1+1/2k} \geq \varepsilon^{1/2k}$ and the commutator structure is bounded from* below *by*

$$\alpha_C(H, 2k)^{1/2k} \in \omega\left(\frac{\log(L)^{\frac{1}{\xi}}}{L}\right). \tag{2.82}$$

Proof: As the conditions for improvement depend on $L_A, \|h_B\|$, and N_B, but we know that $L_A = N_B = \log(L)$, all we need to compute is $\|h_B\|$. This is done using straightforward sums:

$$\|h_B\| = \sum_{i=\log(L)+1}^{L} 2^{-i} = 2^{1-(\log(L)+1)} - 2^{-L} = \frac{1}{L} - 2^{-L} \in \Theta(L^{-1}). \tag{2.83}$$

The total norm can be computed similarly

$$\|h\| = \sum_{i=0}^{L-1} 2^{-i} = 1 - 2^{-L} \in \Theta(1). \tag{2.84}$$

Now we just need to check the conditions on each parameter. We will analyze the requirements for ξ for each parameter instead of doing a case by case analysis for the two regimes of ξ.

Starting with N_B, we find that $N_B = L_A \in \Omega(L_A)$ trivially and that $N_B = \log(L) \in o\left(\frac{L}{(1-q_B)^{1/2k}}\right)$ along with $N_B = L_A \in O(L_A)$ guarantee that N_B meets the conditions in Theorem 2.11 straightforwardly.

We then turn to the next simplest parameter $\|h_B\|$. For $\xi \geq 1$ we require $\|h_B\| \in \|h\|$, and since we computed that $\|h_B\| \in \Theta(L^{-1})$ and $\|h\| \in \Theta(1)$ this condition holds. For $\xi \geq 1$ we require $\|h_B\| \in o\left(\|h\|^{\xi}\left(\frac{\sqrt{\varepsilon}}{t}\right)^{1-\xi}\right)$. This can be propagated to a condition on t as

$$\|h_B\| = \Theta(L^{-1}) \in o\left(\left(\frac{\sqrt{\varepsilon}}{t}\right)^{1-\xi}\right) \iff t \in o\left(L^{\frac{1}{1-\xi}}\sqrt{\varepsilon}\right). \tag{2.85}$$

This makes intuitive sense, as $\xi \to 1$ we have $L^{\frac{1}{1-\xi}} \to \infty$ and our requirement then holds for all t. This follows from the fact that the original requirement, in this limit, boils down to $\|h_B\| \in o(\|h\|)$ which is true.

The last term we will need to address is L_A. For $0 < \xi < 1$ we require $L_A \in o(L)$, which is trivially satisfied. For $\xi \geq 1$ we need a couple results. The first will be a simplification of the requirements, if we assume that $t^{1+1/2k} \geq \varepsilon^{1/2k}$, which should be true for simulations of interest, then we have

$$o\left(L^{\xi}\frac{\alpha_C(H)^{\xi/2k}}{(\alpha_C(A)+\alpha_C(\{A,B\}))^{1/2k}}\right) \in o\left(L^{\xi}\left(\frac{t^{1+1/2k}}{\varepsilon^{1/2k}}\right)^{\xi-1}\frac{\alpha_C(H)^{\xi/2k}}{(\alpha_C(A)+\alpha_C(\{A,B\}))^{1/2k}}\right). \tag{2.86}$$

Using the simplification on the left, we then require $L_A = \log(L) \in o\left(L^{\xi}\frac{\alpha_C(H)^{\xi/2k}}{(\alpha_C(A)+\alpha_C(\{A,B\}))^{1/2k}}\right)$. We could either turn this into a condition on ξ or on $\alpha_C(H)$, but it will be simplest to present as a condition on $\alpha_C(H)$.

Now we can use the bounds on $\alpha_C(A)$ and $\alpha_C(\{A,B\})$ derived in Equation (2.13) and Equation (2.14) to argue

$$\alpha_C(A)+\alpha_C(\{A,B\}) \leq 2^{2k}\left(\|h_A\|^{2k+1} + 2k\|h_A\|^{2k+1}\right) = (2k+1)2^{2k}\|h_A\|^{2k+1}. \tag{2.87}$$

This, along with the fact that $\|h_A\| = 1 - \frac{1}{L} \leq 1$ implies

$$\frac{1}{(2k+1)^{1/2k}} \leq \frac{1}{(\alpha_C(A)+\alpha_C(\{A,B\}))^{1/2k}}. \tag{2.88}$$

Moreover, $2k+1 \geq 1$ implies $1 \leq \frac{1}{(2k+1)^{1/2k}}$. This means that

$$L_A \in o\big(L^{\xi}\alpha_C(H)^{\xi/2k}\big) \tag{2.89}$$

is sufficient to satisfy the asymptotic improvement conditions. Once we have this form we are pretty much done, as the following implication follows directly from the definition of $o(\cdot)$ and $\omega(\cdot)$, and this guarantees Equation (2.89)

$$\alpha_C(H)^{1/2k} \in \omega\left(\frac{\log(L)^{1/\xi}}{L}\right) \iff \log(L) \in o\big(L^{\xi}\alpha_C(H)^{\xi/2k}\big). \tag{2.90}$$

We also point out that this requirement seems to make intuitive sense, if the original Hamiltonian has a closed commutator structure, then it does not make sense to do a Composite channel as a Trotter formula would have no error. Equation (2.90) makes this intuition quantitative. □

2.6 Numerics

In this section we turn to studying the performance of Composite channels on benchmark quantum systems. This work was conductly jointly with Pocrnic et al. in [10] in which the real time Composite simulations we outlined in this chapter were studied numerically and extended to "imaginary time" evolution. If real time evolution is characterized by the map $|\psi\rangle \mapsto e^{-iHt}|\psi\rangle$ then imaginary time is given by the map $|\psi\rangle \mapsto e^{-\beta H}|\psi\rangle$. Application of imaginary time evolution maps can be used to prepare thermal states. If we start with a maximally mixed state, then imaginary evolution for time $\frac{\beta}{2}$ gives $\frac{\mathbb{1}}{\dim} \mapsto e^{-\beta H/2}\frac{\mathbb{1}}{\dim}e^{-\beta H/2} = \frac{e^{-\beta H}}{\dim}$. This is clearly not a quantum channel as the output needs to be properly normalized; dealing with these normalization factors constitutes a significant amount of the analytic work, which was performed by Pocrnic, to extend QDrift and Composite simulations to imaginary time evolution in [10].

To analyze the performance of a Composite channel, real or imaginary, we constructed a library [36] can be used to simulate the dynamics of a product formula channel with a given partitioning, number of QDrift terms N_B, time t, and error ε. Numerically we did not measure the diamond distance of the channel, as this involves a fairly costly maximization. This maximization can be computed via a semidefinite program, this becomes prohibitively costly when used to optimize the "hyperparameters" of the simulation, such as the partitioning. We instead used the trace distance which is easier to compute and avoids the issues of bias found when using infidelity. To find the exact gate count needed we used a search procedure over the

minimal number of time steps, either r for Trotter formulas or N_B for QDrift, needed to meet the error threshold ε.

The main metric we used to analyze the performance of Composite channels is the crossover ratio r_{cross}. As the cost of a QDrift channel scales as $O\left(\frac{t^2}{\varepsilon}\right)$ and Trotter scales as $O\left(\frac{t^{1+1/2k}}{\varepsilon^{1/2k}}\right)$ there exists some time t_{cross} such that $C_{\text{QD}}(H, t_{\text{cross}}, \varepsilon) = C^{(2k)}_{\text{Trot}}(H, t_{\text{cross}}, \varepsilon)$. As this is the simulation time that we expect the most flexibility, and therefore cost improvements, for Composite channels we then define the crossover ratio as

$$r_{\text{cross}} := \frac{C_{\text{QD}}(H, t_{\text{cross}}, \varepsilon)}{C_{\text{comp}}(H, t_{\text{cross}}, \varepsilon)} = \frac{C^{(2k)}_{\text{Trot}}(H, t_{\text{cross}}, \varepsilon)}{C_{\text{comp}}(H, t_{\text{cross}}, \varepsilon)}. \tag{2.91}$$

We then study the performance of this crossover ratio as a function of the partitioning of the channel, which we typically use the `chop` partition with cutoff h_{chop}, and the number of QDrift samples N_B. These parameters were then optimized over using Gradient Boosted Regression Trees (GBRT) in Scikit-learn [37]. A summary of the advantages seen for Composite channels can be found below in Table 2 and afterwards more detailed results for each Hamiltonians studied are presented.

Hamiltonian	r_{cross}	**# Terms**	*** - Time**
Hydrogen-3	2.3	62	Real -
5 Site Jellium	9.2	56	Real -
6 Site Jellium	18.8	94	Real -
7 Site Jellium	10.4	197	Real -
7 Spin Graph	4.1	49	Real -
8 Spin Graph	3.9	64	Real -
8 Spin Heisenberg	3.1	29	Imag. -
Hydrogen-3	2.3	62	Imag. -
6 Site Jellium	18.8	94	Imag. -

Table 2: Summary of gate cost improvements observed via the crossover ratio r_{cross} given in Equation (2.91). We observe that savings tend to somewhat improve as the number of terms increases (within the same model), with the exception of Jellium 7 where GBRT struggles with partitioning due to the number of terms.

2.6.1 Hydrogen Chain

Using OpenFermion [38] and PySCF [39] we were able to compute the Hamiltonian for a chain of 3 Hydrogen atoms equally spaced in a line. OpenFermion is a package for managing electronic structure Hamiltonians, it not only generates the required fermionic creation and annihilation operators but can utilize Jordan-Wigner encodings to make the results amenable to simulation on quantum computers. PySCF is a library used to compute the required molecular orbital integrals that give the actual constants in the final Hamiltonian. We used an active space which was given by the minimal basis and is a byproduct of our minimal spin configuration.

The results of the simulations we conducted are found in Figure 1. Details of the partitioning schemes determined by the

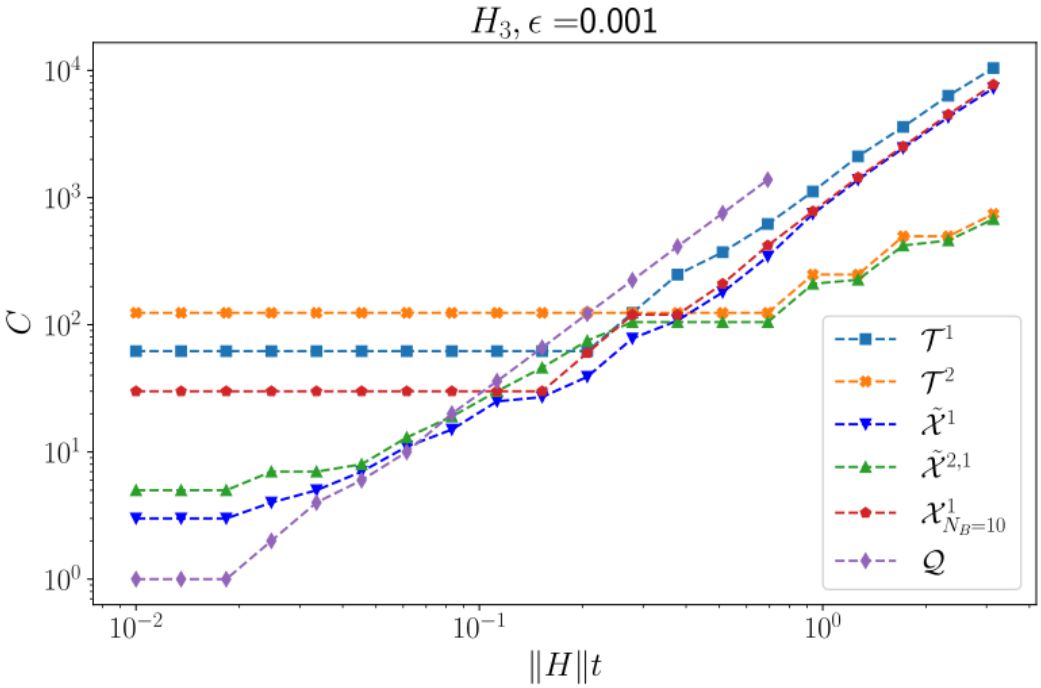
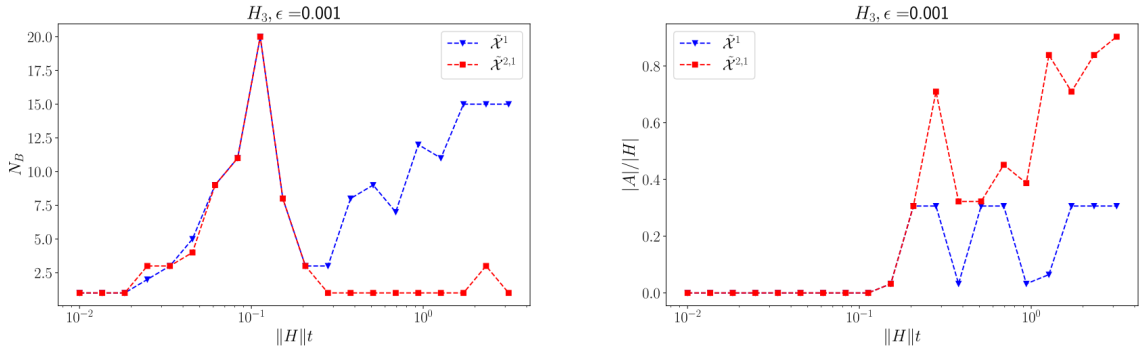


Figure 1: Hydrogen 3 simulation. The crossover time for first order Trotter is around $\|H\|t \approx 0.15$ with a crossover ratio of ≈ 2.3. For second order Trotter the crossover time is ≈ 0.2 with a crossover ratio of ≈ 2. Note that the simulation methods with a tilde denote a GBRT optimized partition and the unmarked method is a hand-tuned `chop` partitioning scheme. **TODO: Replace the $\mathcal{X}$ in the legend with $\mathcal{C}$.**



(a) (b)

Figure 2: (a) Optimal number of QDrift samples N_B for H_3 as determined by GBRT. (b) Spectral weight of the Trotter partition $\|h_A\|$ computed by GBRT applied to h_{chop}, normalized by the total spectral weight of H_3 as a function of simulation time t.

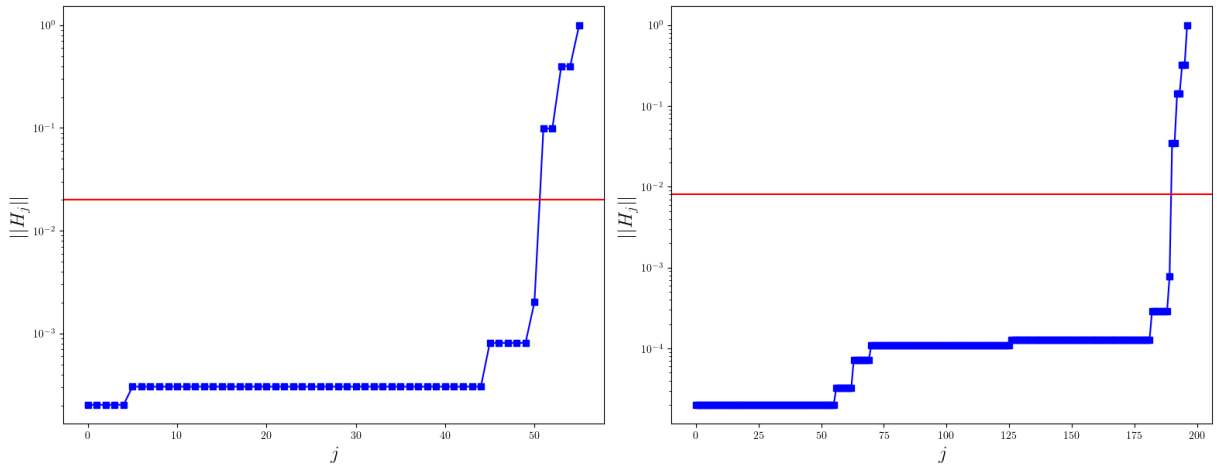
2.6.2 Jellium

Another standard chemistry benchmark system, the Uniform Electron Gas (UEG) which is also known as Jellium, is a collection of free electrons in a solid with a uniform positive potential to serve as nuclei. The Hamiltonian we used is given below

$$H_{\text{Jelly}} = \frac{1}{2}\sum_{p,\sigma} k_p^2 a_{p,\sigma}^\dagger a_{p,\sigma} - \frac{4\pi}{\Omega}\sum_{p\neq q,j,\sigma}\left(\zeta_j \frac{e^{ik_{q-p}\cdot R_j}}{k_{p-q}^2}\right) a_{p,\sigma}^\dagger a_{q,\sigma} + \frac{2\pi}{\Omega}\sum_{(p,\sigma)\neq(q,\sigma'),\nu\neq 0} a_{p,\sigma}^\dagger a_{q,\sigma'}^\dagger a_{q+\nu,\sigma'} \frac{a_{p-\nu,\sigma}}{k_\nu^2}, \tag{2.92}$$

where σ represents a spin, p, q denote momentum eigenvalues, R_j the position of the j^{th} nuclei, ζ_j the atomic number, $k_\nu = 2\pi\nu/\Omega^{\frac{1}{3}}$, and Ω denotes the cell volume. We then use the Jordan-Wigner encoding to represent the creation and annihilation operators as Pauli strings on qubits. For a derivation of this Hamiltonian see Appendix B of [40].

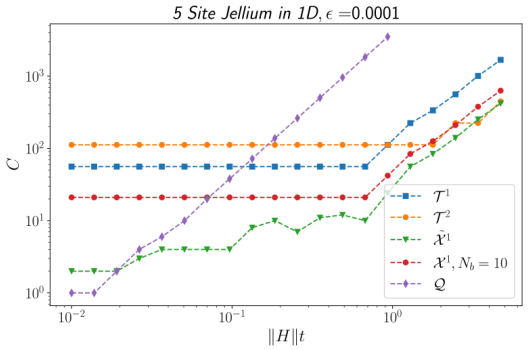
This Hamiltonian serves as a useful benchmark for Composite simulations as there are a lot of terms and the distribution of the spectral norm of each term fits our intuition for Composite channel advantages derived earlier. Figure 3 demonstrates not only the increase in the number of terms as we increase the number of sites used but also how the norms are sharply peaked about the strongest few terms.



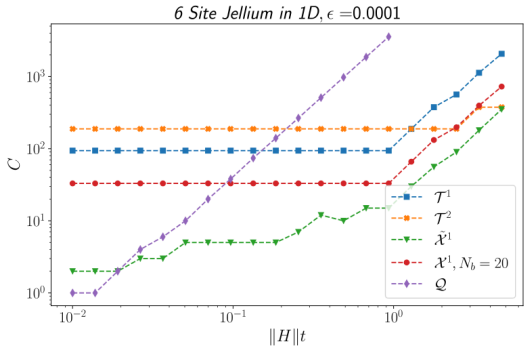
(a) Jellium 5 (b) Jellium 7

Figure 3: Semi-log plots of the spectral norm of the Jellium Hamiltonian. The plots not only show the large increase in the number of terms as we increase te sites but also demonstrate the increasingly concentrated norm in the strongest few terms. The red horizontal line indicates one of the values of h_{chop} used in later simulations.

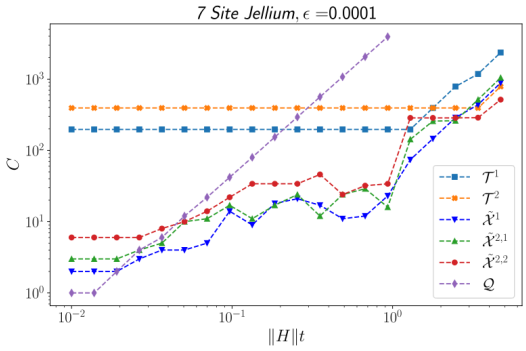
In Figure 4 below we show how the cost of simulating Jellium for various number of sites scales with the normalized simulation time $\|H\|t$. These models are the highest gate cost improvements we observed numerically. For the case of a 6-site Jellium model the Trotter and QDrift cost at t_{cross} is roughly 100 operator exponentials while the Composite channel uses only 7. We find that having more terms in the Hamiltonian allows for greater flexibility in developing partitionings, allowing for more cost savings, but also makes the problem of choosing a partitioning more challenging. We can see this occur with our GBRT chosen `chop` partitioning in Figure 4 (c) where the cost of the Composite channel is non-monotonic with respect to $\|H\|t$.



(a) Jellium 5



(b) Jellium 6



(c) Jellium 7

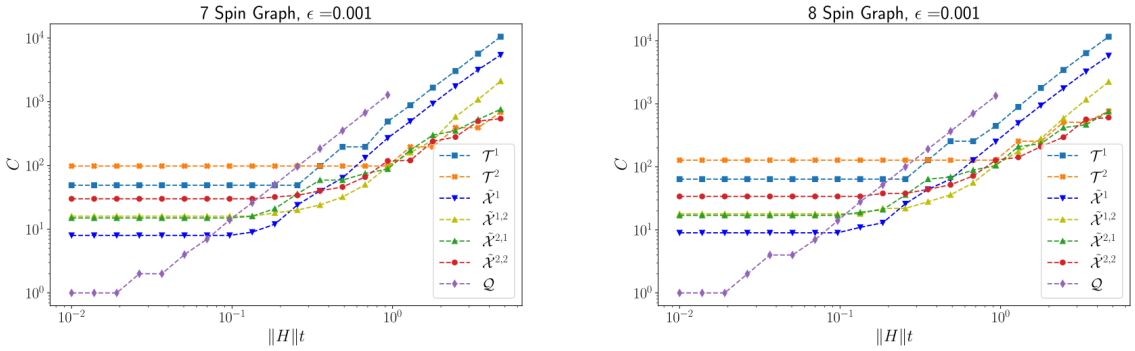
Figure 4: Query costs associated with exact implementation of various product formulas for different Jellium models.

2.6.3 Spin Graphs

The Hamiltonian we explore in this section is a chain of spins on a single line with beyond nearest-neighbor interactions

$$H_{\text{graph}} = \sum_{i>j} e^{-|i-j|} h_{i,j} X_i X_j + \sum_k h_k Z_k, \tag{2.93}$$

where $h_{i,j}$ is a site-dependent coupling constant and h_k is a site-dependent potential. We sampled these values from standard Gaussian random variables to introduce disorder into the system. To keep this system somewhat realistic we require the interactions between sites to decay exponentially with the distance between two sites. This also has the added benefit of introducing some structure into the distribution of the norms of each term in the Hamiltonian. We found modest crossover advantages around $r_{\text{cross}} \approx 4$ for both 7 and 8 spin sites, as seen below in Figure 5.

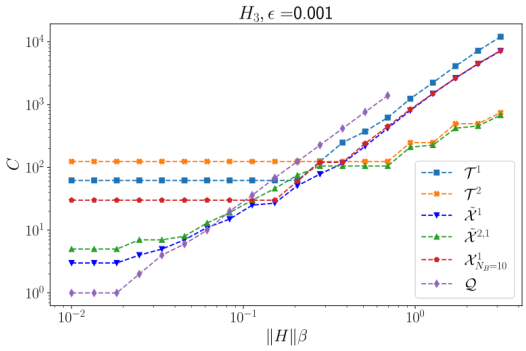


(a) 7 Spin H_{graph} (b) 8 Spin H_{graph}

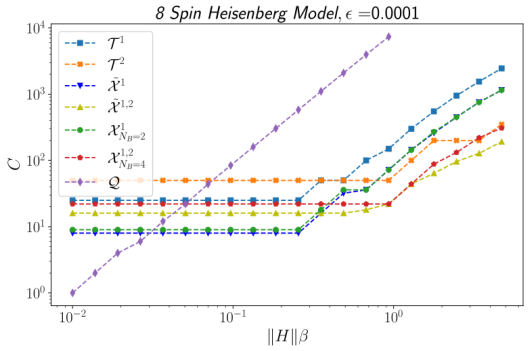
Figure 5: Operator query cost plots for 7 spin model (a) and 8 spin model (b), which have crossover ratios of $r_{\text{cross}} = 4.1$ and $r_{\text{cross}} = 3.9$ respectively.

2.6.4 Imaginary Time Evolutions

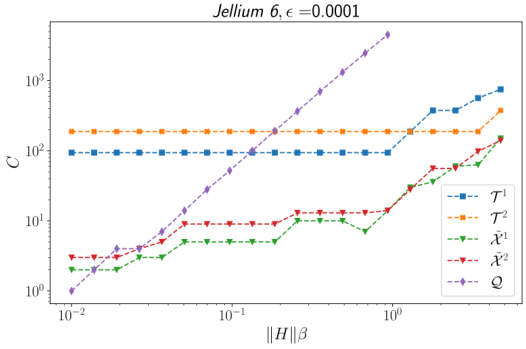
In this section we briefly discuss the application of our Composite simulation approach to implementing imaginary time evolution channels, the results of which are contained below in Figure 6. At a high level we see that the results for imaginary time are comparable to the real time evolutions explored above. We see crossover advantages of similar rates as well, with Composite channels for Jellium outperforming Trotter and QDrift by a factor of ≈ 19, H_3 Composite channels using ≈ 2.3 times less gates, and advantages for a 8 Spin Heisenberg Model are around ≈ 3. The one major distinction we noticed between real and imaginary time simulations came from the 6 site Jellium model at large β, or low-temperature. In this regime we noticed that even the first order Composite channel outperformed a second order Trotter implementation. These simulations suggest that randomized and Composite techniques could be useful in speeding up classical techniques, such as Quantum Monte Carlo [41], which are predominantly based on a Trotter-Suzuki decomposition.



(a) H_3



(b) 8 Spin Heisenberg



(c) 6 Site Jellium

Figure 6: Operator exponential costs for imaginary time simulations. In (a) the crossover advantage is $r_{\text{cross}} = 2.3$, in (b) $r_{\text{cross}} = 3.1$, and in (c) $r_{\text{cross}} = 18.8$.

2.7 Discussion

In this chapter we rigorously showed how to simulate the time evolution of a time-independent Hamiltonian using product formulas. These product formulas are easily implementable on a quantum computer using only single qubit rotations and CNOTs for Hamiltonians that are given as a sum of Pauli operators. We showed how various chemical systems, such as Hydrogen chains and the UEG (Jellium) are naturally expressed in these forms via Jordan-Wigner encodings. The main contribution of this chapter however is the demonstration that splitting these resulting Hamiltonians into two pieces and simulating these two partitions using different product formulas can lead to provably better perfomance. We showed this analytically for systems in which the spectral norm decays exponentially (i.e. $h_i = 2^{-i}$) and gave an explicit partitioning of the terms based on spectral weight, which we denoted `chop`. We verified that these results are not just analytic musings and provided concrete numeric comparisons between each of the

methods, Trotter-Suzuki, QDrift, and Composite, on standard quantum chemistry benchmark systems. We found a range of cost improvements ranging from $2-18$ fold reductions in the number of operator exponentials required.

Chapter 3

Preparing Thermal Quantum States

Thermal states of the form $\frac{e^{-\beta H}}{\mathcal{Z}}$ are ubiquitous in physics. These are the states that we believe physical systems take whenever they are cooled (or heated) to an inverse temperature of β. They are typically used to estimate observables O of interest in physically relevant states, with the thermal average typically denoted $\langle O \rangle_\beta = \mathrm{tr}\left(O \frac{e^{-\beta H}}{\mathcal{Z}}\right)$. These observables could be anything from dipole moments, magnetizations, or two body correlators. Classically these states correspond to the canonical ensemble, a distribution over phase space that tells us the probability of finding a particle at a particular position and momentum when it is in thermal equilibrium. This is not an issue classically, as we have many proofs outlining when systems are ergodic, meaning that the infinite time average is equal to the phase space average.

For quantum systems, however, such results are pretty much nonexistant. One of the first issues one has to deal with is that in a closed quantum system the evolution operator is unitary, meaning that not only is the energy conserved but the distance between two input states is preserved throughout the dynamics. A system ε far away from thermal equilibrium at the beginning will remain ε far away throughout the entire duration of the evolution. This is not what is observed in practice, as a quantum system placed in a dilution refrigerator will eventually cool down to the temperature of the fridge.

To resolve this apparent discrepancy physicists have settled on three main approaches: the Eigenstate Thermalization Hypothesis (ETH), Linbladian based evolutions, and the Repeated Interactions (RI) framework. ETH studies thermalization without giving up the notion of a closed quantum system. In this framework the system is taken to be large, as in thermodynam-

ically large, and thermalization is observed whenever one considers the state of a single particle or a single *local* observable. In chaotic systems these observables can be rigorously shown to appear as if they came from the thermal average $\frac{e^{-\beta H_{\text{local}}}}{Z_{\text{local}}}$, but showing that similar techniques work for non-chaotic systems is a major open question. In fact, the existance of many-body localization serves as a counterexample to a universal ETH, but the existance of these phases of matter in the real world, along with the validity of ETH in general, are heavily debated.

The second approach based on Linbladian schemes abandons the notion of a closed quantum system and instead studies only the effect of the environment on the system. This is a valid idea, as we know from basic quantum information theory that *any* channel on a system can be mimicked with only a quadratically larger Hilbert space. This means that even the effects of an infinitely large environment can in principle be simulated on finite sized quantum devices. This is the approach taken by [3]. Some of the downsides to this approach is that the specific models used for the environment typically make strong assumptions, such as weak coupling, Markovianity, or an infinite number of degrees of freedom, which can break down in specific scenarios.

The last approach is the Repeated Interactions framework which is somewhat of a hybrid of the two previous formalisms. The RI prescription does not abandon the notion of closed quantum systems but instead opts to directly simulate the environment via very small added degrees of freedom. The most straightforward example is to imagine a single photon γ drawn from some black-body thermal spectrum. This single photon then interacts with the system of interest for a brief period, only to then fly off and never interact with the system again. We then can take a new photon from this thermal background and interact with the system again, but with a refreshed photon. The idea is that by repeating this over and over, if the interaction is properly chosen, the many photons can collectively thermalize the system. The main drawback with this approach is that not only does a model for the environment need to be chosen, but physically realistic interactions need to be inserted in order for the system to converge to the thermal state. Existing RI results have shown that single spin environments are sufficient to thermalize single spin systems, or at the largest a three level system.

In this chapter we explore how to lift these restrictions on the RI framework through a randomized interaction approach. The resulting dynamics are not unitary but instead mixed unitary. We show how to analyze this channel in a weak-coupling regime, which is effectively

a Taylor Series with respect to a coupling constant α. This expansion then reveals a Markov chain underlying the resulting channel on the system. By using basic tools from Markov Chain analysis we can then compute the fixed points of the map and bound the number of interactions needed to converge to the fixed point. By showing that the thermal states are the unique fixed points we can then prove thermalization. The beauty of our technique is that it works for any non-degenerate system with or without knowledge of the eigenvalues of the system, although without eigenvalue knowledeg we are only able to show that the thermal state is an approximate fixed point for finite β but exactly fixed in the limit $\beta \to \infty$.

The rest of this chapter is organized as follows. In Section 3.1 we briefly discuss related works in quantum algorithms and provide a summary of the main technical results. In Section 3.2 we develop the weak-coupling expansion and provide necessary analysis about the underlying Markov chain. In Section 3.3 we then take these results and show how to use them to prepare thermal states for single qubit systems and for truncated harmonic oscillators. This section can be viewed as a warmup to the more general results contained in Section 3.4, but we present the results separately as they utilize slightly different techniques and are more readily comparable to existing approaches. In Section 3.4 we study generic systems from two perspectives, one in which no eigenvalue knowledge is present and the other in which eigenvalues are known. Finally we include a discussion on interpretations of these results and possible extensions in Section 3.5.

3.1 Related Work and Main Results

The simulation of quantum systems and materials is among the most promising applications for exponential advantages of digital quantum computers over classical computers [13,42,43]. A critical step in quantum simulation algorithms, as well as other quantum algorithms such as Semi-Definite Program (SDP) solvers [44] and Hamiltonian learning routines [45], is the preparation of good input states, which are typically thermal states $\frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}$. Thermal states at low temperatures (high β) have large overlap with the ground states of the system, indicating that preparing thermal states is just as difficult as the QMA-Hard k-local ground state preparation problem [46].

Many classical algorithms have been developed to estimate the measurement outcomes of quantum experiments with the workhorse behind many of these algorithms typically being some kind of Metropolis-Hastings algorithm [47] to implement a Markov Chain Monte Carlo

(MCMC) program. The Metropolis-Hastings algorithm solves the problem of sampling from arbitrary probability distributions and can be used to estimate partition functions, a #P-Hard problem [48]. Despite the difficulty of the problem it solves and minimal theoretic guarantees on the runtime, the Metropolis-Hastings algorithm has worked resoundingly well in practice. This is in part due to its elegant simplicity and ease of implementation. The algorithm does tend to breakdown in a few important areas though, namely quantum systems with "sign problems" ([49], [50]), very high dimension systems [51], and Hamiltonians with many deep local minima [52]. The sign problem in particular serves as an important impetus for developing quantum computers, which naturally do not have to deal with it. However attempts to naively port the classical Metropolis-Hastings algorithm to quantum computers has been rather difficult due to inherent difficulties with quantum information, such as no-cloning. Initial attempts [53] are rather cumbersome and attempts to deal with the filtering and rejection stages relying on "quantum unwinding" techniques developed by Marriott and Watrous [54]. These complications make the resulting quantum algorithms difficult to analyze.

In recent years new approaches have been developed [4], [5], [55], [56,57], many of which are based on the simulation of Linblad operators from open quantum systems [3]. These algorithms have seen a marked improvement in recent years, ranging from ground state preparation routines with single-ancilla overhead [57] to the first constructions that satisfy a discrete-time detailed balance condition [5]. The correctness of many of these algorithms, such as [6], is based on satisfying the Kubo-Martin-Schwinger (KMS) condition ([58], [59]) which guarantees that the thermal state is a fixed point of the dynamics. The literature on this class of algorithms is already significant and continues to grow, so we point the reader to [5], [60], [4] and [61] for their thorough literature reviews.

One of the main drawbacks to the above approaches is the sheer complexity of the resulting algorithms. These algorithms tend to rely on coherently weighted sums of Heisenberg evolved jump operators and the construction of circuits to simulate the resulting Linbladians is nontrivial, as mentioned in Section 1.2 of [5]. Further, these algorithms tend to require logarithmically more ancilla qubits to allow for the addition of jump operators whereas our routine explicitly utilizes only a single ancilla qubit. Turning to ground states specifically, there exists single ancilla algorithms [57] but we remark that our channel is the first general purpose thermal state preparation routine for finite β that utilizes only one qubit explicitly. Further, our routine

avoids the complication of simulating weighted Linbladians and has incredibly simple circuits only relying on time independent Hamiltonian simulation and Haar 2-designs.

3.1.1 Main Results

The remainder of the paper is split into three main parts. Section Section 3.2 contains a derivation of the weak-coupling expansion in Lemma [??] and outlines the underlying Markov chain behavior in Section [??]. This weak-coupling expansion is presented in as much generality as possible as it may be of use in other applications beyond our thermalization procedure, such as thermometry or spectroscopy. Section Section 3.3 has two theorems concerning single qubit systems and harmonic oscillators, Theorems [??] and [??] respectively, as well as numerics exploring the β and ε dependence of the channel. Section Section 3.4 contains our most general results in Theorems [??] and [??] in which we show that the thermal state is an approximate fixed point for arbitrary Hamiltonians, bound the runtime in terms of a Markovian spectral gap, and finally compute this spectral gap for the ground state limit. The main difference between these two theorems is that one requires only an uppper bound on the spectral norm $\|H_S\|$ while the other takes advantage of eigenvalue knowledge.

One of the key aspects of our thermalization procedure is that the analysis is dependent on the ability to tune the environment gap γ to match the system energy differences. One of our main results in Theorem [??] shows that even if the user cannot tune γ at all and is reduced to uniform guessing within an interval containing all the differences $\Delta_S(i,j)$, then thermalization can still occur. We show that the thermal state at finite β is an approximate fixed state, with the error going to 0 as the coupling constant $\alpha \to 0$. This zero coupling limit can be taken with the opposite limit $t \to \infty$ to yield a nonzero simulation time for the random interaction G. Further, we show that the ground state is exactly the fixed point in the $\beta \to \infty$ limit. In this limit we are also able to bound the total simulation time required as $L \cdot t \in \tilde{O}\left(\frac{\dim_S^{16} \|H_S\|^7}{\delta_{\min}^8 \varepsilon^6}\right)$, where $\delta_{\min}$ represents a "resolution" type distance and is the smallest difference between two distinct eigenvalue differences $|\Delta_S(i,j) - \Delta_S(k,l)|$. When preparing finite β thermal states we pick up an extra factor of $\frac{1}{\tilde{\lambda}_\star(\beta)^7}$ related to the spectral gap of the transition matrix.

3.2 Weak Coupling Expansion

3.2.1 Preliminaries and Notation

We will be working with a bipartite Hilbert space consisting of a system space $\mathcal{H}_S$ with dynamics governed by the Hamiltonian H_S and an environment space $\mathcal{H}_E$ with Hamiltonian H_E. The total space is $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$ with Hamiltonian $H = H_S \otimes \mathbb{1}_E + \mathbb{1}_E \otimes H_E = H_S + H_E$. We will assume without loss of generality that our spaces are encoded in qubits so that $\mathcal{H}_S = \mathbb{C}^{2^n}$ and $\mathcal{H}_E = \mathbb{C}^{2^m}$. We use $\dim_S$ to refer to the dimension of the system's Hilbert space (2^n), $\dim_E$ the environment, and $\dim$ the total Hilbert space. As for the basis we will use for our spaces, we will work directly in the eigenbasis of each Hamiltonian. Besides simplifying our calculations, we can do so because the interaction term we will introduce later is unitarily invariant. We denote these basis in a 1-indexed fashion as

$$H_S = \sum_{i=1}^{2^n} \lambda_S(i)\ |i\rangle\langle i|, H_E = \sum_{j=1}^{2^m} \lambda_E(j)\ |j\rangle\langle j|, H = \sum_{i=1}^{2^n}\sum_{j=1}^{2^m} \lambda(i,j)\ |i,j\rangle\langle i,j|, \tag{3.94}$$

where $\lambda(i,j) = \lambda_S(i) + \lambda_{E(j)}$ and we will sort the eigenvalues in nondecreasing order such that $i > j \Rightarrow \lambda_S(i) \geq \lambda_S(j)$. We note that the ground state in our 1-indexed notation is therefore $|1\rangle\langle 1|$. We also make use of the following notation for the energy differences of the system-environment Hamiltonian and just the system

$$\Delta(i,j|k,l) \coloneqq \lambda(i,j) - \lambda(k,l), \quad \Delta_{S(i,i')} = \lambda_S(i) - \lambda_S(i'), \tag{3.95}$$

and because our eigenvalues are sorted $i > j \Rightarrow \Delta_{S(i,j)} \geq 0$. We will need a few other notations for eigenvalue differences. First we denote the degeneracy of an eigenvalue $\lambda(i,j)$ using $\eta(i,j)$ and the number of times a system eigenvalue *difference* is present as $\eta_\Delta(i,j)$. For example, in a truncated harmonic oscillator with 4 energy levels the lowest gap Δ is present 3 times, so $\eta_\Delta(1,2) = 3$. The second is that we will need to eventually analyze interferences between eigenvalue differences of the system, so we define

$$\delta_{\min} \coloneqq \min_{\Delta_{S(i,j)} \neq \Delta_{S(k,l)}} \left|\Delta_{S(i,j)} - \Delta_{S(k,l)}\right|. \tag{3.96}$$

Note that nothing in this definition prevents one of the summands, say $\Delta_S(k,l)$, from being 0. This implies that $\delta_{\min} \leq \Delta_S(i,j)$ for all i and j.

Currently our dynamics involved a system separated from the environment, so we need to fix this by adding an interaction term $G : \mathcal{H}_S := \mathcal{H}_E \to \mathcal{H}_S \otimes \mathcal{H}_E$. We will choose G randomly via the eigendecomposition

$$G = U_G \Lambda_G U_G^\dagger, U_G \sim \text{Haar}(\mathcal{H}_S \otimes \mathcal{H}_E) \text{ and } \Lambda_G = (-1)^{z_0} Z_1^{z_1} \otimes ... \otimes Z_n^{z_n}, \tag{3.97}$$

where the coefficients z_i are sampled I.I.D via the distribution $\Pr[z_i = 0] = \Pr[z_i = 1] = \frac{1}{2}$. We then add this random interaction term to our system-environment dynamics with a coupling constant α, yielding a total dynamics governed by $H_S + H_E + \alpha G$. We define the following rescaled coupling constant

$$\tilde{\alpha} := \frac{\alpha t}{\sqrt{\dim + 1}}, \tag{3.98}$$

where the dim is the total Hilbert space $\mathcal{H}$ dimension. The rescaling with respect to dim is to capture the factors of $\frac{1}{\dim + 1}$ in the transition amplitudes that appear later and leads to much more compact expressions. This gives a decomposition of expectation values over G into two parts

$$\mathbb{E}_G f(G) = \mathbb{E}_{U_G} \mathbb{E}_{\Lambda_G} f(G) = \mathbb{E}_{\Lambda_G} \mathbb{E}_{U_G} f(G). \tag{3.99}$$

We will use this interaction term to couple our system to an environment prepared in the thermal state $\rho_E(\beta) = \frac{e^{-\beta H_E}}{\mathcal{Z}_E}(\beta)$, where $\mathcal{Z}_E(\beta) = \text{tr}\big(e^{-\beta H_E}\big)$, and then trace out the environment. This gives the definition of our thermalizing channel $\Phi : \mathcal{L}(\mathcal{H}_S) \to \mathcal{L}(\mathcal{H}_S)$ as

$$\Phi(\rho; \alpha, \beta, t) := \text{tr}_{\mathcal{H}_E} \mathbb{E}_G \big[e^{+i(H+\alpha G)t} \rho \otimes \rho_E(\beta) e^{-i(H+\alpha G)t} \big]. \tag{3.100}$$

We will typically drop the implicit parameters of α, β and t. Our goal is to show how this channel can be used to prepare the system in the thermal state $\rho(\beta) = \frac{e^{-\beta H_S}}{\mathcal{Z}(\beta)}$. It will be useful to introduce a fixed-interaction channel $\Phi_G : \mathcal{L}(\mathcal{H}_S \otimes \mathcal{H}_E) \to \mathcal{L}(\mathcal{H}_S \otimes \mathcal{H}_E)$ over the total Hilbert space $\mathcal{H}$ as

$$\Phi_G(\rho \otimes \rho_E; \alpha, t) := e^{+i(H+\alpha t)} \rho \otimes \rho_E e^{-i(H+\alpha G)t}, \tag{3.101}$$

giving us $\Phi(\rho; \alpha, \beta, t) = \text{tr}_{\mathcal{H}_E} \mathbb{E}_G \Phi_G(\rho \otimes \rho_E(\beta); \alpha, t)$. Another alternative notation for Φ that we will use is whenever $\mathcal{H}_E$ is a single qubit with energy gap γ we will use Φ_γ to draw attention to this specific energy gap. We will also make frequent use of indicator functions, denoted $\mathbb{I}[P]$, which is 1 if the predicate P is true and 0 if P is false.

3.2.2 Implementation

In this section we discuss how to implement the channel Φ for a fixed value of γ. First, to prepare the environment thermal state we simply start in the ground state $|0\rangle\langle 0|$ and then with probability $\frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}$ we apply an X gate. This gives the mixed state $\left(1-\frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\right)|0\rangle\langle 0| + \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}|1\rangle\langle 1| = \rho_E(\beta;\gamma)$. Next we need to perform the time evolution $e^{i(H+\alpha G)t}\rho e^{-i(H+\alpha G)t}$. For this we will essentially perform a Composite channel, with H being simulated with Trotter and G with QDrift with the number of samples N_B set to 1. Let $\mathcal{U}_H(t)$ denote the time evolution channel by H for time t and $\mathcal{U}_G(t)$ the same for G. We are going to implement the channel

$$\tilde{\Phi}(t) \coloneqq \mathbb{E}_G \mathcal{U}_G(t) \circ \mathcal{U}_H(t). \tag{3.102}$$

Our goal is to show that this can be repeated

$$\left\| \Phi^{\circ L}(t) - \tilde{\Phi}^{\circ rL}\left(\frac{t}{r}\right) \right\|_\diamond \leq \varepsilon. \tag{3.103}$$

First we can do the following:

$$\left\| \Phi^{\circ L}(t) - \tilde{\Phi}^{\circ rL}\left(\frac{t}{r}\right) \right\|_\diamond \leq L \left\| \Phi(t) - \tilde{\Phi}^{\circ r}\left(\frac{t}{r}\right) \right\|_\diamond, \tag{3.104}$$

so we prove $\left\|\Phi(t) - \tilde{\Phi}^{\circ r}\left(\frac{t}{r}\right)\right\|_\diamond \leq \varepsilon/L$. But this actually doesn't really make sense because we can't really time slice over the expectation value. So how do we prove this?

Since our goal is only to provide an implementation of the thermalizing channel and we are not concerned with minimizing the asymptotic scaling we will use the simpler first-order composite channel given in Lemma 2.5

$$\left\|\mathcal{U}(t) - \mathcal{U}_G(t) \circ \mathcal{U}_H(t)\right\|_\diamond \leq t^2 \left(\sum_{i,j} h_i h_j \left\| \left[H_i, H_j\right] \right\| + \sum_{i,k} h_i g_k \|[H_i, G_k]\| + 4\frac{\|g\|^2}{N_B} \right). \tag{3.105}$$

Now for our interaction we note that it is sufficient to sample a Clifford gate C for G as opposed to a Haar random unitary. This is because all of our results hold for U_G being a $\text{Haar}(n+1)$ 2-design. As the Cliffords are a finite subgroup of the unitaries we can enumerate them C_i, let $|C|$ denote the number of Clifford gates on $n+1$ qubits, which is $2^{O(n^2)}$. For the eigenvalues Λ_G we have 2^{n+2} possible choices. Enumerate all of these via an index k. The spectral norm of a given interaction G_k is 1 as the eigenvalues Λ_{G_k} are all ± 1, which implies the spectral norm of αG_k is α. This gives the norm $\|g\|$ as $\sum_k \|g_k\| = 2^{O(n^2)}\alpha$.

We further will ignore the commutator structure between the Hamiltonian and the interaction and upper bound

$$\sum_{i,k} h_i g_k \|[H_i, G_k]\| \leq \|h\|\|g\|. \tag{3.106}$$

First let's do an outer loop

$$\begin{aligned} \mathbb{E}_G e^{i(H+\alpha G)t} \rho e^{-i(H+\alpha G)t} &= \mathbb{E}_G \left[e^{i\alpha G t} e^{iHt} \rho e^{-iHt} e^{-i\alpha G t} + E_{H,\alpha G} \right] \\ &= \mathbb{E}_G \mathcal{U}_G \circ \mathcal{U}_H(\rho) + \mathbb{E}_G E_{H,\alpha G} \end{aligned} \tag{3.107}$$

Using the triangle inequality allows us to write

$$\left\| \mathbb{E}_G E_{H,\alpha G} \right\| \leq \mathbb{E}_G \left\| E_{H,\alpha G} \right\| \leq \mathbb{E}_G \alpha \frac{t^2}{2} \|[H, G]\| \leq \alpha \|h\| t^2, \tag{3.108}$$

because $\|G\| = 1$ for all G and we can upper bound the commutator $\|[H, G]\| \leq 2\|H\|\|G\|$.

3.2.3 First and Second Order Expansion

In order to understand our thermalizing channel Φ we will compute a Taylor Series for the output of the channel with respect to the coupling constant α. We will perform the α expansion about $\alpha = 0$ and we will use the mean value form of the remainder, in which we are guaranteed a special value $\alpha_\star \in (0, \infty)$ such that the final derivative evaluated at $\alpha_\star$ is the exact amount needed. We use a second-order expansion and will need to explicitly compute terms up to order α^2, which will give the following expansion

$$\Phi(\rho; \alpha) = \Phi(\rho; 0) + \alpha \frac{\partial}{\partial \alpha} \Phi(\rho; \alpha) \mid_{\alpha=0} + \frac{\alpha^2}{2} \frac{\partial^2}{\partial \alpha^2} \Phi(\rho; \alpha) \mid_{\alpha=0} + R_\Phi(\rho; \alpha_\star). \tag{3.109}$$

We use

$$\mathcal{T}(\rho) := \frac{\alpha^2}{2} \frac{\partial^2}{\partial \alpha^2} \Phi(\rho; \alpha) \mid_{\alpha=0} = \frac{\alpha^2}{2} \operatorname{tr}_{\mathcal{H}_E} \mathbb{E}_G \left[\frac{\partial^2}{\partial \alpha^2} \Phi_{G(\rho;\alpha)} \mid_{\alpha=0} \right] \tag{3.110}$$

to denote the transition terms, as it will be revealed that the first two terms do not cause transitions in the system state, and R_Φ to denote the remainder. Further we will often leave the dependence on the α parameter implicit and only include it when necessary.

We start off with the $O(\alpha^0)$ term, which can be trivially computed as

$$\Phi(\rho; 0) = \operatorname{tr}_{\mathcal{H}_E} \int e^{i(H+\alpha G)t} \rho \otimes \rho_{E(\beta)} e^{-i(H+\alpha G)t} dG \mid_{\alpha=0} = e^{iHt} \rho e^{-iHt}. \tag{3.111}$$

We then see that if $[\rho, H] = 0$ then $\Phi(\rho; 0) = \mathbb{1}(\rho)$, and as we restrict ourselves to such input states we will use this throughout the remainder of the paper. The next order correction is the $O(\alpha^1)$ term.

Theorem 3.1 (First Order Φ): *Let Φ be the thermalizing quantum channel given by Equation (3.100) and G a randomly chosen interaction term such that $\mathbb{E}_G[G] = 0$. The $O(\alpha)$ term in the weak-coupling expansion in Equation (3.109) vanishes*

$$\frac{\partial}{\partial \alpha}\Phi(\rho; \alpha) \mid_{\alpha=0} = 0. \tag{3.112}$$

Proof: We start by moving the α derivative through the linear operations of partial tracing and integrals so that it can act on the fixed interaction map Φ_G

$$\begin{aligned}\frac{\partial}{\partial \alpha}\Phi(\rho) \mid_{\alpha=0} &= \frac{\partial}{\partial \alpha}\,\mathrm{tr}_{\mathcal{H}_E}\left(\int \Phi_G(\rho) dG\right) \mid_{\alpha=0} \\ &= \mathrm{tr}_{\mathcal{H}_E}\left(\int \frac{\partial}{\partial \alpha}\Phi_{G(\rho)} dG \mid_{\alpha=0}\right).\end{aligned} \tag{3.113}$$

Now we use the expression for Φ_G in Eq. Equation (3.101) to compute the derivatives,

$$\begin{aligned}\frac{\partial}{\partial \alpha}\Phi_G(\rho) &= \left(\frac{\partial}{\partial \alpha}e^{+i(H+\alpha G)t}\right)\rho \otimes \rho_E e^{-i(H+\alpha G)t} + e^{+i(H+\alpha G)t}\rho \otimes \rho_E\left(\frac{\partial}{\partial \alpha}e^{-i(H+\alpha G)t}\right) \\ &= \left(\int_0^1 e^{is(H+\alpha G)t}(itG)e^{i(1-s)(H+\alpha G)t}ds\right)\rho \otimes \rho_E e^{-i(H+\alpha G)t} \\ &\quad + e^{i(H+\alpha G)t}\rho \otimes \rho_E\left(\int_0^1 e^{-is(H+\alpha G)t}(-itG)e^{-i(1-s)(H+\alpha G)t}ds\right).\end{aligned} \tag{3.114}$$

We can set $\alpha = 0$ in the above and introduce the expectation over G that will be required

$$\begin{aligned}&\mathbb{E}_G\left[\frac{\partial}{\partial \alpha}\Phi_{G(\rho)} \mid_{\alpha=0}\right] \\ &= it\mathbb{E}_G\int_0^1 e^{isHt}Ge^{-isHt}ds\rho \otimes \rho_E - ite^{+iHt}\rho \otimes \rho_E\mathbb{E}_G\int_0^1 e^{-isHt}Ge^{-i(1-s)Ht}ds \\ &= it\int_0^1 e^{isHt}\mathbb{E}_G[G]e^{-isHt}ds\rho \otimes \rho_E - ite^{+iHt}\rho \otimes \rho_E\int_0^1 e^{-isHt}\mathbb{E}_G[G]e^{-i(1-s)Ht}ds \\ &= 0,\end{aligned} \tag{3.115}$$

where the last step follows from our assumption that $\mathbb{E}_G[G] = 0$. □

Now we move on to the $O(\alpha^2)$ term in the weak-coupling expansion of Φ. We first will compute the combined system-environment output of a generic system-environment basis state and we note that this result holds for an arbitrary dimension environment. We will use this to draw two results: the first being for a single qubit environment the transition amplitudes of just the system can be split into on-resonance and off-resonance terms based on the tuning of the environment qubit Hamiltonian. The second result is that coherences are not introduced to the state at this order of Φ, meaning if an input density matrix ρ is diagonal then $(\mathbb{1} + \mathcal{T})(\rho)$ will also be diagonal. This will be crucial for our later understanding of the channel as a Markov chain.

Lemma 3.2: *Given a system Hamiltonian H_S, an environment Hamiltonian H_E, a simulation time t, and coupling coefficient α, let Φ_G denote the time evolution channel under a fixed interaction term G, let χ denote the following coherence prefactor*

$$\chi(i,j) := \sum_{a,b:\Delta(i,j,|a,b)\neq 0} \frac{1 - i\Delta(i,j|a,b)t - e^{-i\Delta(i,j|a,b)t}}{\Delta(i,j|a,b)^2}, \tag{3.116}$$

and let $\eta(i,j)$ denote the degeneracy of the $(i,j)^{\text{th}}$ eigenvalue of $H = H_S + H_E$. Then the $O(\alpha^2)$ term of $\mathbb{E}_G \Phi_G$, where G is any random interaction with Haar random eigenvectors and random eigenvalues such that $\mathbb{E}_G[G] = 0$ and the eigenvalues satisfy $\mathbb{E}_G[\lambda_G(i)\lambda_G(j)] = \delta_{i,j}$, in a weak-coupling expansion is given by

$$\begin{aligned}
&\frac{\alpha^2}{2}\mathbb{E}_G\left[\frac{\partial^2}{\partial\alpha^2}\Phi_G(|i,j\rangle\langle k,l|)|_{\alpha=0}\right] \\
&= -\frac{\alpha^2 e^{i\Delta(i,j|k,l)t}}{\dim + 1}\left(\chi(i,j) + \chi(k,l)^* + \frac{t^2}{2}(\eta(i,j) + \eta(k,l))\right)|i,j\rangle\langle k,l| \\
&\quad + \langle i,j|k,l\rangle \frac{\alpha^2 t^2}{\dim + 1}\sum_{a,b}\operatorname{sinc}^2\left(\Delta(i,j|a,b)\frac{t}{2}\right)|a,b\rangle\langle a,b|
\end{aligned} \tag{3.117}$$

For $|i,j\rangle = |k,l\rangle$ the above expression simplifies to

$$\begin{aligned}
&\frac{\alpha^2}{2}\mathbb{E}_G\left[\frac{\partial^2}{\partial\alpha^2}\Phi_G(|i,j\rangle\langle i,j|)|_{\alpha=0}\right] \\
&= \tilde{\alpha}^2 \sum_{(a,b)\neq(i,j)} \operatorname{sinc}^2\left(\Delta(i,j|a,b)\frac{t}{2}\right)(|a,b\rangle\langle a,b| - |i,j\rangle\langle i,j|)
\end{aligned} \tag{3.118}$$

which also demonstrates that $\operatorname{tr}\mathcal{T}(\rho) = 0$ for ρ such that $[\rho, H_S] = 0$.

The proof of this lemma uses similar techniques to the proof of Theorem 3.1 but is significantly more technical and can be found in Appendix B.1.

Next we will compute the effects of the channel on just the system alone which involves computing the partial trace $\operatorname{tr}_{\mathcal{H}_E}$. We can either do this for a generic environment, which results in summations over $\mathcal{H}_E$ floating around, or specialize to a specific choice of $\mathcal{H}_E$ and compute the summation. For the remainder of this paper we will choose the latter option with a single qubit environment $\mathcal{H}_E = \mathbb{C}^2$ and denote the Hamiltonian $H_E = \begin{pmatrix} 0 & 0 \\ 0 & \gamma \end{pmatrix}$. Our environment input states then become

$$\rho_E(\beta) = \frac{e^{-\beta H_E}}{\mathcal{Z}_E(\beta)} = \frac{1}{1+e^{-\beta\gamma}}\ |0\rangle\langle 0| + \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\ |1\rangle\langle 1| := q(0)\ |0\rangle\langle 0| + q(1)\ |1\rangle\langle 1|, \tag{3.119}$$

where we will use the environment qubit probabilities $q(0)$ and $q(1)$ in calculations for brevity. It will turn out that the value chosen for γ is highly critical to the convergence of our algorithm, tuning it to match eigenvalue *differences* of the system H_S will allow us to analyze the convergence of the algorithm. As we can see in Equation (3.117) there will be a lot of sinc functions used, we will characterize a sinc function as being on-resonance or off-resonance if the inputs are sufficiently close to zero (the max for sinc). As for how close "sufficiently close" actually is will depend on various parameters, such as t, α, ε, and the spectral properties of H_S.

Theorem 3.3 (Second-Order Expansion $\mathcal{T}$): *Let $\mathcal{T}$ denote the second-order correction for a weak coupling expansion for a thermalizing channel Φ with a single qubit environment. The following properties hold for the second order term.*

1. *The transition element from $|i\rangle\langle i|$ to $|j\rangle\langle j|$, for $i \neq j$, is given by*

$$\langle j|\mathcal{T}(|i\rangle\langle i|)|j\rangle = \tilde{\alpha}^2\left(\operatorname{sinc}^2\left(\Delta_S(i,j)\frac{t}{2}\right) + \frac{1}{1+e^{-\beta\gamma}}\operatorname{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right)\right.$$
$$\left. + \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\operatorname{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right)\right). \tag{3.120}$$

2. *For same-state transitions $|i\rangle\langle i|$ to $|i\rangle\langle i|$ we have*

$$\langle i|\mathcal{T}(|i\rangle\langle i|)|i\rangle = -\sum_{j\neq i}\langle j|\mathcal{T}(|i\rangle\langle i|)|j\rangle, \tag{3.121}$$

   which follows from $\operatorname{tr}\mathcal{T}(\rho) = 0$ as shown in Lemma 3.2.

3. *There are no coherences, or off-diagonal density matrix elements, introduced in the system up to $O(\alpha^2)$, or mathematically*

$$j \neq k \Longrightarrow \langle j|\mathcal{T}(|i\rangle\langle i|)|k\rangle = 0. \tag{3.122}$$

Before we prove this result we will introduce the concept of on- and off-resonant transitions which we give below.

Definition 3.4 (On and Off Resonant Transitions): *The transition elements in Equation (3.120) can be divided into on-resonance and off-resonance transitions based on the arguments to the* sinc *function. We define the on-resonance transitions as*

$$\begin{aligned}\langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle \coloneqq \tilde{\alpha}^2\frac{1}{1+e^{-\beta\gamma}}\mathbb{I}[|\Delta_S(i,j)-\gamma|\leq\delta_{\min}]\,\text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right)\\ +\tilde{\alpha}^2\frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\mathbb{I}[|\Delta_S(i,j)+\gamma|\leq\delta_{\min}]\,\text{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right)\end{aligned} \tag{3.123}$$

and the off-resonance terms as

$$\begin{aligned}\langle j|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|j\rangle \coloneqq \tilde{\alpha}^2\frac{1}{1+e^{-\beta\gamma}}\mathbb{I}[|\Delta_S(i,j)-\gamma|>\delta_{\min}]\,\text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right)\\ +\tilde{\alpha}^2\frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\mathbb{I}[|\Delta_S(i,j)+\gamma|>\delta_{\min}]\,\text{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right)\\ +\tilde{\alpha}^2\,\text{sinc}^2\left(\Delta_S(i,j)\frac{t}{2}\right).\end{aligned} \tag{3.124}$$

For the same-state transitions $|i\rangle\langle i|$ to $|i\rangle\langle i|$ the on- and off-resonance transitions are equal to

$$\begin{aligned}\langle i|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|i\rangle &= -\sum_{j\neq i}\langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle\\ \text{and}\quad \langle i|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|i\rangle &= -\sum_{j\neq i}\langle j|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|j\rangle.\end{aligned} \tag{3.125}$$

We will now use these definitions to prove Theorem 3.3.

Proof of Theorem 3.3: The bulk of this proof will be based on straightforward reductions from Equation (3.117). To start we will first show that no off-diagonal elements are introduced to the density matrix. By taking the (j,k) matrix element of the output from Equation (3.117) we see

$$\begin{aligned}\langle j|\mathcal{T}(|i\rangle\langle i|)|k\rangle &= \sum_{l,m}\frac{e^{-\beta\lambda_E(m)}}{1+e^{-\beta\lambda_E(m)}}\langle j,l|\frac{\alpha^2}{2}\mathbb{E}_G\left[\frac{\partial^2}{\partial\alpha^2}\Phi_G(|i,m\rangle\langle i,m|)\;|_{\alpha=0}\right]|k,l\rangle\\ &= -\sum_{l,m}q(m)\tilde{\alpha}^2(\chi(i,m)+\chi(i,m)^*+t^2\eta(i,m))\langle j,l|i,m\rangle\langle i,m|k,l\rangle\\ &\quad+\sum_{l,m}q(m)\sum_{a,b}\tilde{\alpha}^2\,\text{sinc}^2\left(\Delta(i,m|a,b)\frac{t}{2}\right)\langle j,l|a,b\rangle\langle a,b|k,l\rangle\\ &= 0,\end{aligned} \tag{3.126}$$

where we introduce $q(m)$ for $m = 0, 1$ to be a placeholder for the prefactors in Equation (3.117) and the last equality is due to the fact that $j \neq k$ implies that $\langle j, l|i, m\rangle$ and $\langle i, m|k, l\rangle$ cannot both be nonzero and likewise for $\langle j, l|a, b\rangle$ and $\langle a, b|k, l\rangle$.

Since we have shown that coherences are not introduced to our system we can focus on the transitions from diagonal entries to diagonal entries in ρ. We make heavy use of Eq. eqref{eq:el_gigante_dos} which tells us that for $i \neq k$ the system-environment transition amplitude is

$$\frac{\alpha^2}{2}\langle k, l|\mathbb{E}_G\left[\frac{\partial^2}{\partial\alpha^2}\Phi_G(|i, j\rangle\langle i, j|)\ |_{\alpha=0}\right]|k, l\rangle = \tilde{\alpha}^2\,\mathrm{sinc}^2\Big(\Delta(i, j \mid k, l)\frac{t}{2}\Big). \tag{3.127}$$

Now because all the operations present in the above expression are linear we can compute this map for the initial environment state $\rho_E(\beta)$ straightforwardly. Taking the output of this linear combination and computing the trace over the environment then gives us the expression for $\mathcal{T}$ using the assumption that the environment is a single qubit we find using the definition of γ and Δ_S in Equation (3.95)

$$\begin{aligned}
\langle j|\mathcal{T}(|i\rangle\langle i|)|j\rangle &= \sum_{k,l} q(k)\frac{\alpha^2}{2}\langle j, l|\mathbb{E}_G\left[\frac{\partial^2}{\partial\alpha^2}\Phi_{G(|i,k\rangle\langle i,k|)}\ |_{\alpha=0}\right]|j, l\rangle \\
&= \tilde{\alpha}^2\sum_{k,l} q(k)\,\mathrm{sinc}^2\left(\frac{\Delta(i, k \mid j, l)t}{2}\right) \\
&= \tilde{\alpha}^2\left(q(0)\,\mathrm{sinc}^2\left(\frac{\Delta(i, 0 \mid j, 0)t}{2}\right) + q(0)\,\mathrm{sinc}^2\left(\frac{\Delta(i, 0 \mid j, 1)t}{2}\right)\right) \\
&\quad +\tilde{\alpha}^2\left(q(1)\,\mathrm{sinc}^2\left(\frac{\Delta(i, 1 \mid j, 0)t}{2}\right) + q(1)\,\mathrm{sinc}^2\left(\frac{\Delta(i, 1 \mid j, 1)t}{2}\right)\right) \\
&= \tilde{\alpha}^2\left(q(0)\,\mathrm{sinc}^2\left(\frac{\Delta_S(i, j)t}{2}\right) + q(0)\,\mathrm{sinc}^2\left(\frac{(\Delta_S(i, j) - \gamma)t}{2}\right)\right) \\
&\quad +\tilde{\alpha}^2\left(q(1)\,\mathrm{sinc}^2\left(\frac{(\Delta_S(i, j) + \gamma)t}{2}\right) + q(1)\,\mathrm{sinc}^2\left(\frac{\Delta_S(i, j)t}{2}\right)\right),
\end{aligned} \tag{3.128}$$

where we see that combining the terms with $\mathrm{sinc}^2\big(\Delta_{S(i,j)}\frac{t}{2}\big)$, as $q(0) + q(1) = 1$, we immediately get Equation (3.120).

To classify these terms as on-resonance or off-resonance we will focus on the argument to the sinc function, which is of the form $\Delta_{S(i,j)}\frac{t}{2}$ or $\big(\Delta_{S(i,j)} \pm \gamma\big)\frac{t}{2}$. The idea is that we will take t large enough so that only the energy differences that are less than $\delta_{\min}$, as defined in

Equation (3.96), will be non-negligible. Clearly the term $\tilde{\alpha}^2 \operatorname{sinc}^2\left(\frac{\Delta_S(i,j)t}{2}\right)$ will always be off-resonance, as $\delta_{\min} \leq \Delta_S(i,j)$.

Now we have three terms to classify as either on-resonance or off-resonance, we refer to each term by their argument to the sinc function. The first we can categorically declare as being off-resonance is the $\Delta_{S(i,j)}$ term. By [??] we know $\operatorname{sinc}^2\left(\Delta_{S(i,j)}\frac{t}{2}\right) \leq \frac{4}{\delta_{\min}^2 t^2}$, which we will make arbitrarily small in later sections. The other two can only be classified as on or off resonance depending if $\Delta_{S(i,j)}$ is positive or negative. If $i > j$ then we know that $\Delta_{S(i,j)} \geq 0$ and therefore $\operatorname{sinc}^2\left(\left(\Delta_{S(i,j)} - \gamma\right)\frac{t}{2}\right)$ term can be close to 1 if $\gamma \approx \Delta_{S(i,j)}$, which also shows the $\Delta_{S(i,j)} + \gamma$ term is off-resonance for all γ. We say that the $\Delta_{S(i,j)} - \gamma$ term in this scenario is on-resonance if $|\Delta_{S(i,j)} - \gamma| \leq \delta_{\min}$. This classification is best described symbolically as

$$i > j \text{ and } |\Delta(i,j) - \gamma| \leq \delta_{\min} \implies \langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle = \tilde{\alpha}^2 q(0) \operatorname{sinc}^2\left((\Delta_S(i,j) - \gamma)\frac{t}{2}\right). \tag{3.129}$$

The $q(0)$ prefactor indicates that the ancilla started in it's low energy state and since sinc^2 is symmetric we can write the argument as $\gamma\Delta_S(i,j)$ which can be remembered as the ancilla gaining γ amount of energy and the system losing $\Delta_S(i,j)$. In this scenario the $\Delta_S(i,j) + \gamma$ term is therefore put in the off-resonance map

$$\begin{aligned} &i > j \text{ and } |\Delta_S(i,j) - \gamma| \leq \delta_{\min} \\ \implies &\langle j|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|j\rangle = \tilde{\alpha}^2\left(\operatorname{sinc}^2\left(\Delta_S(i,j)\frac{t}{2}\right) + q(1)\operatorname{sinc}^2\left((\Delta_S(i,j) + \gamma)\frac{t}{2}\right)\right). \end{aligned} \tag{3.130}$$

Now for $i < j$ we find that the on-resonance term is

$$i < j \text{ and } |\Delta_S(i,j) + \gamma| \leq \delta_{\min} \implies \langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle = \tilde{\alpha}^2 q(1) \operatorname{sinc}^2\left((\Delta_S(i,j) + \gamma)\frac{t}{2}\right). \tag{3.131}$$

Similarly to before the $q(1)$ prefactor tells us the ancilla starts in the excited state. This matches with the energy argument by noting that $\Delta_S(i,j) \leq 0$ and that the argument to sinc is symmetric, which allows us to write it as $|\Delta_S(i,j)| - \gamma$; indicating that the system gains energy $|\Delta_S(i,j)|$ and the ancilla energy *drops* by $-\gamma$ (therefore increases by γ). In this scenario the $\Delta_S(i,j) - \gamma$ term is off-resonance and we have

$$\begin{aligned} &i < j \text{ and } |\Delta_S(i,j) + \gamma| \leq \delta_{\min} \\ \implies &\langle j|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|j\rangle = \tilde{\alpha}^2\left(\operatorname{sinc}^2\left(\Delta_S(i,j)\frac{t}{2}\right) + q(0)\operatorname{sinc}^2\left((\Delta_S(i,j) - \gamma)\frac{t}{2}\right)\right) \end{aligned} \tag{3.132}$$

Now to compute the $i = j$ case, it is sufficient to utilize our results from the $i \neq j$ scenario. This is because our second order correction has zero trace $\mathrm{tr}(\mathcal{T}(\rho)) = 0$ from , so we can define the on-resonance and off-resonance terms as the following

$$\begin{aligned}\langle i|\mathcal{T}(|i\rangle\langle i|)|i\rangle &= -\tilde{\alpha}^2 \sum_{k\neq i} \langle k|\mathcal{T}(|i\rangle\langle i|)|k\rangle \\ &= -\tilde{\alpha}^2 \sum_{k\neq i} \langle k|(\mathcal{T}_{\text{on}}(|i\rangle\langle i|) + \mathcal{T}_{\text{off}}(|i\rangle\langle i|))|k\rangle \\ &=: \langle i|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|i\rangle + \langle i|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|i\rangle \end{aligned} \tag{3.133}$$

By plugging in Equation (3.129) and Equation (3.131) to the above we are done with the self-transition terms. □

3.2.4 Markovian Dynamics and Error Terms

Now that we have fully computed the significant contributors to the output of our channel Φ, we move on to characterize the behavior of the channel as a Markov chain with noise. A Markov chain is a random process that involves a walker transitioning to vertices on a graph wherein the probability of transition does not depend on the history of the walker. Specifically, in this context we view the vertices in this graph as the eigenstates of the Hamiltonian. The repeated interaction model because of the lack of coherences in the weak coupling limit can be interpreted as a Markov process over these eigenstates with transitions probabilities given by the above analysis.

Specifically, the Markov chain is dictated by the $\Phi(\rho; 0)$ and $\mathcal{T}_{\text{on}}$ terms in the weak-coupling expansion, for $[\rho, H_S] = 0$ we showed that $\Phi(\rho; 0) = \mathbb{1}(ho)$, so from now on we will specifically only deal with such density matrices and characterize the zeroth order term as an identity map. As for the Markov chain, we will use normal font to denote matrices, such as I for the identity matrix and T for the transition term added on. We use $\vec{e}_i$ to denote the basis vector associated with the quantum state $|i\rangle\langle i|$ and p to denote the probability vector for ρ associated with its eigenvalues.

Lemma 3.5 (Quantum Dynamics to Classical Markov Chain): *Let T be the matrix defined by*

$$\vec{e}_i^{\top} T \vec{e}_j := \langle i|\mathcal{T}_{\text{on}}(|j\rangle\langle j|)|i\rangle. \tag{3.134}$$

The matrix $I+T$ is a column stochastic matrix and models the Markovian dynamics of our thermalizing channel up to $O(\alpha^2 t^2)$,

$$\langle j|(\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ L}(|i\rangle\langle i|)|j\rangle = \vec{e}_j^{\top}(I+T)^L \vec{e}_i. \tag{3.135}$$

By linearity of $\mathbb{1}+\mathcal{T}_{\text{on}}$ this identity extends to any diagonal density matrix input $\rho = \sum_i p(i)\ |i\rangle\langle i|$.

Proof: We prove this inductively on L. The base case of $L=1$ is trivial from the defintion of T

$$\langle j|(\mathbb{1}+\mathcal{T}_{\text{on}})(|i\rangle\langle i|)|j\rangle = \delta_{i,j} + \langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle = \vec{e}_j^{\top}(I+T)\vec{e}_i. \tag{3.136}$$

For the inductive step we will rely on the fact that there are no off-diagonal elements for diagonal inputs.

$$\langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|k\rangle = \delta_{j,k}\langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle \Longrightarrow \langle j|\mathcal{T}_{\text{on}}^{\circ L}(|i\rangle\langle i|)|k\rangle = \delta_{j,k}\langle j|\mathcal{T}_{\text{on}}^{\circ L}(|i\rangle\langle i|)|j\rangle. \tag{3.137}$$

This is again by induction where the case $L=1$ is proved in Theorem and the inductive step is

$$\begin{aligned}
\langle j|\mathcal{T}_{\text{on}}^{\circ L}(|i\rangle\langle i|)|k\rangle &= \langle j|\mathcal{T}_{\text{on}}\big(\mathcal{T}_{\text{on}}^{\circ L-1}(|i\rangle\langle i|)\big)|k\rangle \\
&= \sum_{m,n}\langle j|\mathcal{T}_{\text{on}}\big(|m\rangle\langle m|\ \mathcal{T}_{\text{on}}^{\circ L-1}(|i\rangle\langle i|)\ |n\rangle\langle n|\big)|k\rangle \\
&= \sum_{m,n}\delta_{m,n}\langle m|\mathcal{T}_{\text{on}}^{\circ L-1}(|i\rangle\langle i|)|m\rangle\langle j|\mathcal{T}_{\text{on}}(|m\rangle\langle m|)|k\rangle \\
&= \sum_{m}\langle m|\mathcal{T}_{\text{on}}^{\circ L-1}(|i\rangle\langle i|)|m\rangle\delta_{j,k}\langle j|\mathcal{T}_{\text{on}}(|m\rangle\langle m|)|j\rangle \\
&= \delta_{j,k}\langle j|\mathcal{T}_{\text{on}}^{\circ L}(|i\rangle\langle i|)|j\rangle.
\end{aligned} \tag{3.138}$$

This argument points the way towards how we will prove the inductive step in our stochastic conversion, starting with

$$\begin{aligned}
\langle j|(\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ L}(|i\rangle\langle i|)|j\rangle &= \langle j|\Big((\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ L-1}(|i\rangle\langle i|) + \mathcal{T}_{\text{on}}\circ(\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ L-1}(|i\rangle\langle i|)\Big)|j\rangle \\
&= \vec{e}_j^{\top}(\mathbb{1}+T)^{L-1}\vec{e}_i + \langle j|\mathcal{T}_{\text{on}}\circ(\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ L-1}(|i\rangle\langle i|)|j\rangle.
\end{aligned} \tag{3.139}$$

We can use the inductive hypothesis on the term on the left and we now have to break down the $\mathcal{T}_{\text{on}}$ term.

$$\begin{aligned}\langle j|\mathcal{T}_{\text{on}} \circ (\mathbb{1} + \mathcal{T}_{\text{on}})^{\circ L-1}(|i\rangle\langle i|)|j\rangle &= \sum_{m,n}\langle j|\mathcal{T}_{\text{on}}\Big(|m\rangle\langle m|\ (\mathbb{1} + \mathcal{T}_{\text{on}})^{\circ L-1}(|i\rangle\langle i|)\ |n\rangle\langle n|\Big)|j\rangle \\ &= \sum_{m}\langle j|\mathcal{T}_{\text{on}}(|m\rangle\langle m|)|j\rangle \vec{e}_m^{\top}(I+T)^{L-1}\vec{e}_i \\ &= \sum_{m}\vec{e}_j^{\top}T\vec{e}_m\vec{e}_m^{\top}(I+T)^{L-1}\vec{e}_i \\ &= \vec{e}_j^{\top}T(I+T)^{L-1}\vec{e}_i. \end{aligned} \tag{3.140}$$

Substituting this into Equation (3.139) yields

$$\langle j|(\mathbb{1} + \mathcal{T}_{\text{on}})^{\circ L}(|i\rangle\langle i|)|j\rangle = \vec{e}_j^{\top}(I+T)^{L}\vec{e}_i. \tag{3.141}$$

Our final step in the proof is to show that $I+T$ is column-stochastic. This is straightforward from our definition of T

$$\sum_i \vec{e}_i^{\top}(I+T)\vec{e}_j = 1 + \sum_i \langle i|\mathcal{T}_{\text{on}}(|j\rangle\langle j|)|i\rangle. \tag{3.142}$$

Now we use the fact that $\langle j|\mathcal{T}_{\text{on}}(|j\rangle\langle j|)|j\rangle = -\sum_{i\neq j}\langle i|\mathcal{T}_{\text{on}}(|j\rangle\langle j|)|i\rangle$ from Equation (3.125) to conclude that $I+T$ is column stochastic. □

Since we will be effectively reducing our quantum dynamics to classical dynamics over the eigenbasis for H_S we will need bounds on the convergence of Markov chains. This is a very deep area of research, with many decades of results, so we point interested readers to the comprehensive book by Levin and Peres [62]. As we will be dealing with non-reversible Markov chains we unfortunately cannot use the relatively well-developed theory for reversible Markov chains. Luckily, we will only need the following theorem due to Jerison.

Theorem 3.6 (Jerison's Markov Relaxation Theorem [63]): *Let $M : \mathbb{R}^N \to \mathbb{R}^N$ be an ergodic Markov transition matrix acting on an N dimensional state space with absolute spectral gap $\lambda_\star := 1 - \max_{i>1}|\lambda_i(M)|$, where the eigenvalues of M are ordered $1 = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq -1$. Given this gap, if the number of steps L in the Markov chain satisfies the following bound*

$$L \geq \frac{N}{\lambda_\star}\left(2\log\frac{1}{\lambda_\star} + 4(1+\log 2) + \frac{1}{N}\left(2\log\left(\frac{1}{\varepsilon}\right) - 1\right)\right) =: \frac{N}{\lambda_\star}J, \tag{3.143}$$

where J is the collection of logarithmic and constant terms that we will typically ignore in asymptotic notation, then the resulting state $M^L(x)$ is ε close to the fixed point

$$\forall \vec{x} \text{ s.t. } x_i \geq 0 \text{ and } \sum_i x_i = 1, \quad \left\|\vec{\pi} - M^L\vec{x}\right\|_1 \leq \varepsilon. \tag{3.144}$$

We use $\vec{\pi}$ to denote the unique eigenvector of eigenvalue 1 for M.

Now that we have an idea of how long it takes for our Markov chain to converge to the fixed points we need to show which states are actually fixed points. We demonstrate that for finite β any fixed point must satisfy a summation of detailed-balance terms. This fixed point is unique if the Markov chain is ergodic, which we do not argue in this lemma as an arbitrary thermalization channel Φ may not be ergodic. For the ground state limit of $\beta \to \infty$ we show that the Markov matrix $I + T$ is upper triangular, which is crucial to our analysis of the spectral gap of the Markov chain in later results. We also demonstrate that the ground state is a fixed point in this limit nearly trivially.

Lemma 3.7 (Markov Chain Fixed Points): *Let T be the transition matrix with sum zero columns $\sum_j \vec{e}_j^\top T \vec{e}_i$ for all i, negative diagonal entries $\vec{e}_i^\top T \vec{e}_i \leq 0$, and off-diagonals smaller than 1 $\vec{e}_j^\top T \vec{e}_i \leq 1$ for $j \neq i$, associated with the on-resonance term $\mathcal{T}_{\text{on}}$ of an arbitrary thermalizing channel Φ. A vector $\vec{p}$ is a fixed point of the Markovian dynamics $I + T$ if and only if it is in the kernel of T. This holds for finite β if the following is satisfied for all j*

$$\sum_{i \neq j} \frac{e^{-\beta \lambda_S(i)}}{\mathcal{Z}_S(\beta)} \vec{e}_j^\top T \vec{e}_i - \frac{e^{-\beta \lambda_S(j)}}{\mathcal{Z}_S(\beta)} \vec{e}_i^\top T \vec{e}_j = 0. \tag{3.145}$$

In the $\beta \to \infty$ limit the ground state $\vec{e}_1$ is a fixed point and T is upper triangular

$$\lim_{\beta \to \infty} (I + T) \vec{e}_1 = \vec{e}_1 \quad \text{and} \quad i > j \implies \lim_{\beta \to \infty} \vec{e}_i^\top T \vec{e}_j = 0. \tag{3.146}$$

Proof: To show that the thermal state is the fixed point of the zero knowledge thermalizing channel we need to show that $T\vec{p}_\beta = 0$ and that the Markov chain is ergodic. Ergodicity will be easy to prove so we focus on showing that $T\vec{p}_\beta = 0$. This condition can be expressed as

$$\vec{e}_j^\top T \vec{p}_\beta = \sum_i \frac{e^{-\beta \lambda_S(i)}}{\mathcal{Z}_S(\beta)} \vec{e}_j^\top T \vec{e}_i = 0. \tag{3.147}$$

We can make a quick substitution as we know the diagonal elements must equal the sum of the remainder of the column

$$\vec{e}_i^\top T \vec{e}_i = -\sum_{j \neq i} \vec{e}_j^\top T \vec{e}_i, \tag{3.148}$$

which we can then pull out the $i = j$ term from the sum in Equation (3.147)

$$\vec{e}_j^\top T \vec{p}_\beta = \sum_{i \neq j} \frac{e^{-\beta \lambda_S(i)}}{\mathcal{Z}_S(\beta)} \vec{e}_j^\top T \vec{e}_i - \frac{e^{-\beta \lambda_S(j)}}{\mathcal{Z}_S(\beta)} \sum_{k \neq j} \vec{e}_k^\top T \vec{e}_j, \tag{3.149}$$

which is 0 if and only if $\vec{p}_\beta$ is a fixed point of $I + T$.

We now show the $\beta \to \infty$ case. We can show that T is upper triangular using Theorem 3.3 which gives us the on-resonance transition amplitude. We assume $i < j$, which implies $\Delta_S(i, j) \leq 0$, and get

$$\begin{aligned}\lim_{\beta\to\infty}\vec{e}_j^{\top}T\vec{e}_i &= \lim_{\beta\to\infty}\langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle \\ &= \tilde{\alpha}^2\lim_{\beta\to\infty}\left[\frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\mathbb{I}[|\Delta_S(i,j)+\gamma|\leq\delta_{\min}]\,\text{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right)\right] \\ &= \tilde{\alpha}^2\mathbb{I}[|\Delta_S(i,j)+\gamma|\leq\delta_{\min}]\,\text{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right)\lim_{\beta\to\infty}\frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}} \\ &= 0. \end{aligned} \tag{3.150}$$

This further shows that the ground state is a fixed point, as every other eigenvector must have higher energy and therefore all on-resonance transitions *from* the ground state must be 0

$$\lim_{\beta\to\infty}\vec{e}_1^{\top}T\vec{e}_1 = \lim_{\beta\to\infty}\langle 1|\mathcal{T}_{\text{on}}(|1\rangle\langle 1|)|1\rangle = -\sum_{j>1}\lim_{\beta\to\infty}\langle j|\mathcal{T}_{\text{on}}(|1\rangle\langle 1|)|j\rangle = 0. \tag{3.151}$$

This then shows that the ground state is fixed

$$(I+T)\vec{e}_1 = \vec{e}_1, \tag{3.152}$$

and completes the proof. □

Using the decomposition from Theorem 3.3 and intermediate expressions in its proof we can now show why the off-resonance map $\mathcal{T}_{\text{off}}$ is named "off-resonance"; even in the worst case scenario of choosing a bad value of γ such that all terms in $\mathcal{T}$ end up in $\mathcal{T}_{\text{off}}$ the trace norm of its output is always controllably small via α.

Corollary 3.2.4.1: *The induced trace norm of the off-resonance map $\mathcal{T}_{\text{off}}(\rho)$, for all density matrices ρ such that $[\rho, H_S] = 0$ and $\dim \geq 2$, is upper bounded for all choices of the environment Hamiltonian γ by*

$$\|\mathcal{T}_{\text{off}}(\rho)\|_1 \leq \frac{8\alpha^2}{\delta_{\min}^2}. \tag{3.153}$$

Proof: This result follows from applying bounds on the sinc function from Lemma 2.1 (given in Appendix B.1) to the worst-case scenario off-resonance terms given in Equation (3.124).

$$i\neq j \implies |\langle j|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|j\rangle| \leq \tilde{\alpha}^2\frac{4}{\delta_{\min}^2 t^2}(1+q(0)+q(1)) = \frac{8\alpha^2}{\delta_{\min}^2(\dim+1)}. \tag{3.154}$$

This allows us to bound the off-resonance self-transition term in Theorem 3.3 as

$$|\langle i|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|i\rangle| = \left|-\sum_{j\neq i}\langle j|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|j\rangle\right| \leq (\dim_S - 1)\frac{8\alpha^2}{\delta_{\min}^2(\dim+1)} \leq \frac{4\alpha^2}{\delta_{\min}^2}. \tag{3.155}$$

Now we can use this, along with our no off-diagonal output elements of $\mathcal{T}$, to compute the trace norm of the off-resonance map

$$\begin{aligned}
\left\|\mathcal{T}_{\text{off}}(\rho)\right\|_1 &= \sum_j |\langle j|\mathcal{T}_{\text{off}}(\rho)|j\rangle| \\
&\leq \sum_{i,j} \rho_{i,i} |\langle j|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|j\rangle| \\
&= \sum_i \rho_{i,i} \left(\sum_{j\neq i} |\langle j|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|j\rangle| + |\langle i|\mathcal{T}_{\text{off}}(|i\rangle\langle i|)|i\rangle| \right) \\
&\leq \sum_i \rho_{i,i} \left((\dim_S - 1) \frac{8\alpha^2}{\delta_{\min}^2(\dim + 1)} + \frac{4\alpha^2}{\delta_{\min}^2} \right) \\
&\leq \frac{8\alpha^2}{\delta_{\min}^2}. \qquad (3.156)
\end{aligned}$$

□

The last result in this section that we will need is a bound on the trace norm of the remainder term, which we state in the following theorem.

Theorem 3.1 (Remainder Bound): *Let $R_\Phi(\rho)$ be the remainder term for the second-order Taylor series expansion of the quantum channel Φ acting on an input state ρ about $\alpha = 0$ defined in Equation (3.109) where the Schätten 1-norm of the remainder operator is bounded by*

$$\left\|R_\Phi(\rho;\alpha)\right\|_1 \leq 4\dim(\alpha t)^3. \qquad (3.157)$$

The proof of the remainder bound follows from the triangle inequality and remainder bounds on Taylor series and is given in Appendix B.1.

3.3 Single Qubit and Truncated Harmonic Oscillator

The first system we study is the qubit $\mathcal{H}_S = \mathbb{C}^2$. This system is simple enough that we can explicitly write the dynamics as a 2×2 transition matrix, which makes it easy to compute required simulation times and easy for the reader to follow. Although this system could be viewed as a warmup to the more general systems in Section ref{sec:general_systems}, as the proof techniques are very similar, we remark that this system does have some unique properties.

The biggest difference is that we do not assume any kind of belief distribution of the eigenvalue gap Δ of the system. We only require that a window of width 2σ is known that contains Δ. We can then characterize the runtime in terms of σ and in addition to determining runtime we find it determines an upper bound on the β that can be prepared at low error.

The other unique phenomenon with the single qubit scenario is that the total simulation time needed is *independent* of β. Although this may seem incorrect, as most existing thermal state preparation algorithms tend to scale at least linearly with β, this is in fact a property of the underlying Markov chain. The rate of convergence of the Markov chain is dictated by the spectral gap, which for this system is shown to be $\tilde{\alpha}^2$. The only aspect of the Markov chain that changes with β is what the fixed point is and the Markov Relaxation Theorem 3.6 provides relaxation guarantees regardless of initial or final state.

Theorem 3.2: *Let H_S be an arbitrary single qubit Hamiltonian with eigenvalue gap Δ, ρ any input state that commutes with H_S, and L the number of interactions simulated. Given a window of width 2σ that is promised to contain Δ and satisfies the inequality*

$$\sigma \leq \min\left\{\frac{\varepsilon}{2\beta}, \Delta\sqrt{\frac{\varepsilon}{2}}\right\}, \tag{3.158}$$

then the following parameter choices

$$\begin{aligned}
\alpha &= \frac{1}{t^3(\Delta+\sigma)^2}, \\
t &\in \frac{1}{\sigma}\left[\sqrt{1-\sqrt{1-\frac{2\sigma^2}{\varepsilon\Delta^2}}}, \sqrt{1+\sqrt{1-\frac{2\sigma^2}{\varepsilon\Delta^2}}}\right], \\
\text{and}\quad L &= \left\lceil \frac{10}{\alpha^2 t^2(1-\sigma^2t^2/2)}\left(2\log\left(\frac{5}{\alpha^2t^2\,\mathrm{sinc}^2(|\Delta-\gamma|\,t/2)}\right)\right.\right. \\
&\qquad \left.\left. +4\log(2e) - \frac{1}{2} + \log\left(\frac{2}{\varepsilon}\right)\right)\right\rceil,
\end{aligned} \tag{3.159}$$

are sufficient to guarantee thermalization of the form $\left\|\rho_S(\beta) - \Phi^{\circ L}(\rho)\right\|_1 \in \tilde{O}(\varepsilon)$. In the limit as $\sigma \to 0$, the total simulation time required scales as

$$\lim_{\sigma\to 0} L\cdot t \in \tilde{O}\left(\frac{1}{\Delta\varepsilon^{2.5}}\right). \tag{3.160}$$

Proof: The proof will be structured into three parts. First, we will need a bound on how close the fixed point of the Markov chain is to the thermal state, because the fixed point is exactly the thermal state only when $\gamma = \Delta$ and our window of width σ is sufficiently small given our error budget. Second, once we have these bounds we then need to determine the number of interactions L that will be necessary to reach the fixed point within trace distance ε. Lastly, we use this value of L to bound the accumulative error from the off-resonance mapping $\mathcal{T}_{\text{off}}$ and remainder term R_Φ.

We start by breaking down the trace distance into three components, one for the fixed-point distance from the thermal state, one for the Markov dynamics distance to the fixed-point, and lastly the remainder terms

$$\begin{aligned}
&\left\|\rho_S(\beta;\Delta) - \Phi^{\circ L}(\rho)\right\|_1 \\
&\leq \left\|\rho_S(\beta;\Delta) - \rho_S(\beta;\gamma)\right\|_1 + \left\|\rho_S(\beta;\gamma) - \Phi^{\circ L}(\rho)\right\|_1 \\
&\leq \left\|\rho_S(\beta;\Delta) - \rho_S(\beta;\gamma)\right\|_1 + \left\|\rho_S(\beta;\gamma) - (\mathbb{1} + \mathcal{T}_{\text{on}})^{\circ L}(\rho)\right\|_1 + \left\|(\mathbb{1} + \mathcal{T}_{\text{on}})^{\circ L}(\rho) - \Phi^{\circ L}(\rho)\right\|_1 \\
&\leq \left\|\rho_S(\beta;\Delta) - \rho_S(\beta;\gamma)\right\|_1 + \left\|\rho_S(\beta;\gamma) - (\mathbb{1} + \mathcal{T}_{\text{on}})^{\circ L}(\rho)\right\|_1 + L\Big(\left\|\mathcal{T}_{\text{off}}(\rho)\right\|_1 + \left\|R_\Phi\right\|_1\Big). \qquad (3.161)
\end{aligned}$$

We proceed with the leftmost term first. The trace distance can be computed explicitly for a single qubit state as

$$\begin{aligned}
&\left\|\rho_S(\beta;\gamma) - \rho_S(\beta;\Delta)\right\|_1 \\
&= |\langle 1|\rho_S(\beta;\gamma)|1\rangle - \langle 1|\rho_S(\beta;\Delta)|1\rangle| + |\langle 2|\rho_S(\beta;\gamma)|2\rangle - \langle 2|\rho_S(\beta;\Delta)|2\rangle| \\
&= |\langle 1|\rho_S(\beta;\gamma)|1\rangle - \langle 1|\rho_S(\beta;\Delta)|1\rangle| + |1 - \langle 1|\rho_S(\beta;\gamma)|1\rangle - 1 + \langle 1|\rho_S(\beta;\Delta)|1\rangle| \\
&= 2|\langle 1|\rho_S(\beta;\gamma)|1\rangle - \langle 1|\rho_S(\beta;\Delta)|1\rangle|. \qquad (3.162)
\end{aligned}$$

Now we expand $\langle 1|\rho_S(\beta;\gamma)|1\rangle$ about $\gamma = \Delta$

$$\begin{aligned}
\langle 1|\rho_S(\beta;\gamma)|1\rangle = \frac{1}{1+e^{-\beta\gamma}} &= \frac{1}{1+e^{-\beta\Delta}} + (\gamma-\Delta)\beta\frac{1}{1+e^{-\beta\gamma_\star}}\frac{e^{-\beta\gamma_\star}}{1+e^{-\beta\gamma_\star}} \\
&= \langle 1|\rho_S(\beta;\Delta)|1\rangle + (\gamma-\Delta)\beta\frac{1}{1+e^{-\beta\gamma_\star}}\frac{e^{-\beta\gamma_\star}}{1+e^{-\beta\gamma_\star}}, \qquad (3.163)
\end{aligned}$$

where $\gamma_\star$ denotes the special value of γ that is guaranteed to make the above equation hold by Taylor's Remainder Theorem. Since the rightmost factors can be upper bounded by $\frac{1}{1+e^{-\beta\gamma_\star}}\frac{e^{-\beta\gamma_\star}}{1+e^{-\beta\gamma_\star}} \leq 1$, this can be rearranged and plugged into Equation (3.162) to give the upper bound

$$\|\rho_S(\beta,\gamma) - \rho_S(\beta,\Delta)\|_1 \leq 2\beta|\Delta-\gamma|. \tag{3.164}$$

Since we require this distance to be less than ε, we can upper bound $|\Delta-\gamma| \leq \sigma$ and require

$$\sigma \leq \frac{\varepsilon}{2\beta}. \tag{3.165}$$

Now we move on to the second stage of the proof: computing the number of interactions needed to reach the fixed point of the Markov chain. As the Markov transition matrix is only 2×2 we will compute it explicitly. To do so, we need the matrix elements for T, which can be computed using Theorem 3.3

$$\begin{aligned}
\vec{e}_1^\top T \vec{e}_1 &= \langle 1|\mathcal{T}_{\text{on}}(|1\rangle\langle 1|)|1\rangle = -\tilde{\alpha}^2 \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}} \operatorname{sinc}^2\left(\frac{(-\Delta+\gamma)t}{2}\right) \\
\vec{e}_2^\top T \vec{e}_1 &= \langle 2|\mathcal{T}_{\text{on}}(|1\rangle\langle 1|)|2\rangle = \tilde{\alpha}^2 \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}} \operatorname{sinc}^2\left(\frac{(-\Delta+\gamma)t}{2}\right) \\
\vec{e}_1^\top T \vec{e}_2 &= \langle 1|\mathcal{T}_{\text{on}}(|2\rangle\langle 2|)|1\rangle = \tilde{\alpha}^2 \frac{1}{1+e^{-\beta\gamma}} \operatorname{sinc}^2\left(\frac{(\Delta-\gamma)t}{2}\right) \\
\vec{e}_2^\top T \vec{e}_2 &= \langle 2|\mathcal{T}_{\text{on}}(|2\rangle\langle 2|)|2\rangle = -\tilde{\alpha}^2 \frac{1}{1+e^{-\beta\gamma}} \operatorname{sinc}^2\left(\frac{(\Delta-\gamma)t}{2}\right).
\end{aligned} \tag{3.166}$$

This gives us the total Markov chain matrix as

$$I + T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \tilde{\alpha}^2 \operatorname{sinc}^2\left((\Delta-\gamma)\frac{t}{2}\right) \frac{1}{1+e^{-\beta\gamma}} \begin{pmatrix} -e^{-\beta\gamma} & 1 \\ e^{-\beta\gamma} & -1 \end{pmatrix}, \tag{3.167}$$

where it can be seen that the fixed point is

$$(I+T)\vec{p}_{\beta,\gamma} = \vec{p}_{\beta,\gamma} = \frac{1}{1+e^{-\beta\gamma}}\vec{e}_1 + \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\vec{e}_2. \tag{3.168}$$

To show convergence we will need the spectral gap of Equation (3.167), which is given as $\lambda_\star = \tilde{\alpha}^2 \operatorname{sinc}^2((\Delta-\gamma)\frac{t}{2})$. Plugging this into the Markov Relaxation Theorem 3.6 we can compute a lowe bound on the number of interactions needed

$$L \geq \frac{2J}{\tilde{\alpha}^2 \operatorname{sinc}^2(|\Delta-\gamma|\frac{t}{2})}, \tag{3.169}$$

where J captures subleading logarithmic factors.

ur next goal is to simplify these bounds so that we can propagate them to our final error requirements. We first use Lemma 2.1 to produce a bound on sinc whenever γ is within our window and $|\Delta-\gamma| \leq \sigma$

$$\operatorname{sinc}^2\left(|\Delta-\gamma|\ \frac{t}{2}\right) \geq 1-\frac{\sigma^2 t^2}{2}, \tag{3.170}$$

provided that $t\sigma \leq \sqrt{2}$ to make the bound meaningful. Recalling that the dimension of the system is 4, we can then create a new lower bound for L by plugging this expression for sinc in to Equation (3.169)

$$L \geq \frac{10J}{\alpha^2 t^2(1-\sigma^2 t^2/2)} \tag{3.171}$$

which is larger than our bound in Equation (3.169). If we choose L to be twice this bound we will for sure meet the Markov chain error requirements.

The third stage of the proof utilizes the above bound on L to bound the off-resonance and remainder terms. The magnitiude of the total off-resonance contributions are $L\|\mathcal{T}_{\text{off}}\|_1 \leq 8/\Delta^2$, given by Corollary 3.2.4.1, and the remainder term is $L\|R_\Phi\|_1 \leq 32\sqrt{2/\pi}\alpha t^3$ from Theorem 3.1. Setting $\alpha = 1/(t^3(\Delta+\sigma)^2) \leq 1/(t^3\Delta^2)$ allows us to make the following inequalities

$$\begin{aligned} L\Big(\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1\Big) &\leq \frac{20J}{t^2(1-\sigma^2t^2/2)}\left(\frac{8}{\Delta^2}+32\sqrt{\frac{2}{\pi}}\alpha t^3\right) \\ &\leq \frac{20J}{t^2\Delta^2(1-\sigma^2t^2/2)}\left(8+32\sqrt{\frac{2}{\pi}}\right). \end{aligned} \tag{3.172}$$

The last step is then to show that the above is $\tilde{O}(\varepsilon)$. As J contains only logarithmic factors, it is sufficient to show that there exists a t such that $t^2(1-\sigma^2t^2/2) \leq \varepsilon$. Rearranging this expression reveals a quadratic in t^2that must satisfy the following

$$\Delta^2 t^2\left(1-\frac{\sigma^2t^2}{2}\right)-\frac{1}{\varepsilon} \geq 0. \tag{3.173}$$

The roots of this quadratic are

$$t^2 = \frac{1}{\sigma^2}\left(1 \pm \sqrt{1-\frac{2\sigma^2}{\Delta^2\varepsilon}}\right), \tag{3.174}$$

meaning that if t lies between these two roots then our bound in Equation (3.172) is $\tilde{O}(\varepsilon)$.

The first observation to make about these roots is that we require $\sigma \leq \Delta\sqrt{\varepsilon/2}$ in order to keep the roots real and not become complex. As $\sigma \to 0$ we note that the larger root $\frac{1}{\sigma^2}\left(1+\sqrt{1-\frac{2\sigma^2}{\Delta^2\varepsilon}}\right)$ approaches infinity and the smaller root approaches $1/\Delta^2\varepsilon$. This means that Equation (3.172) has valid solutions provided σ is sufficiently small. This means that we have successfully bounded

all 3 error terms present in the original decomposition Equation (3.161) by $\tilde{O}(\varepsilon)$. We have done so by setting the following parameters

- $\alpha = \frac{1}{(\Delta+\sigma)^2 t^3}$,
- $t \in \frac{1}{\sigma^2}\left[1 - \sqrt{1 - \frac{2\sigma^2}{\Delta^2 \varepsilon}}, 1 + \sqrt{1 - \frac{2\sigma^2}{\Delta^2 \varepsilon}}\right]$,
- and $L \geq \tilde{\Omega}\left(\frac{1}{\alpha^2 t^2 (1 - \sigma^2 t^2/2)}\right)$.

Substituting in derived expressions for these parameters is sufficient to yield the theorem statement. □

3.3.1 Harmonic Oscillator

Now that we have explored the thermalization channel completely for the single qubit case we turn our attention to a more complicated system: a truncated harmonic oscillator. For this scenario we will assume that the oscillator gap, Δ, is known. This is mostly to simplify proofs of ergodicity and should not be an issue in practice, as evidenced by later theorems that show thermalization without eigenvalue knowledge. The reason behind this proof requirement is that by tuning γ to be the spectral gap we can create a ladder“ transition matrix in which states can move one level up or down. The proof of ergodicity relies on this ladder. Once we remove knowledge of Δ if γ has some probability of being close to 2Δ this special ladder structure is destroyed. To avoid this annoyance and focus on the special structure granted by the harmonic oscillator we will assume $\gamma = \Delta$.

This system also represents a transition from the single qubit to more general settings discussed later as the guarantees on total simulation time as a function of β are similar. For the harmonic oscillator we are only able to bound the spectral gap in the ground state limit as $\beta \to \infty$, meaning that the convergence time for finite β has to be characterized in terms of the spectral gap of the Markov chain. For infinite β we are able to compute the spectral gap exactly, as the Markov transition matrix is upper triangular. The following theorem introduces this technique in a straightforward setting before it is used later for more complicated transition matrices.

Theorem 3.3 (Harmonic Oscillator): *Let H_S denote a truncated harmonic oscillator with $\dim_S$ energy levels that are separated by Δ, giving $\lambda_S(k) = k\Delta$ for $1 \leq k \leq \dim_S$, let γ be chosen to match the eigenvalue gap $\gamma = \Delta$ exactly, and let ρ be any input state that commutes with H_S. Setting the following parameters for the thermalizing channel Φ*

$$\alpha = \frac{\varepsilon^{1.5}\tilde{\lambda_\star}(\beta)^{1.5}\Delta}{\dim_S^4}, t = \dim_S\left(\Delta\sqrt{\varepsilon\tilde{\lambda_\star}(\beta)}\right), \text{ and } \; L \in \tilde{O}\left(\frac{\dim_S^2}{\alpha^2 t^2 \tilde{\lambda_\star}(\beta)}\right), \tag{3.175}$$

where $\tilde{\lambda_\star}(\beta)$ is the spectral gap of the scaled transition matrix $T/\tilde{\alpha}^2$, is sufficient for thermaliziation for arbitrary β as

$$\left\|\rho_S(\beta) - \Phi^{\circ L}(\rho)\right\|_1 \in \tilde{O}(\varepsilon). \tag{3.176}$$

This gives the total simulation time required as

$$L \cdot t \in \tilde{O}\left(\frac{\dim_S^9}{\Delta\varepsilon^{2.5}\tilde{\lambda_\star}(\beta)^{2.5}}\right). \tag{3.177}$$

In the limit $\beta \to \infty$ the above settings for α, t, and L are valid for preparing the ground state with the spectral gap of the scaled transition matrix is further given by

$$\lim_{\beta\to\infty} \tilde{\lambda_\star}_\beta = 1. \tag{3.178}$$

Proof: We first show that the thermal state is the unique fixed point for finite β. This will be done by computing the nonzero on-resonance transitions and plugging in to Lemma 3.7. As $\gamma = \Delta$, $\Delta_{S(i,j)} = (i-j)\Delta$, and $\delta_{\min} = \Delta$, we can deduce that the on-resonance transitions will only be nonzero for adjacent states $|i\rangle\langle i|$ and $|i\pm 1\rangle\langle i\pm 1|$. This can be seen explicitly for $i \neq j$ by evaluating the transition elements given by Definition 3.4.

$$\begin{aligned}
\langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle &= \tilde{\alpha}^2\frac{1}{1+e^{-\beta\gamma}}\mathbb{I}[|\Delta_S(i,j)-\gamma| \leq \delta_{\min}]\,\text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right) \\
&\quad + \tilde{\alpha}^2\frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\mathbb{I}[|\Delta_S(i,j)+\gamma| \leq \delta_{\min}]\,\text{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right) \\
&= \tilde{\alpha}^2 q(0)\mathbb{I}[j = i-1]\,\text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right) \\
&\quad + \tilde{\alpha}^2 q(1)\mathbb{I}[j = i+1]\,\text{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right) \\
&= \tilde{\alpha}^2(q(0)\mathbb{I}[j=i-1] + q(1)\mathbb{I}[j=i+1])
\end{aligned} \tag{3.179}$$

We now plug this expression into Equation (3.145) of Lemma 3.7 and use the fact that $\Delta_S(i, i+1) = \Delta$ for the harmonic oscillator

$$\begin{aligned}
&\sum_{i\neq j} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)}\langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\langle i|\mathbb{T}_{\text{on}}(|j\rangle\langle j|)|i\rangle \\
&= \tilde{\alpha}^2\frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\sum_{i\neq j}\big[\mathbb{I}[j=i-1]\big(q(0)e^{-\beta\Delta_S(i,j)} - q(1)\big) + \mathbb{I}[j=i+1]\big(q(1)e^{-\beta\Delta_S(i,j)} - q(0)\big)\big] \\
&= \tilde{\alpha}^2\frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\big(\big(e^{-\beta\Delta}q(0) - q(1)\big) + \big(e^{+\beta\Delta}q(1) - q(1)\big)\big) \\
&= \tilde{\alpha}^2\frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\left(\left(e^{-\beta\Delta}\frac{1}{1+e^{-\beta\gamma}} - \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}}\right) + \left(e^{+\beta\Delta}\frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}} - \frac{1}{1+e^{-\beta\gamma}}\right)\right) \\
&= 0,
\end{aligned} \tag{3.180}$$

where the final equality comes from setting $\gamma = \Delta$. By Lemma 3.7 this is sufficient for $\rho_S(\beta)$ to be a fixed point of $\mathbb{1} + \mathcal{T}_{\text{on}}$.

To show that $\rho_S(\beta)$ is the unique fixed point of the Markov chain it suffices to show that the walk is ergodic. This means that we need to show that the walk can generate transitions between any two sites, or in other words, the hitting time for any two states $i \neq j$ is nonzero. We prove this by induction on $i - j$ first for $i > j$. For $i = j + 1$ we have

$$\langle j|(\mathbb{1} + \mathcal{T}_{\text{on}})(|i\rangle\langle i|)|j\rangle = \langle j|(\mathbb{1} + \mathcal{T}_{\text{on}})(|j+1\rangle\langle j+1|)|j\rangle = \tilde{\alpha}^2 q(0), \tag{3.181}$$

which is nonzero and therefore the base case holds. Assuming $i = j + n$ holds we show that the hitting time for $i = j + n + 1$ is nonzero. Let p denote the probability of transitioning from j to $j + n$ after n applications of $\mathbb{1} + \mathcal{T}_{\text{on}}$. We show that the probability of transitioning to $j + n + 1$ is nonzero in a few steps starting with the reduction

$$\begin{aligned}
&\langle j|(\mathbb{1} + \mathcal{T}_{\text{on}})^{\circ n+1}(|j+n+1\rangle\langle j+n+1|)|j\rangle \\
&= \sum_{k_1,k_2}\langle j|(\mathbb{1} + \mathcal{T}_{\text{on}})^{\circ n} \circ (|k_1\rangle\langle k_1|\ (\mathbb{1} + \mathcal{T}_{\text{on}})(|j+n+1\rangle\langle j+n+1|)\ |k_2\rangle\langle k_2|)|j\rangle.
\end{aligned} \tag{3.182}$$

We then can set $k_1 = k_2 = k$ as we know that $\mathbb{1} + \mathcal{T}_{\text{on}}$ does not add coherences, meaning it maps diagonal operators $(|j+n+1\rangle\langle j+n+1|)$ to diagonal operators $(|k\rangle\langle k|)$. We can then use the fact that one application of $\mathbb{1} + \mathcal{T}_{\text{on}}$ can map $j + n + 1$ to $j + n$ and take only that term out of the sum over k

$$\begin{aligned}
&\sum_k \langle j|(\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ n} \circ (|k\rangle\langle k|\ (\mathbb{1}+\mathcal{T}_{\text{on}})(|j+n+1\rangle\langle j+n+1|)\ |k\rangle\langle k|)|j\rangle \\
&\geq \langle j|(\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ n} \circ (\langle j+n|j+n\rangle(\mathbb{1}+\mathcal{T}_{\text{on}})(|j+n+1\rangle\langle j+n+1|)\ |j+n\rangle\langle j+n|)|j\rangle \\
&= \tilde{\alpha}^2 q(0)\langle j|(\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ n}(|j+n\rangle\langle j+n|)|j\rangle \\
&= \tilde{\alpha}^2 q(0)p, \qquad (3.183)
\end{aligned}$$

which is clearly greater than 0. To prove the case where $i < j$ the same inductive argument above can be repeated but this time factors of $q(1)$ accumulate as opposed to $q(0)$.

ow that we have shown that the thermal state is the fixed point we would like to bound the total simulation time needed. To do so we first decompose our error into two parts, a Markov chain error and an off-resonance and remainder error

$$\left\|\rho_S(\beta) - \Phi^{\circ L}(\rho)\right\|_1 \leq \left\|\rho_S(\beta) - (\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ L}(\rho)\right\|_1 + L\left(\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1\right) \qquad (3.184)$$

We first bound the number of interactions, L, needed for the output of the Markov chain to be ε close to the fixed point and then use this bound on L to upper bound the off-resonance and remainder error. Unfortunately in the finite β scenario we are unable to determine the spectral gap of T, the entries of which are given in Eq. eqref{eq:harmonic_oscillator_t_matrix}. The spectral gap of T is necessary to use Jerison's Markov Relaxation Theorem 3.6 which poses a problem for our understanding of the evolution time needed. Instead, we will pull out the overall factor of $\tilde{\alpha}^2$ and let $\tilde{\lambda_\star}(\beta)$ denote the spectral gap of $T/\tilde{\alpha}^2$. This then allows us to use Theorem 3.6 but we will have to leave the number of interactions required in terms of $\tilde{\lambda_\star}(\beta)$.

Theorem 3.6 tells us that requiring

$$L \geq \frac{\dim_S}{\tilde{\alpha}^2\tilde{\lambda_\star}(\beta)} J \in \tilde{O}\left(\frac{\dim_S^2}{\alpha^2 t^2 \tilde{\lambda_\star}(\beta)}\right) \qquad (3.185)$$

is sufficient for the total variational distance between the stationary distribution to be ε-small, in other words $\left\|\rho_S(\beta) - (\mathbb{1}+\mathcal{T}_{\text{on}})^{\circ L}(\rho)\right\|_1 \in \tilde{O}(\varepsilon)$. Now we use this expression for L to bound the off-resonance and remainder errors. To do so we first want to asymptotically bound the two contributions, which can be found in Corollary 3.2.4.1 and Theorem 3.1. The sum of the two errors is given by

$$\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1 \leq \frac{8\alpha^2}{\Delta^2} + 16\sqrt{\frac{\pi}{2}}\dim_S (\alpha t)^3. \qquad (3.186)$$

by setting $\alpha = 1/(\dim_S \Delta^2 t^3)$ we can simplify the above as

$$\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1 \leq \frac{\alpha^2}{\Delta^2}\left(8 + 16\sqrt{\frac{\pi}{2}}\right). \tag{3.187}$$

Using the sub-additivity property of the trace distance the total error scales as

$$\begin{aligned} L\left(\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1\right) &\leq \frac{\dim_S^2 \alpha^2 J}{\alpha^2 t^2 \tilde{\lambda}_\star(\beta)\Delta^2}\left(8 + 16\sqrt{\frac{\pi}{2}}\right) \\ &\in \tilde{O}\left(\frac{\dim_S^2}{t^2\Delta^2\tilde{\lambda}_\star(\beta)}\right). \end{aligned} \tag{3.188}$$

We can make this $\tilde{O}(\varepsilon)$ by setting $t = \frac{\dim_S}{\Delta\sqrt{\varepsilon\tilde{\lambda}_\star(\beta)}}$. This then gives the total simulation time as

$$L \cdot t \in \tilde{O}\left(\frac{\dim_S^2}{\alpha^2 t\tilde{\lambda}_\star(\beta)}\right) = \tilde{O}\left(\frac{\dim_S^9}{\varepsilon^{2.5}\tilde{\lambda}_\star(\beta)^{3.5}}\right). \tag{3.189}$$

Now that we have analyzed the finite β regime, we turn to the $\beta \to \infty$ limit. Our proof above fo the fixed points only works for finite β, but Lemma 3.7 tells us that in the $\beta \to \infty$ limit the ground state is a fixed point. We will show it is the unique fixed point by directly computing the spectrum of T, which will be rather easy to do. Lemma 3.7 further tells us that as $\beta \to \infty$ the matrix T is upper triangular, which means we can compute the spectrum simply by just computing the diagonal elements. We do so via Equation (3.179) and Equation (3.119), which says that for $1 < i < \dim_S$ we have

$$\begin{aligned} \vec{e}_i^\top T \vec{e}_i &= \langle i|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|i\rangle \\ &= -\sum_{j\neq i}\langle j|\mathcal{T}_{\text{on}}(|i\rangle\langle i|)|j\rangle \\ &= -\tilde{\alpha}^2\sum_{j\neq i}(q(0)\mathbb{I}[j = i-1] + q(1)\mathbb{I}[j = i+1]) \\ &= -\tilde{\alpha}^2(q(0) + q(1)) \\ &= -\tilde{\alpha}^2, \end{aligned} \tag{3.190}$$

where the summation is only nonzero for $j = i \pm 1$. For $i = 1$ we note that because $\vec{e}_1$ is a fixed point we have $\vec{e}_j^\top T\vec{e}_1 = 0$, so the diagonal entry is 0. The computation for $i = \dim_S$ is similar to the above but yields from Equation (3.119)

$$\lim_{\beta\to\infty}\langle \dim_S|\mathcal{T}_{\text{on}}(|\dim_S\rangle\langle\dim_S|)|\dim_S\rangle = -\tilde{\alpha}^2\lim_{\beta\to\infty}q(0) = -\tilde{\alpha}^2. \tag{3.191}$$

This shows us that the zero temperature limit of the transition matrix T is

$$\lim_{\beta\to\infty} T = \tilde{\alpha}^2 \begin{pmatrix} 0 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & & \\ & & & \ddots & \\ & & & & 1 \\ & & & & -1 \end{pmatrix}. \tag{3.192}$$

We can compute the spectrum via the characteristic polynomial $\det(\lambda I - T)$. This is because T is upper triangular and the determinant we need to compute is

$$\det\left(\lambda I - \lim_{\beta\to\infty} T\right) = \begin{vmatrix} \lambda & -\tilde{\alpha}^2 & & & \\ & \lambda+\tilde{\alpha}^2 & -\tilde{\alpha}^2 & & \\ & & \lambda+\tilde{\alpha}^2 & & \\ & & & \ddots & \\ & & & & -\tilde{\alpha}^2 \\ & & & & \lambda+\tilde{\alpha}^2 \end{vmatrix}. \tag{3.193}$$

The roots of the above characteristic polynomial gives the spectrum of T as 0 and $-\tilde{\alpha}^2$ with multiplicity $\dim_S - 1$. This not only gives the spectral gap of $\tilde{\alpha}^2$ but further shows that the ground state is the unique fixed point because 0 only has multiplicity 1. This shows that $\lim_{\beta\to\infty} \tilde{\lambda}_\star(\beta) = 1$.

We now can use this to repeat the simulation time bound arguments from the finite β case. The decomposition in Equation (3.184) is still valid and we can use the Markov Relaxation Theorem 3.6 to bound

$$L \geq \frac{\dim_S}{\tilde{\alpha}^2 \lim_{\beta\to\infty} \tilde{\lambda}_\star(\beta)} J \in \tilde{\Theta}\left(\frac{\dim_S^2}{\alpha^2 t^2}\right). \tag{3.194}$$

The arguments for the off-resonance and remainder error bounds are the exact same and tell us that it suffices to set

$$\alpha = \frac{1}{\dim_S \Delta^2 t^3} \quad \text{and} \quad t = \frac{\dim_S}{\Delta\sqrt{\varepsilon}}. \tag{3.195}$$

This gives the total simulation time needed as

$$L \cdot t \in \tilde{O}\left(\frac{\dim_S^9}{\varepsilon^{2.5}\Delta}\right). \tag{3.196}$$

□

3.3.2 Numerics

Now that we have rigorous bounds on each of the parameters α, t and L needed to prepare thermal states of simple systems, we turn to numerics to test these bounds. The first question we explore is how the total simulation time $L \cdot t$ behaves as a function of α and t. After, we examine the dependence of the total simulation time on the inverse temperature β and we observe a Mpemba-like effect where we find higher temperature states can cool faster than lower temperature ones [64]. Finally, we demonstrate how our proof techniques could be leading to worse ε scaling than appears numerically necessary. Throughout these experiments we have the same numeric method of starting with the maximally mixed state $\rho_S(0)$ and performing a search on the minimal number of interactions needed for the mean trace distance over all samples to be less than the target ε. The number of samples is increased until the variance in the trace distance is less than an order of magnitude below the mean.

In Figure 7 we explore the total simulation time needed to prepare a thermal state with $\beta = 2.0$ and $\varepsilon = 0.05$ for a single qubit system. We plot the total simulation time $L \cdot t$ needed as a function of t for various settings of α. We find that increasing both parameters tends to decrease the overall cost until a saturation point is reached, which is at a value of t slightly larger $\frac{1}{\alpha}$. For a fixed value of α this initial decrease in $L \cdot t$ is inverse with t, in agreement with our finding of $L \in \tilde{O}(t^{-2})$ in Eq. eqref{eq:single_qubit_l_bound_2} for $\sigma = 0$. However, this process of decreasing the cost by increasing t can only scale so far and appears to run into a minimum number of interactions L required to thermalize. After this saturation point $L \cdot t$ scales linearly with t, indicating that the number of interactions L has reached a minimum.

Another major take away from Figure 7 is that it demonstrates that our thermalizing channel is exceptionally robust beyond the weak-coupling expansion in which we can theoretically analyze it. The values of αt used in the far right of the plot completely break our weak-coupling expansion, as we have values of $\tilde{\alpha}$ that reach up to 500. One interesting phenomenon that we do not have an explanation for is the clumping“ of various settings of α in the large t limit. As αt dictates the amount of time that the random interaction term G is simulated for, it could be that once a minimum amount of randomness is added via this interaction it is no longer beneficial in causing transitions among system eigenstates.

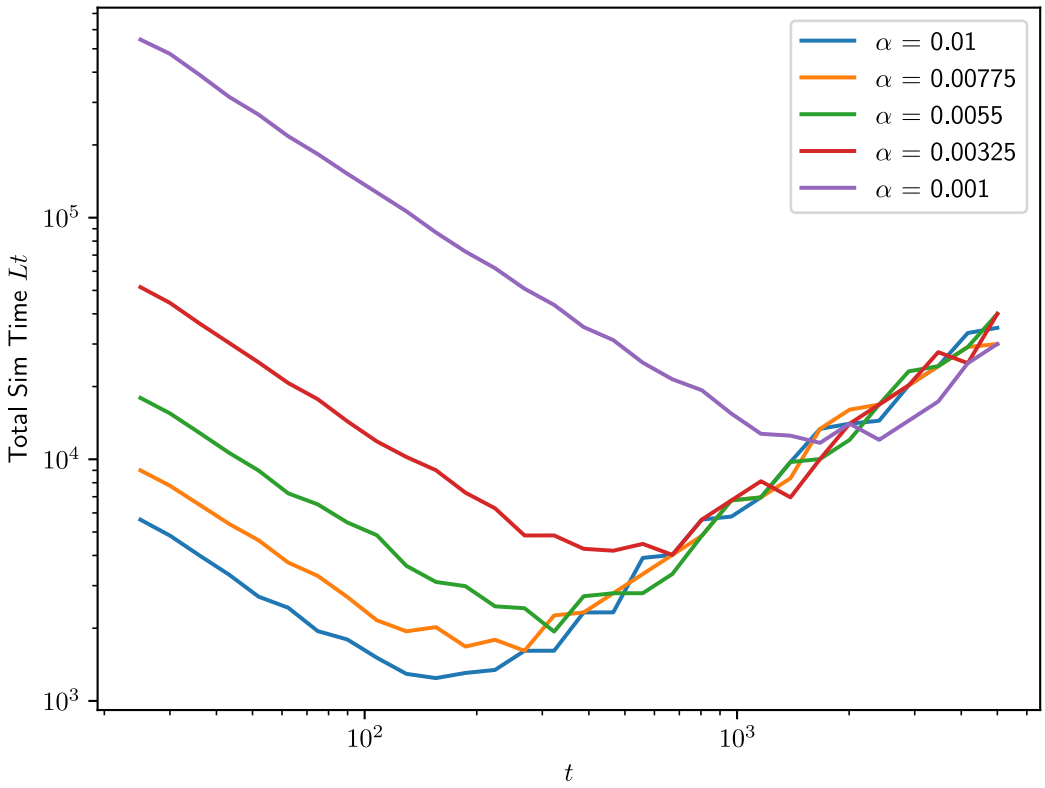
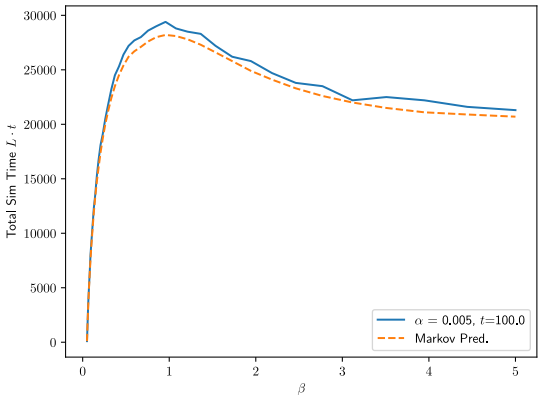


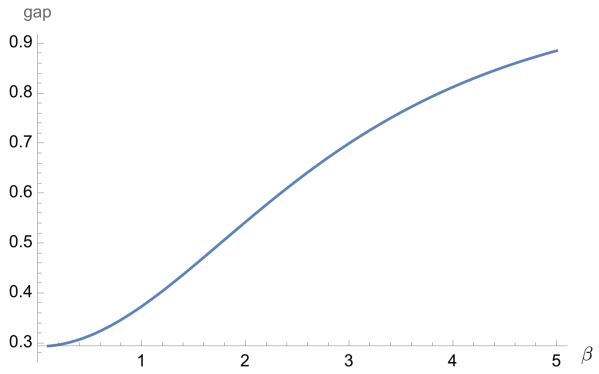
Figure 7: Total simulation time for a single qubit system to reach within trace distance of 0.05 of the thermal state for $\beta = 2$ as a function of per-interaction simulation time t. The slope of the large t asymptote is ≈ 1.01.

The next task we have is to examine the β dependence. For the harmonic oscillator Theorem 3.3 is helpful for giving an idea of the total simulation time for the ground state but we cannot extend it to finite β due to the special structure of the transition matrix in the $\beta \to \infty$ limit. Perturbation theory could possibly be used to extend the computation of the spectral gap to the low temperature regime, but even then it would break down for large temperature (small β). For generic β the structure of the harmonic oscillator transition matrix is tridiagonal but it is not quite Toeplitz, as the main diagonals deviate in the upper left and bottom right corners. We could try to pull these deviations into a separate matrix and treat them as perturbations to a fully Toeplitz matrix, which we can then compute the spectrum of. The issue with this approach is that these deviations are on the order of $\tilde{\alpha}^2 q(0)$ and $\tilde{\alpha}^2 q(1)$, which are comparable to the eigenvalues of the unperturbed matrix.

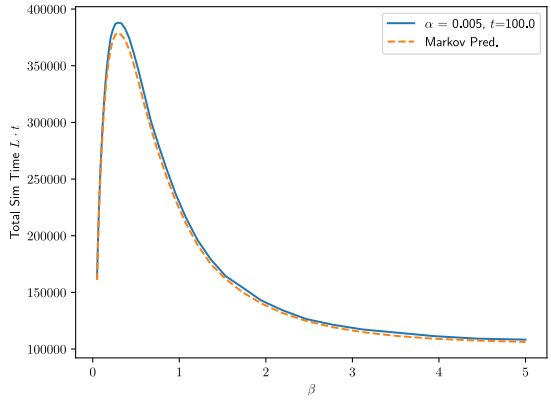
In Figure 8 we are able to probe the total simulation time and spectral gap of the harmonic oscillator as a function of β. We reveal a rather surprising Mpemba-like phenomenon where it takes longer for an infinite temperature initial state (the maximally mixed state) to cool to intermediate temperatures than low temperature states. The Mpemba effect [65] is a classical phenomenon related to the time needed to freeze hot water compared to room temperature water with mentions going all the way back to Aristotle. This phenomenon has been extended to quantum thermodynamics and observed in both theory [66,67] and in recent experimental research [68]. Our observations are not only a further analytic observation, but we are able to provide a proposed mechanism that explains the behavior. It is clear that the distance of our initial state to the target thermal state $\|\rho_S(\beta) - \rho_S(\infty)\|_1$ increases monotonically with β but what is not obvious is that the spectral gap of the underlying Markov chain is emph{also} increasing. As larger spectral gaps lead to quicker convergences this acts in an opposite way on the total simulation time. The end result is that for small β the increase in initial distance is stronger than the increase in the spectral gap and $L \cdot t$ increases. After some amount of β these forces flip and the spectral gap effects become stronger than the initial state distance increasing, leading to a reduction in $L \cdot t$. This phenomenon appears to become more pronounced as the dimension of the harmonic oscillator increases, as can be seem in the $\dim_S = 10$ data. Two things remain unclear: the first is what parameters affect the position and height of the peak in total simulation time and the second is if this behavior is present in Hamiltonians with more complicated eigenvalue difference structure than the harmonic oscillator.



(a) dim = 4 truncated harmonic oscillator



(b) dim = 4 spectral gap vs β



(c) dim = 10 truncated harmonic oscillator

Figure 8: Demonstration of β dependence of the thermalizing channel Φ for the truncated harmonic oscillator. The environment gap γ was tuned to match the system gap Δ exactly. The minimal number of interactions was found by binary search over values of L that have an average error of less than $\varepsilon = 0.05$ with 100 samples.

The analytic proofs given in Theorem 3.2 and Theorem 3.3 are entirely based on our weak-coupling expansion derived in Section 3.2. The high level picture of this expansion is that we have a remainder error that scales like $\tilde{O}(\alpha t)^3)$ and an off-resonance error that scales as $O(\alpha^2)$. To balance these two terms we then set $\alpha = O(\frac{1}{t^3})$. However, as seen in Figure 7 our thermalization routine appears to be quite robust beyond this weak-coupling expansion, which could lead to significant improvements in runtime. In our derivation for the $O(\alpha)$ and $O(\alpha^2)$ terms we relied on our eigenvalues being I.I.D Gaussian variables, with the first and second order expressions containing factors with the first and second moments respectively of the Gaussian distribution. This would suggest that the third order term in a weak coupling expansion might also be 0,

similarly to the first order term. This would lead to a supposed remainder error of $O(\alpha^4 t^4)$, which after balancing with the off-resonance error would give $\alpha = O(1/t^2)$. If the number of interactions then scales like $O(1/(\alpha^2 t^2))$, which is consistent with the spectral gap of $\mathcal{T}_{\text{on}}$ scaling as $O(\alpha^2 t^2)$, then to make the total error of order $O(\varepsilon)$ we would require $t \in \tilde{O}(1/\varepsilon^{0.5})$ as in Theorem 3.2 and Theorem 3.3. This conjecture then leads to a total simulation time of order $O(1/\varepsilon^{1.5})$.

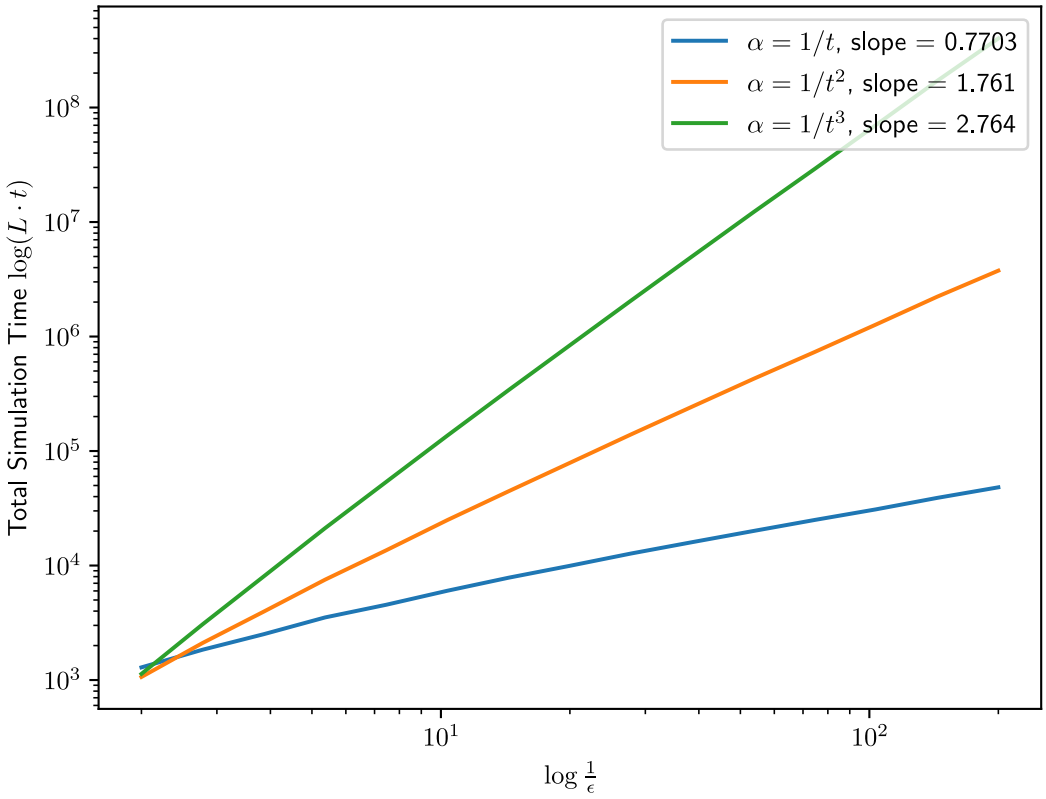


Figure 9: Scaling of $L \cdot t$ to prepare a harmonic oscillator thermal state with $\beta = \dim_S = 4$ with respect to $1/\varepsilon$ in a log-log plot. For each line in the plot we scaled α by a constant value to make $\tilde{\alpha}^2 \approx 0.05$ for the largest value of ε. Each of these slopes was obtained via a least squares fitting of a power-law to $L \cdot t$ and $1/\varepsilon$ and are consistently larger by 0.25-0.27 compared to stated predictions.

An even further conjecture would be to keep $\alpha \cdot t$ as a small constant, in this case we are essentially saying that the randomized dynamics $e^{i\alpha t G}$ are beneficial and should not be thought of as some remainder error to be minimized. If the αt constant is small enough then the dynamics will still be approximated by the Markov chain $\mathcal{T}_{\text{on}}$. Our spectral gap will still scale as $O((\alpha t)^2)$

and t as $O(1/\varepsilon^{0.5})$. This would lead to our total simulation time scaling as $O(1/\varepsilon^{0.5})$. In Figure 9 we numerically explore these various scalings of α for the harmonic oscillator with $\beta = \dim_S = 4$. Our first remark is that the $\alpha = O(1/t^3)$ scaling as dictated by Theorem 3.3 is numerically supported. Specifically, the theorem suggests that we should observe $O(1/\varepsilon^{2.5})$ scaling for $L \cdot t$. Our experiment suggesting $L \cdot t \in O(1/\varepsilon^{2.764})$ which is approximately consistent and deviations from this scaling may arise from the inclusion of data in the fit from outside of the weak coupling limit which is the only regime where we anticipate this scaling.

3.4 Generic Systems

We now extend our thermalization techniques to arbitrary Hamiltonians with no degenerate eigenvalues. The first major difficulty that we run into is how to choose our environment gap γ. If one does not have any knowledge whatsoever about where the eigenvalues of H_S may lie then we are reduced to uniform guessing. In Section ref{sec:zero_knowledge} we show that even in this scenario the thermal state is an approximate fixed point for finite β and the exact fixed point for ground states and we provide a bound on the total simulation time required. However, we show that this generality does come at a cost. If one has complete knowledge of the eigenvalue differences we show in Section ref{sec:perfect_knowledge} that the total simulation time markedly decreases. Further, with complete knowledge the thermal state is an exact fixed point for all β. Finally, in Section ref{sec:general_numerics} we study these impacts on small Hydrogen chain systems and observe the quantitative effects of noise added to γ.

The assumption on non-degenerate eigenvalues is required for fairly technical conditions. In the $\beta \to \infty$ limit for our proof of the spectral gap we use the fact that the transition matrix T is upper triangular in Lemma 3.7. This is where the non-degeneracy is required because degenerate eigenvalues always have a non-zero transition amplitude that scales as $\tilde{O}(\alpha^2)$ without a factor of sinc. This means within the degenerate subspace in the transition matrix there is a uniform block. This makes computing the transition matrix spectrum a little more complicated than necessary, so we avoid this issue by requiring no degeneracies. This restriction could likely be lifted through an intelligent choice of eigenbasis for the degenerate subspace, or through better spectrum calculations of the resulting transition matrix, but we leave such explorations for future work.

3.4.1 Zero Knowledge

We now move on to show how our channel performs if one has no knowledge about the eigenvalue differences $\Delta_{S(i,j)}$ apart from a bound on the maximum value of these differences. This is represented by choosing γ uniformly from the interval $[0, 4\|H_S\|]$, which technically constitutes an upper bound on the largest $\Delta_S(i,j)$, but estimates of $\|H_S\|$ are often readily attainable from the specification of the Hamiltonian using the triangle inequality. We also assume that an input state that commutes with the Hamiltonian can be provided, the maximally mixed state is sufficient as would a random eigenstate yielded by the quantum phase estimation algorithm.

Theorem 3.4 (Zero Knowledge Thermal State Prep): *Let H_S be a Hermitian matrix of dimension $\dim_S$ with no degenerate eigenvalues, ρ any input state that commutes with H_S, and γ a random variable distributed uniformly in the interval $[0, 4\|H_S\|]$ and let ρ_{fix} denote the unique fixed point of the transition dynamics $\mathbb{1} + \mathbb{E}_\gamma \mathcal{T}_{\text{on}}^{(\gamma)}$ where $\mathcal{T}_{\text{on}}^{(\gamma)}$ is the on-resonance transition matrix used above with the dependence on γ made explicit. The following statements then hold.*

1. *For finite β the thermal state is an approximate fixed point of the thermalizing channel $\mathbb{E}_\gamma \Phi_\gamma$ with a deviation of*

$$\left\|\rho_S(\beta) - \mathbb{E}_\gamma \Phi_\gamma(\rho_S(\beta))\right\|_1 \leq \alpha^2 t e^{\beta \delta_{\min}} \|H_S\|^{-1} \pi + 8\frac{\alpha^2}{\delta_{\min}} + 16\sqrt{\frac{\pi}{2}} \dim_S (\alpha t)^3 \tag{3.197}$$

2. *The parameter settings*

$$\alpha = \frac{\delta_{\min}^4 \varepsilon^3 \tilde{\lambda}_\star(\beta)^3}{\dim_S^7 \|H_S\|^3}, \quad t = \frac{\dim_S^2 \|H_S\|}{\varepsilon \tilde{\lambda}_\star(\beta) \delta_{\min}^2}, \quad \text{and} \quad L \in \tilde{O}\left(\frac{\dim_S^{14} \|H_S\|^6}{\varepsilon^5 \delta_{\min}^6 \tilde{\lambda}_\star(\beta)^6}\right) \tag{3.198}$$

are sufficient for any $\beta \in [0, \infty]$ and error tolerance $\varepsilon \in (0, 2]$ to guarantee $\left\|\rho_{\text{fix}} - \left(\mathbb{E}_\gamma \Phi_\gamma\right)^{\circ L}(\rho)\right\|_1 \in \tilde{O}(\varepsilon)$. The total simulation time needed is therefore

$$L \cdot t \in \tilde{O}\left(\frac{\dim_S^{16} \|H_S\|^7}{\delta_{\min}^8 \varepsilon^6 \tilde{\lambda}_\star(\beta)^7}\right). \tag{3.199}$$

3. *The fixed point is the ground state In the $\beta \to \infty$ limit and the spectral gap, $\tilde{\lambda}_\star(\beta)$, of the rescaled transition matrix $\mathbb{E}_\gamma T_\gamma \cdot \left(\frac{2\|H_S\|(\dim + 1)}{\alpha^2 t}\right)$ is lower bounded by a constant, giving the two limits*

$$\lim_{\beta \to \infty} \rho_{\text{fix}} = |1\rangle\langle 1| \quad \text{and} \quad \lim_{\beta \to \infty} \tilde{\lambda}_\star(\beta) = 2 \int_0^{-\delta_{\min} \frac{t}{2}} \text{sinc}^2(u) du \geq 2.43. \tag{3.200}$$

Proof: We start by understanding the fixed points of $\mathbb{1} + \mathbb{E}_\gamma \mathcal{T}_{\text{on}}^{(\gamma)}$, conditions for the thermal state being fixed are given in Lemma 3.7. As the condition boils down to a detailed balance like condition, we need to compute the off-diagonal transition elements first. Starting with $i > j$ we have from Definition 3.4 that

$$
\begin{aligned}
&\mathbb{E}_\gamma \langle j|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|j\rangle \\
&= \tilde{\alpha}^2 \mathbb{E}_\gamma \frac{1}{1+e^{-\beta\gamma}} \mathbb{I}[|\Delta_S(i,j)-\gamma| \leq \delta_{\min}] \text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right) \\
&+\tilde{\alpha}^2 \mathbb{E}_\gamma \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}} \mathbb{I}[|\Delta_S(i,j)+\gamma| \leq \delta_{\min}] \text{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right) \\
&= \tilde{\alpha}^2 \mathbb{E}_\gamma \frac{1}{1+e^{-\beta\gamma}} \mathbb{I}[|\Delta_S(i,j)-\gamma| \leq \delta_{\min}] \text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right) \\
&= \tilde{\alpha}^2 \frac{1}{4\|H_S\|} \int_0^{4\|H_S\|} \frac{1}{1+e^{-\beta\gamma}} \mathbb{I}[|\Delta_S(i,j)-\gamma| \leq \delta_{\min}] \text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right) d\gamma \\
&= \frac{\tilde{\alpha}^2}{4\|H_S\|} \int_{\Delta_S(i,j)-\delta_{\min}}^{\Delta_S(i,j)+\delta_{\min}} \frac{1}{1+e^{-\beta\gamma}} \text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right) d\gamma \\
&= \frac{\tilde{\alpha}^2}{2t\|H_S\|} \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \frac{1}{1+e^{-\beta(\Delta_S(i,j)-2\frac{u}{t})}} \text{sinc}^2(u) du
\end{aligned} \tag{3.201}
$$

The exact same calculation holds for $i < j$, which after repeating the steps that led to Equation (3.201) we arrive at a similar result with a slightly different integrand

$$
\mathbb{E}_\gamma \langle j|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|j\rangle = \frac{\tilde{\alpha}^2}{2t\|H_S\|} \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \frac{e^{-\beta(\Delta_S(j,i)-2\frac{u}{t})}}{1+e^{-\beta(\Delta_S(j,i)-2\frac{u}{t})}} \text{sinc}^2(u) du, \tag{3.202}
$$

as we pick up a factor of $q(1)$ as opposed to $q(0)$. Note that we have also shown that $\mathbb{E}_\gamma T_\gamma$ is ergodic, as there is a nonzero probability for any state $|i\rangle\langle i|$ to transition to any other state $|j\rangle\langle j|$ in one iteration on average over γ.

For finite β the condition for $\rho_S(\beta)$ being a fixed point is given in Equation (3.145), repeated here as

$$
\sum_{j \neq i} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{e}_i - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \vec{e}_i^\top \mathbb{E}_\gamma T_\gamma \vec{e}_j = 0, \tag{3.203}
$$

for all j. We can plug in our calculation for the transition coefficients for summands with $i > j$ first

$$\begin{aligned}
&\frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)}\vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{e}_i - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\vec{e}_i^T \mathbb{E}_\gamma T_\gamma \vec{e}_j \\
&= \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\Big(e^{-\beta\Delta_S(i,j)}\mathbb{E}_\gamma \langle j|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|j\rangle - \langle i|\mathcal{T}_{\text{on}}^{(\gamma)}(|j\rangle\langle j|)|i\rangle\Big) \\
&= \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\frac{\tilde{\alpha}^2}{2t\|H_S\|}e^{-\beta\Delta_S(i,j)}\int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \frac{1-e^{\beta 2\frac{u}{t}}}{1+e^{-\beta(\Delta_S(i,j)-2\frac{u}{t})}}\text{sinc}^2(u)du \\
&= \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)}\frac{\tilde{\alpha}^2}{2t\|H_S\|}\int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \frac{1-e^{\beta 2\frac{u}{t}}}{1+e^{-\beta(\Delta_S(i,j)-2\frac{u}{t})}}\text{sinc}^2(u)du. \qquad (3.204)
\end{aligned}$$

For $i < j$ we have the very similar

$$\begin{aligned}
&\frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)}\vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{e}_i - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\vec{e}_i^\top \mathbb{E}_\gamma T_\gamma \vec{e}_j \\
&= \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\frac{\tilde{\alpha}^2}{2t\|H_S\|}\int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \frac{e^{\beta 2\frac{u}{t}}-1}{1+e^{-\beta(\Delta_S(j,i)-2\frac{u}{t})}}\text{sinc}^2(u)du. \qquad (3.205)
\end{aligned}$$

Unfortunately these integrals are not 0, which can be verified numerically, and it is unclear how to make the summation over $in = j$ equal to 0.

Our work around this is that instead of showing that the thermal state is exactly the fixed point we can use these results to show that it is an approximate fixed point. There are a few ways we could proceed. The first way could be to compute a Taylor series for the integrand and isolate the limits in which the remainder goes to 0. Unfortunately due to the $\text{sinc}^2(u) = \frac{\sin(u)^2}{u^2}$ term this means that the overall scaling will go like $\frac{1}{t}$, making the total expression independent of t. Instead the route we will take will be to upper bound the norm $\left\|\vec{p}_\beta - \mathbb{E}_\gamma(I + T_\gamma)\vec{p}_\beta\right\|_1 = \left\|\mathbb{E}_\gamma T_\gamma \vec{p}_\beta\right\|_1$, as this norm is only 0 if and only if $\vec{p}_\beta$ is a fixed point. We reduce this to computations we have already performed as

$$\begin{aligned}
\left\|\mathbb{E}_\gamma T_\gamma \vec{p}_\beta\right\|_1 &= \sum_j \left|\vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{p}_\beta\right| \\
&= \sum_j \left|\sum_i \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)}\vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{e}_i\right| \\
&= \sum_j \left|\sum_{i\neq j} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)}\vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{e}_i - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)}\vec{e}_i^\top \mathbb{E}_\gamma T_\gamma \vec{e}_j\right|. \qquad (3.206)
\end{aligned}$$

This is essentially the derivation for the fixed point conditions derived in Lemma 3.7. We now plug in Equation (3.204) and Equation (3.205) into the above and upper bound the integral as

$$\sum_j \left| \sum_{i \neq j} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{e}_i - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \vec{e}_i^\top \mathbb{E}_\gamma T_\gamma \vec{e}_j \right|$$

$$\leq \sum_j \left| \sum_{i<j} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{e}_i - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \vec{e}_i^\top \mathbb{E}_\gamma T_\gamma \vec{e}_j \right| + \sum_j \left| \sum_{i>j} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \vec{e}_j^\top \mathbb{E}_\gamma T_\gamma \vec{e}_i - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \vec{e}_i^\top \mathbb{E}_\gamma T_\gamma \vec{e}_j \right|$$

$$= \frac{\tilde{\alpha}^2}{2t\|H_S\|} \sum_j \left| \sum_{i<j} \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \frac{e^{\beta 2\frac{u}{t}} - 1}{1 + e^{-\beta(\Delta_S(j,i) - 2\frac{u}{t})}} \mathrm{sinc}^2(u) du \right|$$

$$+ \frac{\tilde{\alpha}^2}{2t\|H_S\|} \sum_j \left| \sum_{i>j} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \frac{1 - e^{\beta 2\frac{u}{t}}}{1 + e^{-\beta(\Delta_S(i,j) - 2\frac{u}{t})}} \mathrm{sinc}^2(u) du \right|$$

$$\leq \frac{\tilde{\alpha}^2}{2t\|H_S\|} \sum_j \sum_{i<j} \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \left| \frac{e^{\beta 2\frac{u}{t}} - 1}{1 + e^{-\beta(\Delta_S(j,i) - 2\frac{u}{t})}} \right| \mathrm{sinc}^2(u) du$$

$$+ \frac{\tilde{\alpha}^2}{2t\|H_S\|} \sum_j \sum_{i>j} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \left| \frac{1 - e^{\beta 2\frac{u}{t}}}{1 + e^{-\beta(\Delta_S(i,j) - 2\frac{u}{t})}} \right| \mathrm{sinc}^2(u) du$$

$$\leq \frac{\tilde{\alpha}^2}{2t\|H_S\|} e^{\beta\delta_{\min}} \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \mathrm{sinc}^2(u) du \left(\sum_j \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \sum_{i<j} 1 + \sum_j \sum_{i>j} \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \right)$$

$$\leq \frac{\tilde{\alpha}^2 \dim_S}{t\|H_S\|} e^{\beta\delta_{\min}} \pi$$

$$\leq \alpha^2 t e^{\beta\delta_{\min}} \|H_S\|^{-1} \pi. \tag{3.207}$$

For this we can have αt, which represents the total simulation time multiplied by the strength of the random interaction G, be constant and still take $\alpha \to 0$ to achieve arbitrarily small error.

Now we turn to bounding the total simulation time. We will let ρ_{fix} denote the fixed point of the dynamics. As before, we break the error into two pieces

$$\left\| \rho_{\text{fix}} - \left(\mathbb{E}_\gamma \Phi_\gamma \right)^{\circ L}(\rho) \right\|_1 \leq \left\| \rho_{\text{fix}} - \left(\mathbb{1} + \mathbb{E}_\gamma \mathcal{T}_{\text{on}}^{(\gamma)} \right)^{\circ L}(\rho) \right\|_1 + L\left(\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1 \right). \tag{3.208}$$

Let $\tilde{\lambda}_\star(\beta)$ denote the spectral gap for the rescaled transition matrix $\mathbb{E}_\gamma T_\gamma \cdot \left(\frac{2\|H_S\|(\dim+1)}{\alpha^2 t} \right)$, as this is the dimensionful prefactor in front of the transitions derived in Equation (3.205) and Equation (3.205). Jerison's Markov Relaxation Theorem 3.6 tells us that taking L to satisfy

$$L \geq \frac{\dim_S}{\lambda_\star} J \in \tilde{O}\left(\frac{\dim_S^2 \|H_S\|}{\alpha^2 t \tilde{\lambda}_\star(\beta)} \right) \tag{3.209}$$

is sufficient to guarantee $\left\| \rho_{\text{fix}} - \left(\mathbb{1} + \mathbb{E}_\gamma \mathcal{T}_{\text{on}}^{(\gamma)} \right)^{\circ L}(\rho) \right\|_1 \in \tilde{O}(\varepsilon)$. Now we balance the off-resonance and remainder errors

$$\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1 \leq \frac{8\alpha^2}{\delta_{\min}^2} + 16\sqrt{\frac{\pi}{2}}\dim_S (\alpha t)^3 = \frac{\alpha^2}{\delta_{\min}^2}\left(8 + 16\sqrt{\frac{\pi}{2}}\right), \tag{3.210}$$

and we see that setting $\alpha = 1/(\dim_S \delta_{\min}^2 t^3)$ makes the parenthesis a constant. To bound the total off-resonance and remainder error we take the product

$$L\Big(\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1\Big) \in \tilde{O}\left(\frac{\dim_S^2 \|H_S\|}{\alpha^2 t\tilde{\lambda}_\star(\beta)}\frac{\alpha^2}{\delta_{\min}^2}\right) = \tilde{O}\left(\frac{\dim_S^2 \|H_S\|}{t\delta_{\min}^2\tilde{\lambda}_\star(\beta)}\right). \tag{3.211}$$

Observe that setting

$$t = \frac{\dim_S^2 \|H_S\|}{\varepsilon\delta_{\min}^2\tilde{\lambda}_\star(\beta)} \tag{3.212}$$

is sufficient to make the above product $L\Big(\|\mathcal{T}_{\text{off}}\|_1 + \|R_\Phi\|_1\Big) \in \tilde{O}(\varepsilon)$.

We now turn to the $\beta \to \infty$ limit. For this we note that Lemma 3.7 guarantees that the ground state is a fixed point and that $\mathbb{E}_\gamma T_\gamma$ is upper triangular. We will show the ground state is unique by computing the spectrum of $\mathbb{E}_\gamma T_\gamma$. For this we take the $\beta \to \infty$ limit of the transitions in Equation (3.201) and Equation (3.202), which will give us the diagonal elements and then the spectrum. Starting with $i > j$ given in Equation (3.201) we get

$$\lim_{\beta\to\infty} \mathbb{E}_\gamma \langle j|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|j\rangle = \frac{\tilde{\alpha}^2}{2t\|H_S\|}\int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \text{sinc}^2(u)du \tag{3.213}$$

and for $i < j$ from Equation (3.202) we have

$$\lim_{\beta\to\infty} \mathbb{E}_\gamma \langle j|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|j\rangle = 0. \tag{3.214}$$

We denote the sinc integration above as

$$I_{\text{sinc}}(t) \coloneqq \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \text{sinc}^2(u)du, \tag{3.215}$$

and we will show later that this is constant for $\dim_S \geq 3$. Now these transitions allow us to compute the diagonal elements

$$\begin{aligned}
\lim_{\beta\to\infty} \mathbb{E}_\gamma \langle i|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|i\rangle &= -\sum_{j\neq i}\lim_{\beta\to\infty} \mathbb{E}_\gamma \langle j|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|j\rangle \\
&= -\sum_{j<i}\lim_{\beta\to\infty} \mathbb{E}_\gamma \langle j|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|j\rangle \\
&= -\frac{\tilde{\alpha}^2}{2t\|H_S\|}(i-1)I_{\text{sinc}}(t).
\end{aligned} \tag{3.216}$$

This gives a spectrum for $\mathbb{E}_\gamma T_\gamma$ as 0 and $-\frac{\tilde{\alpha}^2}{2t\|H_S\|}(i-1)I_{\text{sinc}}(t)$ for $i > 1$. This shows the ground state is the unique fixed point as 0 has multiplicity 1 in the spectrum. Further the spectral gap of the rescaled transition matrix $\tilde{\lambda_\star}(\beta)$ is then given by

$$\lim_{\beta\to\infty} \tilde{\lambda_\star}(\beta) = I_{\text{sinc}}(t). \tag{3.217}$$

We can repeat the analysis for finding suitable values for α, t, and L to guarantee thermalization and we find that

$$\alpha = \frac{1}{\dim_S \delta_{\min}^2 t^3}, \;\; t = \frac{4\dim_S^2 \|H_S\|}{\varepsilon\delta_{\min}^2}, \text{ and } \;\; L = \tilde{O}\left(\frac{\dim_S^2\|H_S\|}{\alpha^2 t}\right) \tag{3.218}$$

are sufficient to guarantee $\left\||1\rangle\langle 1| - \left(\mathbb{E}_\gamma \Phi_\gamma\right)^{\circ L}(\rho)\right\|_1 \in \tilde{O}(\varepsilon)$. Substituting this into Equation (3.209) yields

$$L \cdot t \in \tilde{O}\left(\frac{\dim_S^{16}\|H_S\|^7}{\delta_{\min}^8 \varepsilon^6 \tilde{\lambda_\star}(\beta)^7}\right) \tag{3.219}$$

as stated in the second claim of the theorem.

Our final task is to justify the third claim of the theorem, which involves showing that $I_{\text{sinc}}(t)$ as constant is valid. Using the choice of t directly above

$$I_{\text{sinc}}(t) = \int_{-\delta_{\min}\frac{t}{2}}^{\delta_{\min}\frac{t}{2}} \text{sinc}^2(u)du = 2\int_0^{\dim_S^2 4\|H_S\|/(\varepsilon\delta_{\min})} \text{sinc}^2(u)du. \tag{3.220}$$

Now we note that this integral is monotonic with respect to the upper limit of integration with a final value of $\lim_{t\to\infty} I_{\text{sinc}}(t) = \pi$. We note that we can capture a significant amount of this integral by just requiring the upper limit to be greater than the first zero of sinc located at $\frac{\pi}{2}$, which is true if $\varepsilon \leq \frac{4\dim_S^2\|H_S\|}{\pi\delta_{\min}}$. This value can be computed as $2\int_0^{\frac{\pi}{2}} \text{sinc}^2(u)du \geq 2.43$. This can be guaranteed by noting that ε can be at most 2, so the upper limit in Equation (3.220) is satisfied if

$$\varepsilon \leq 2 \leq \frac{3^2}{\pi} \leq \frac{\dim_S^2}{\pi} \leq \frac{\dim_S^2 4\|H_S\|}{\pi\delta_{\min}}, \tag{3.221}$$

as $\delta_{\min} \leq 4\|H_S\|$. This shows that for our choice of t then $|I_{\text{sinc}}(t) - \pi| \leq 0.71$, rendering it asymptotically constant as claimed. □

There are a few points that need to be addressed with the above theorem. The first is that our proof of the approximate fixed point utilizes rather poor bounds, resulting in diverging behavior as $\beta \to \infty$. For finite β our bounds on the change in the thermal state scales as $e^{\beta\delta_{\min}}$, which

diverges as β goes to ∞, but in this exact same limit we are able to show that the ground state is the emph{exact} fixed point of the Markov chain. This clear divergence in approximation error is a result of loose bounds and could be a potential avenue for improvement. The second point we would like to address is the rather high asymptotic scaling. This is the byproduct of a few things, the most important of which is the introduction of a $\frac{1}{t}$ in the reduction of the sinc integral. This causes a downstream effect of increasing the degree of each asymptotic parameter. To improve this one would need some kind of knowledge of the eigenvalues to prevent a uniform integration of each sinc term. We study the limiting case of this by assuming sample access to the exact eigenvalue differences $\Delta_S(i,j)$ in Section 3.4.2 and obtain much improved scaling. The second source of inflation in our asymptotic scaling could be our weak-coupling approach to studying the channel. As explored numerically in Section 3.3.2 we find that using different α scalings with respect to t can greatly effect the ε scaling of the total simulation time $L \cdot t$. A higher order analysis of this channel could lead to anytically better guarantees on the thermalization time required, even in this zero knowledge scenario.

3.4.2 Perfect Knowledge

Oftentimes when studying a system some knowledge of the eigenvalue gaps may be present. Our goal in this section is to study the extreme case of this scenario where one has knowledge of the exact eigenvalue differences. This is unlikely to happen with realistic quantum materials but instead serves as an ideal scenario for our channel to benchmark the effects of eigenvalue knowledge. Further, we note for some computational tasks, such as amplitude amplification, the eigenvalues may be explicitly computable and the real task is to find the dominant eigenvectors. A more realistic model for studying the impacts of eigenvalue knowledge on the total simulation time might be to place Gaussians at each of the $\Delta_S(i,j)$ values with some width σ. This is the model we use for numeric investigations in Section 3.4.3, but we were unable to compute the total simulation time required analytically. We find that our model of perfect knowledge allows us to show a reduced total simulation time budget, with the ratio of zero knowledge to perfect knowledge scaling as $\tilde{O}\left(\frac{\|H_S\|^7}{\delta_{\min}^7 \varepsilon^{3.5} \widetilde{\lambda}_\star(\beta)^{3.5}}\right)$, which gives an explicit worst-case simulation time bound for ground state preparation.

Theorem 3.5 (Perfect Knowledge Thermal State Prep): *Let H_S be a Hermitian matrix of dimension $\dim_S$ with no degenerate eigenvalues, ρ any input state that commutes with H_S, and let γ be a random variable with distribution $\Pr[\gamma = \Delta_S(i,j)] = \frac{\eta_{\Delta_S}(i,j)}{\binom{\dim_S}{2}}$ where $\eta_{i,j}$ is the number of times a particular eigenvalue difference appears. For any $\beta \in [0,\infty]$ the thermal state can be prepared with controllable error*

$$\left\| \rho_S(\beta) - \left(\mathbb{E}_\gamma \Phi_\gamma\right)^{\circ L}(\rho) \right\|_1 \in \tilde{O}(\varepsilon) \tag{3.222}$$

with the following parameter settings

$$\alpha = \frac{\delta_{\min}\varepsilon^{1.5}\tilde{\lambda}_\star(\beta)^{1.5}}{\dim_S^7}, t = \frac{\dim_S^2}{\delta_{\min}\varepsilon^{0.5}\tilde{\lambda}_\star(\beta)^{0.5}}, \text{ and } \; L \in \tilde{O}\left(\frac{\dim_S^{14}}{\varepsilon^2\tilde{\lambda}_\star(\beta)^3}\right), \tag{3.223}$$

where $\tilde{\lambda}_\star(\beta)$ is the spectral gap of the rescaled transition matrix $\mathbb{E}_\gamma T_\gamma \cdot \frac{\binom{\dim_S}{2}}{\tilde{\alpha}^2}$. This gives the total simulation time required as

$$L \cdot t \in \tilde{O}\left(\frac{\dim_S^{16}}{\delta_{\min}\varepsilon^{2.5}\tilde{\lambda}_\star(\beta)^{3.5}}\right). \tag{3.224}$$

All of the above conditions hold in the ground state limit as $\beta \to \infty$ and further we can compute a lower bound on the spectral gap of the rescaled transition matrix as

$$\lim_{\beta\to\infty} \tilde{\lambda}_\star(\beta) = \min_{i>1} \sum_{j<i} \eta_{\Delta(i,j)} \geq 1. \tag{3.225}$$

Proof: This proof structure is structurally similar to the proof of Theorem 3.5. To show that the thermal state is the fixed point we will need to compute transition factors of the form $\mathbb{E}_\gamma \langle j|\mathcal{T}_{\text{on}}^{(\gamma)}(|i\rangle\langle i|)|j\rangle$ for use in Lemma 3.7. Using the on-resonance definition in Equation (3.123) we have for $i > j$

$$
\begin{aligned}
&\mathbb{E}_\gamma \langle j|\mathcal{T}^{(\gamma)}_{\text{on}}(|i\rangle\langle i|)|j\rangle \\
&= \tilde{\alpha}^2 \mathbb{E}_\gamma \frac{1}{1+e^{-\beta\gamma}} \mathbb{I}[|\Delta_S(i,j)-\gamma| \leq \delta_{\min}] \text{sinc}^2\left((\Delta_S(i,j)-\gamma)\frac{t}{2}\right) \\
&+\tilde{\alpha}^2 \mathbb{E}_\gamma \frac{e^{-\beta\gamma}}{1+e^{-\beta\gamma}} \mathbb{I}[|\Delta_S(i,j)+\gamma| \leq \delta_{\min}] \text{sinc}^2\left((\Delta_S(i,j)+\gamma)\frac{t}{2}\right) \\
&= \tilde{\alpha}^2 \sum_{\Delta_S(k,l)} \Pr[\gamma = \Delta_S(k,l)] \frac{\mathbb{I}[|\Delta_S(i,j)-\Delta_S(k,l)| \leq \delta_{\min}]}{1+e^{-\beta\Delta_S(k,l)}} \text{sinc}^2\left((\Delta_S(i,j)-\Delta_S(k,l))\frac{t}{2}\right) \\
&= \tilde{\alpha}^2 \frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}} \frac{1}{1+e^{-\beta\Delta_S(i,j)}}.
\end{aligned} \tag{3.226}
$$

$i < j$ can be computed similarly as

$$
\mathbb{E}_\gamma \langle j|\mathcal{T}^{(\gamma)}_{\text{on}}(|i\rangle\langle i|)|j\rangle = \tilde{\alpha}^2 \frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}} \frac{e^{-\beta\Delta_S(k,l)}}{1+e^{-\beta\Delta_S(k,l)}}. \tag{3.227}
$$

This allows us to compute the detailed-balance like condition in Equation (3.145) for $i > j$

$$
\begin{aligned}
&\frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \mathbb{E}_\gamma \langle j|\mathcal{T}^{(\gamma)}_{\text{on}}(|i\rangle\langle i|)|j\rangle - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \langle i|\mathcal{T}^{(\gamma)}_{\text{on}}(|j\rangle\langle j|)|i\rangle \\
&= \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \tilde{\alpha}^2 \frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}} \frac{1}{1+e^{-\beta\Delta_S(i,j)}} - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \tilde{\alpha}^2 \frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}} \frac{e^{-\beta\Delta_S(i,j)}}{1+e^{-\beta\Delta_S(i,j)}} \\
&= \frac{\tilde{\alpha}^2}{\mathcal{Z}_S(\beta)} \frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}} \left(\frac{e^{-\beta\lambda_S(i)}}{1+e^{-\beta\Delta_S(i,j)}} - e^{-\beta\lambda_S(j)} \frac{e^{-\beta\Delta_S(i,j)}}{1+e^{-\beta\Delta_S(i,j)}} \right) \\
&= 0.
\end{aligned} \tag{3.228}
$$

For $i < j$ we can repeat the same steps to argue that detailed balance also holds in this case.

$$
\begin{aligned}
&\frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \mathbb{E}_\gamma \langle j|\mathcal{T}^{(\gamma)}_{\text{on}}(|i\rangle\langle i|)|j\rangle - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \langle i|\mathcal{T}^{(\gamma)}_{\text{on}}(|j\rangle\langle j|)|i\rangle \\
&= \frac{e^{-\beta\lambda_S(i)}}{\mathcal{Z}_S(\beta)} \tilde{\alpha}^2 \frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}} \frac{e^{-\beta\Delta_S(j,i)}}{1+e^{-\beta\Delta_S(j,i)}} - \frac{e^{-\beta\lambda_S(j)}}{\mathcal{Z}_S(\beta)} \tilde{\alpha}^2 \frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}} \frac{1}{1+e^{-\beta\Delta_S(j,i)}} \\
&= \frac{\tilde{\alpha}^2}{\mathcal{Z}_S(\beta)} \frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}} \left(\frac{e^{-\beta\lambda_S(j)}}{1+e^{-\beta\Delta_S(j,i)}} - \frac{e^{-\beta\lambda_S(j)}}{1+e^{-\beta\Delta_S(j,i)}} \right) \\
&= 0.
\end{aligned} \tag{3.229}
$$

This is sufficient to show that the thermal state $\rho_S(\beta)$ is a fixed point via Lemma 3.7. As we have also shown that the probability of transitioning from any state $|i\rangle\langle i|$ to any other state $|j\rangle\langle j|$ is nonzero this gives a nonzero expected hitting time for any pair of states. This implies the Markov chain is ergodic and that $\rho_S(\beta)$ is the *unique* fixed point.

Next we bound the total simulation time required. For reasons similar to the harmonic oscillator in Section 3.3.1 we are unable to compute the spectral gap of the Markov matrix. We start the analysis in a similar manner by using the decomposition

$$\left\|\rho_S(\beta) - \left(\mathbb{E}_\gamma \Phi_\gamma\right)^{\circ L}(\rho)\right\|_1 \leq \left\|\rho_S(\beta) - \left(\mathbb{E}_\gamma \mathbb{1} + \mathcal{T}_{\text{on}}^{(\gamma)}\right)^{\circ L}(\rho)\right\|_1 + L\left(\left\|\mathbb{T}_{\text{off}}\right\|_1 + \left\|R_\Phi\right\|_1\right). \tag{3.230}$$

We bound the Markov error via Theorem 3.6. This theorem guarantees that choosing L to satisfy

$$L \geq \frac{\dim_S^{\binom{\dim_S}{2}}}{\tilde{\alpha}^2 \tilde{\lambda}_\star(\beta)} J \in \tilde{O}\left(\frac{\dim_S^4}{\alpha^2 t^2 \tilde{\lambda}_\star(\beta)}\right), \tag{3.231}$$

where $\tilde{\lambda}_\star(\beta)$ is the spectral gap of the rescaled transition matrix $\mathbb{E}_\gamma T_\gamma \cdot \frac{\binom{\dim_S}{2}}{\tilde{\alpha}^2}$, is sufficient for $\left\|\rho_S(\beta) - \left(\mathbb{E}_\gamma \mathbb{1} + \mathcal{T}_{\text{on}}^{(\gamma)}\right)^{\circ L}(\rho)\right\|_1 \in \tilde{O}(\varepsilon)$. We now use this to bound the total off-resonance and remainder error after balancing the two contributions asymptotically

$$\left\|\mathcal{T}_{\text{off}}\right\|_1 + \left\|R_\Phi\right\|_1 \leq \frac{8\alpha^2}{\delta_{\min}^2} + 16\sqrt{\frac{\pi}{2}}\dim_S(\alpha t)^3 = \frac{\alpha^2}{\delta_{\min}^2}\left(8 + 16\sqrt{\frac{\pi}{2}}\alpha \dim_S \delta_{\min}^2 t^3\right). \tag{3.232}$$

Setting $\alpha = \frac{1}{\dim_S \delta_{\min}^2 t^3}$ is sufficient to make the parenthesis a constant. Lastly to get the total error in $\tilde{O}(\varepsilon)$ we multiply the above by the L chosen before

$$L\left(\left\|\mathbb{T}_{\text{off}}\right\|_1 + \left\|R_\Phi\right\|_1\right) \in \tilde{O}\left(\frac{\dim_S^4}{\alpha^2 t^2 \tilde{\lambda}_\star(\beta)} \frac{\alpha^2}{\delta_{\min}^2}\right) = \tilde{O}\left(\frac{\dim_S^4}{t^2 \delta_{\min}^2 \tilde{\lambda}_\star(\beta)}\right). \tag{3.233}$$

Choosing

$$t = \frac{\dim_S^2}{\delta_{\min}\sqrt{\varepsilon \tilde{\lambda}_\star(\beta)}} \tag{3.234}$$

is sufficient to guarantee $L\left(\left\|\mathbb{T}_{\text{off}}\right\|_1 + \left\|R_\Phi\right\|_1\right) \in \tilde{O}(\varepsilon)$ and that the total error $\left\|\rho_S(\beta) - \left(\mathbb{E}_\gamma \Phi_\gamma\right)^{\circ L}(\rho)\right\|_1 \in \tilde{O}(\varepsilon)$. Combining the above results for α, L and t yields the theorem statement for finite β.

We now show how to calculate $\tilde{\lambda}_\star(\beta)$ in the $\beta \to \infty$ limit. From Lemma 3.7 we know that $\mathbb{E}_\gamma T_\gamma$ will be upper triangular, implying again that we can compute the spectrum if we can compute the diagonal elements of the matrix. Using our computation of the off-diagonal elements from the proof of Theorem 3.4 we have for $i > 1$

$$\lim_{\beta\to\infty}\mathbb{E}_\gamma\langle i|\mathcal{T}^{(\gamma)}_{\text{on}}(|i\rangle\langle i|)|i\rangle = -\lim_{\beta\to\infty}\sum_{j\neq i}\langle j|\mathcal{T}^{(\gamma)}_{\text{on}}(|i\rangle\langle i|)|j\rangle$$

$$= -\lim_{\beta\to\infty}\sum_{j<i}\tilde{\alpha}^2\frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}}\frac{1}{1+e^{-\beta\Delta_S(i,j)}} - \lim_{\beta\to\infty}\sum_{j>i}\tilde{\alpha}^2\frac{\eta_{\Delta(i,j)}}{\binom{\dim_S}{2}}\frac{e^{-\beta\Delta_S(j,i)}}{1+e^{-\beta\Delta_S(j,i)}}$$

$$= -\frac{\tilde{\alpha}^2}{\binom{\dim_S}{2}}\sum_{j<i}\eta_{\Delta(i,j)}. \tag{3.235}$$

For $i = 1$ as we know the ground state is fixed we have $\lim_{\beta\to\infty}\mathbb{E}_\gamma\langle 1|\mathcal{T}^{(\gamma)}_{\text{on}}(|1\rangle\langle 1|)|1\rangle = 0$. This gives the spectrum of $\mathbb{E}_\gamma T_\gamma$ as 0 and $-\frac{\tilde{\alpha}^2}{\binom{\dim_S}{2}}\sum_{j<i}\eta_{\Delta(i,j)}$ for all $i > 1$. From this spectrum we can conclude that the ground state is the *unique* fixed point as 0 has multiplicity 1 in the spectrum and further that the spectral gap can be bounded from below as

$$\lim_{\beta\to\infty}\tilde{\lambda}_\star(\beta) = \min_{i>1}\sum_{j<i}\eta_{\Delta(i,j)} \geq 1. \tag{3.236}$$

□

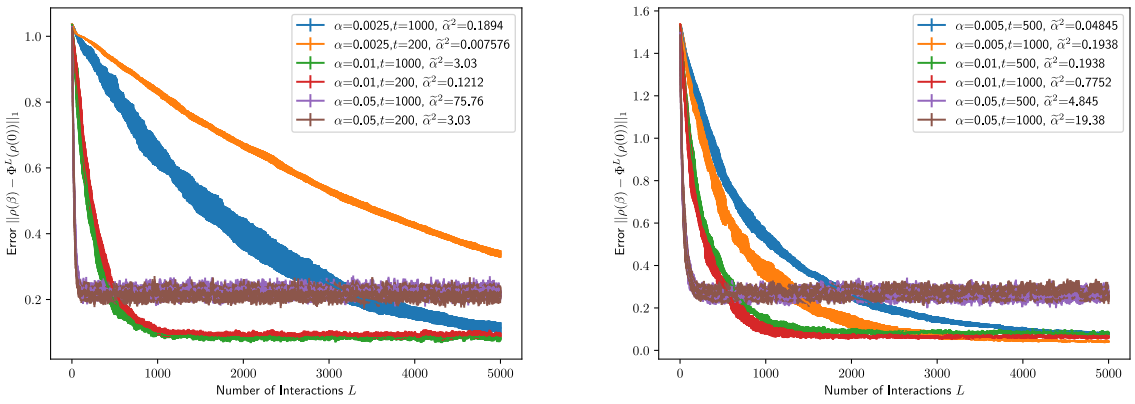
The above theorem shows that if we sample our transitions strategically rather than randomly then we can achieve much faster convergence to the groundstates in our upper bounds. Importantly, the scaling of the total simulation time is also independent of the norm of H_S in this case, whereas the time required by the zero knowledge case does. Unfortunately, the dimensional scaling of $\dim_S^{16}$ is prohibitive for all but the smallest dimensional systems. This scaling is again likely loose because of a number of assumptions that we make above and also a result of our insistence that the channel always operate inside the regime of weak coupling. In contrast, we will see below that equilibration can be much faster if strong coupling is assumed. Finally, it is worth noting that although perfect knowledge is assumed, a cooling schedule is not used. By changing the distribution depending on the temperature of the Gibbs state it is possible that even better scaling may be achievable.

3.4.3 Numerics

The analytic results developed in the previous two sections provide strong guarantees on the correctness of our routine for most quantum systems, however the bounds on the total simulation are fairly high degree polynomials in the parameters of interest. One crucial interpretation of the two different results is that knowledge of the eigenvalue differences of H_S can lead to significantly better simulation time bounds, but this knowledge is not emph{crucial} for thermalization. Another important takeaway is that we cannot bound the simulation time or number of interac-

tions required for finite β as we cannot bound the spectral gap of the expected transition matrix $\mathbb{E}_\gamma T_\gamma$. The purpose of this section is to investigate these two theoretic takeaways numerically with small Hydrogen chain systems. These systems are some of the smallest chemical systems that still display some real-world chemical behavior, and as a result are typically used in many numeric benchmarks for quantum routines.

Our first experiment conducted is to study the effects of changing α and t on the trace distance error as a function of L. The theory developed in prior sections is very prescriptive; to reach a specific trace distance of ε all of our theorems give a value of α, L, and t that guarantee a distance of at most $\tilde{O}(\varepsilon)$ but say nothing about what this convergence looks like. In Figure ref{fig:h_chain_error} we study the effects of different choices of α and t on this convergence rate. To generate the Hamiltonians used in these experiments we created a small chain of equally spaced hydrogen nuclei with an STO-3G active space for the electrons. Hamiltonian creation was done with OpenFermion [38] and PySCF [39]. Once the Hamiltonians were generated, the distance to the thermal state $\rho_S(\beta)$ for each was tracked over $L = 5000$ interactions. For both Hydrogen 2 and Hydrogen 3 we chose $\beta = 4$ for consistency, this gave a ground state overlap of around 0.56 for Hydrogen 2 and 0.26 for Hydrogen 3. **TODO: Need to replace this with actual plots.**



(a) (b)

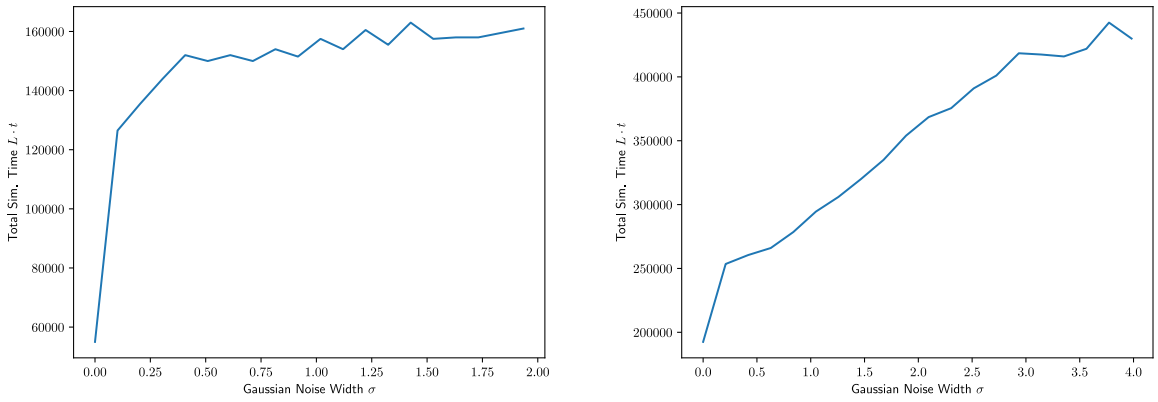
Figure 10: These plots show the distance to the target thermal state for Hydrogen 2 and Hydrogen 3 chains as the number of interactions L increases. For both Hydrogen 2 and 3 we set $\beta = 4.0$, which gives a ground state overlap of greater than 0.5 for Hydrogen 2 and 0.25 for Hydrogen 3. γ for both (a) and (b)) was generated by placing a Gaussian at the average energy $\frac{\mathrm{tr}(H_S)}{\dim_S}$ with a width of $\frac{\|H_S\|}{2}$. We note that a variety of $\tilde{\alpha}^2$ values were chosen to demonstrate the faster convergence, but higher error, of strong coupling.

There are a few key takeaways from Figure ref{fig:h_chain_error}. The first is that we observe increasing α and t tend to increase the convergence rate, with α visuallly appearing more important. At higher values of α changes in t appear to make less of an impact on the error. We also observe that our channel is seemingly robust beyond our weak-interaction analysis. For values of $\tilde{\alpha}^2$, the weak-interaction expansion parameter, we observe that values as high as $\tilde{\alpha}^2 \approx 3$ can have rapid convergence to fairly low error floors. We note that these coupling values that go beyond weak-interaction seem to lead to faster convergence of the dynamics at the cost of a larger error floor. It remains an open question if dynamic choices of α and t could lead to better performance of the overall routine, one could use very large $\tilde{\alpha}$ initially to quickly thermalize with large error and then decrease $\tilde{\alpha}$ to fine-tune the final state.

The second observation we make is on the choice of the environment gaps γ. For both Hydrogen 2 and 3 we selected γ randomly from a Gaussian with mean $\frac{\mathrm{tr}(H_S)}{\dim_S}$ and standard deviation of $\frac{\|H_S\|}{2}$. This choice of γ is completely heuristic and was intended to have a large overlap with what the typical eigenvalue differences may look like with a large enough deviation to pick up potentially large differences. Although this heuristic works well enough to show

convergence, it leads us to question if the error convergence or floors can be improved with better choices of γ.

In Figure 11 we demonstrate that better choices of γ do in fact reduce the total simulation time needed for thermalization. To generate the data we compute the number of interactions needed at a fixed coupling constant α and a fixed time t as a function of the noise added to our samples for γ. We generate one sample of γ by first computing the eigenvalue spectrum of H2 or H3 exactly, then by choosing two non-equal eigenvalues, and finally sampling a Gaussian centered at the absolute value of the difference. The width of this Gaussian then serves as a proxy for the amount of knowledge one may have about the system's eigenvalues. We plot the total simulation time with respect to this width as it varies from 0 to the spectral width $\max_i \lambda_S(i) - \min_j \lambda_S(j)$. The results align well with our theoretic analysis: having knowledge of the eigenvalues of the system can be used to speed up the thermalization routine but if one does not have any knowledge at all the thermal state can still be prepared. It is an open question if the dependence of the total simulation time $L \cdot t$ on the noise level σ can be determined analytically.



(a) Hydrogen 2 (b) Hydrogen 3

Figure 11: In these plots the amount of total simulation time needed to prepare a $\beta = 2.0$ thermal state with $\alpha = 0.01$ and $t = 500$ is tracked as a function of the noise added to samples of γ. A sample for γ is generated by choosing two non-equal eigenvalues from the system spectrum and adding a Gaussian random variable with standard deviation σ. For each value of σ the resulting state needs to have an average trace distance of less than 0.05 for 100 samples.

3.5 Discussion

Thermal state preparation is likely to be a crucial preparation step for the simulation of quantum systems on digital quantum computers. We have presented a thermalization routine for this task that has an optimally minimal number of overhead ancilla qubits and compiles to remarkably simple circuits of time independent Hamiltonian evolution of the unprocessed system Hamiltonian, with no filtering or rejection steps and no Fourier weighted jump operators of Linbladians. Our routine is based on relatively recent classical Monte Carlo techniques, specifically Hamiltonian Monte Carlo [69] and the end result bears striking resemblance to the Repeated Interactions framework in open quantum systems [70]. In the Hamiltonian Monte Carlo algorithm thermal states over a position coordinate q is prepared by sampling momentum p from the Boltzmann distribution for Gaussians $e^{-\beta\frac{p^2}{2}m}$ followed by time evolution. Classical Hamiltonian dynamics is enough to couple the position and momentum, leading to the Boltzmann distribution over q with enough time and samples. In the repeated interactions framework a quantum system interacts with many small environments, typically a single photon, that is repeated until the system thermalizes.

Our work extends these procedures to quantum algorithms. For Hamiltonian Monte Carlo, instead of adding in momentum variables we add in a single ancilla qubit to serve as our extra state space. We do not have the luxury of classical Hamiltonian dynamics that couples these two spaces or registers, so we add in a randomized interaction term to the Hamiltonian. After simulating the time dynamics of this system-ancilla pair and repeating multiple times we are able to thermalize the system to the same β. On the other hand, the Repeated Interactions framework typically is concerned with thermodynamic limits, such as infinite time or interactions, and specific system-interactions pairs. As our procedure is intended to be used as a subroutine for quantum computers our techniques work for arbitrary, non-degenerate, Hamiltonians and purposefully use randomized interactions as opposed to a fixed interaction model.

One benefit of our thermalization procedure is that it can be compiled all the way down to the circuit level with minimal overhead in complexity. The elements of the channel that need to be compiled are: the ancilla qubit state preparation, the initial state preparation for the system, the time evolution of $H + \alpha G$, and the partial trace. The ancilla state preparation can be done starting from the ground state with a Pauli-X rotation. The initial system state can be any

state that commutes with H_S, the two leading choices are the maximally mixed state or a Haar random pure state. The maximally mixed state can be prepared using just CNOT gates at a cost of higher qubit count and a Haar random pure state can be prepared with no additional qubit overhead by applying a deep enough random circuit [71].

The time evolution of $H + \alpha G$ can be broken down into simulation of H_S and then αG by one application of Trotter. The simulation of H_S can be implemented in two different ways depending on the access model for the system Hamiltonian H_S. If H_S is provided as a block-encoding then there exist optimal simulation techniques that add only a single extra ancilla qubit [21]. If H_S is provided as a sum of k-local Pauli strings or as a sparse matrix then product formula techniques can be used [26] at zero extra overhead in ancilla qubits. To simulate the time evolution of αG we can use the following breakdown $e^{i\alpha G t} = e^{i\alpha U_G D U_G^\dagger t} = U_G e^{i\alpha D t} U_G^\dagger$. This unitary can be implemented with a 2-design to approximate the U_G, this is due to the fact that we only expand the channel to second order in α. Random Clifford circuits cite{webb2015clifford} are sufficient for this purpose. To simulate $e^{i\alpha D t}$ random Z rotations and controlled-Z rotations should be sufficient, as we only ever rely on the eigenvalues being pairwise independent in our analysis. In the worst case the number of random gates in the Z basis that would need to be applied would scale with the dimension of the system, adding an overall factor of dim_S to the preparation. In the product formula case we further remark that $H + \alpha G$ could be simulated in total with a composite technique of using Trotter for H and randomized compilation for αG [35].

In classical Hamiltonian Monte Carlo it is well known that sharp gradients in the Hamiltonian require longer simulation time and more samples to converge. Our quantum routine has a much more subtle dependence on the structure of the Hamiltonian. As our single ancilla qubit only has one energy difference γ, we have to tune this energy difference to allow for energy to be siphoned off from the system into the ancilla. This would present a conundrum, as knowing spectral gaps is as difficult or harder than preparing ground states of arbitrary quantum Hamiltonians, but we are able to prove that our routine is robust to complete ignorance of these differences. We show that this ignorance comes at an asymptotic cost in the amount of resources needed to prepare the thermal state. We numerically verify that knowledge of the eigenvalue differences can be used to speed up the total simulation time, as demonstrated in Figure 10. We posit that this behavior serves as a crucial entry point for heuristics about Hamiltonian spectra into thermal state preparation algorithms. No prior thermal state preparation routines have had

such an explicit demonstration of the utility of such knowledge. It was our hope to analytically quantify the speedups gained as a function of the relative entropy between a heuristic guess for the eigenvalue differences and the true spectra, but our numeric evidence will have to suffice until future work can clarify this dependence.

We would like to make a few remarks on potential improvements for the analysis of this channel. As we have demonstrated numerically, our guaranteed analytic values of α and t that lead to thermalization are drastically overestimated. We conjecture that this is due to our truncation of the weak-coupling expansion and in Figure 9 we demonstrate that taking $\alpha \propto \frac{1}{t}$ and $t \propto \frac{1}{\sqrt{\varepsilon}}$ drastically outperforms our analytically derived bounds of $\alpha \propto \frac{1}{t^3}$, by almost 4 orders of magnitude at $\varepsilon \approx 0.005$. It is an open question of how to analyze this channel in the strong-coupling regime, and our numeric results suggest that such an analysis may indicate better performance of our protocol than a weak-coupling expansion can show. It is also an open question of whether dynamically chosen values of α and t, such as having strong coupling and low time at the beginning and gradually decreasing α and increasing t, can outperform static α and t. We also suspect that the Markov relaxation theorem we used greatly overestimates the number of interactions needed. It remains to be seen if better Markov theory is needed or if the convergence time could be characterized based on the overlap of the initial state with the thermal state, which is a property that a few ground state preparation algorithms demonstrate. Another potential avenue for improving the analysis of this channel is whether different randomized interactions or even eigenvector heuristics can be beneficial. For example, in the harmonic oscillator if one has knowledge of the creation and annihilation operators $a^\dagger$ and a, could one simply use the interaction $a^\dagger \otimes (X + iY) + a \otimes (X - iY)$ instead of involving a randomized G that relies on a Haar average? The last potential improvement is to extend our spectral gap computations using perturbation theory. We are only able to compute the spectral gap $\tilde{\lambda}_\star(\beta)$ in the limit of $\beta \to \infty$, but it should be possible to compute a perturbation on the order of $\frac{1}{\beta}$. This would give the simulation time needed to prepare low-temperature thermal states as opposed to zero-temperature states.

Lastly, we would like to speculate on possible applications of this routine to other quantum information processing tasks. The first question that arises is if these techniques could be used in the training of quantum Boltzmann machines, which are essentially thermal states. It is an open question if our thermalizing techniques could be used to either train models or to generate

output samples from an already trained model. Through the process of demonstrating that this channel prepares the system in the thermal state we have calculated the output of our channel for both the system and the environment registers, and for much larger environments than single qubits. We can turn this protocol on it's head and ask how much information about the system are these ancilla qubits carrying away with them? Preliminary explorations suggest that given knowledge about eigenvalue gaps one can use transition statistics in the ancilla qubits to infer what the inverse temperature β is of the system, assuming the system is in a thermal state. Could this thermalizing channel instead be used to develop a Bayesian model to update beliefs about Hamiltonian spectra and system temperatures? This would represent an interaction agnostic model for performing quantum thermometry or spectroscopy, which to the best of our knowledge has not been developed yet.

Chapter 4

Conclusion

After 30+ years of intense theoretic development, the most advanced end-to-end pipeline we have for utilizing potential quantum computers beyond the capabilities of classical computers is the simulation of quantum systems. Factoring large integers via Shor's algorithm is one provable advantage, assuming widely adopted complexity theoretic conjectures such as $\mathsf{BQP} \neq \mathsf{BPP}$. However, the advancing maturity [72] of lattice-based cryptosystems implies that using quantum computers solely for factoring integers has decreasing utility as the time to reach thousands of logical qubits grows. Many other provable end-to-end quantum speedups, such as Grover's algorithm for unstructured search or Brandao's Semi-Definite Program solver, only offer quadratic improvements over classical techniques. Evidence suggests that these speedups would have to be applied to unfeasibly large instance sizes in order for these speedups to provide an advange over classical computers.

Recent efforts have also explored larger polynomial improvements, such as the quadratic improvement for planted noisy kXOR problem [73], but progress remains sparse. There exist other exponential speedups for problems such as Glued Trees traversal or Sunflower graph traversal, but these problems are relatively contrived and have yet to find many "real world" applications. The most difficult to analyze speedups are those that rely only on existing classical algorithms, such as normalized Betti number estimation for clique homology [74], optimization, and linear systems solving. This is due to the lack of good classical lower bounds, and sometimes the same traits that make a problem amenable to fast quantum algorithms can lead to better classical algorithms. More recently, efforts have been made to probe the approximation ratios

provided by quantum computers on difficult combinatorial optimization problems [75], which also suffers the same problem that benchmarks can only be made against existing classical algorithms.

Simulating quantum systems has the advantage of being one of the oldest quantum algorithms [34] and as a result has seen much more intense scrutiny. Further, the problem of estimating observables of quantum states has been studied for many decades longer [47] and we have a decent understanding of when classical methods tend to fail. This understanding is constantly changing [76] with the advent of new classical algorithms, such as Density Matrix Renormalization Group (DMRG) methods [77] and tensor networks [78], but many papers [79] expect early quantum computers to be useful for the electronic structure problem in medium sized molecules with strong electron correlation. Condensed matter systems and models such as the Fermi-Hubbard model [80,81] are some of the most promising candidates for quantum advantage. This is due to their inherent symmetries, which makes for easy to analyze quantum algorithms, and the strong correlation between electrons makes classical algorithms difficult to scale [82].

In this thesis, we presented two new quantum algorithms that can be added to the quantum simulation toolkit. The first is an extension of product formulas to include both random and deterministic sections. We were able to provide generic conditions on when these composite simulations can provide advantages over their constituent methods. We find that systems where the strength of the Hamiltonian terms decays exponentially, i.e. if $H = \sum_i H_i$ then $\|H_i\| \propto 2^{-i}$, have a relatively large parameter window for improvement. These analytically derived cost advanteges were then verified numerically for small systems. Further, this algorithm was extended to imaginary time evolution and provides new avenues for improving classical estimations of quantum observables.

The second contribution we make is an algorithm for preparing thermal states $\frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}$ on a quantum computer. This algorithm assumes access to the Hamiltonian only via time evolution operators e^{iHt} and adds only a single ancilla qubit. At a high level, our technique works by randomizing over all possible interaction terms that could be present between this ancilla qubit and the system of interest. By randomly choosing the energy gap of the Hamiltonian dictating the time evolution of the ancilla qubit we can rigorously show that the thermal state is an approximate fixed point at finite inverse temperature β if no knowledge of the Hamiltonian spectrum is assumed. We can show that the thermal state is the unique fixed point exactly in the

ground state limit ($\beta \to \infty$) or if the eigenvalue differences of H are known. Remarkably, we are able to bound the number of interactions, the coupling strength, and the time per interaction needed in order to thermalize the system state to within trace distance $\tilde{O}(\varepsilon)$ of the thermal state.

Taken as a whole, this thesis provides a blueprint for studying thermal equilibrium for quantum states on digital fault tolerant quantum computers. The two main contributions we presented work very well together: our thermal state preparation algorithm reduces the problem of thermalization to quantum simulation with a deterministic system hamiltonian and a randomly chosen interaction. Our composite simulation algorithm analyzes this situation precisely and provides detailed error estimates and oracle query costs. This gives a direct method for compiling quantum circuits for estimating thermal observables using simple primitives that are likely to be highly optimized for quantum computers.

For the rest of this conclusion I would like to discuss possible extensions to these methods. First with our composite simulation algorithm. The most obvious extension we could make is to propose better partitioning schemes that can take advantage of commutativity between subsections of the Hamiltonian. A classic example of this is that when simulating the most basic Fermi-Hubbard models the on-site potentials are typically simulated for all sites and then the tunnelling terms are simulated. Ideally, one would like to be able to collect some level of commutator data, such as all pairs $\left\| \left[H_i, H_j\right] \right\|$ or even higher levels, and use this data along with spectral norm data to construct a better partition than the straightforward `chop`. This may incorporate more stages than just a single partition into $A + B$.

To extend these ideas further, one could incorporate different simulation techniques than product formulas into a composite simulation. In [35] we proposed including block-encoding techniques, such as multiproduct formulas [22] or qubitization [21]. It remains an open question if block-encoding constants, the dominant cost contributor for these techniques, can be significantly reduced by taking terms out of the block-encoding partition and into a product formula partition. The final possible improvement one could make to partitioning schemes is to include information about entanglement. This was recently shown to improve the analysis for Trotter formulas [83] and extending these ideas to inform partitions for composite simulations is a natural extension.

In regards to our thermal state preparation routine there are a few avenues for improvement, which is needed as our provable scaling of $\dim^{16}$ is prohibitively high for a naive implementation. To eliminate the dimensionful factors contributing to the cost there are three areas that need to be addressed. The first is that the Haar integration over the entire system and environment can be reduced based on symmetries of the system. For example, similar algorithms have been studied for specific systems with less general interaction terms [84,85] and have seen empirical success. The stringent requirements on the randomness of the interaction could be reduced if knowledge of the symmetries of the system are known. For example on 1D and 2D lattice models that are translationally invariant we could require that our interaction also be translationally invariant. This would make our interaction scale invariant with respect to the number of lattice sites and eliminate one of the factors of dimension, and efficient mixing times have been shown for Linbladian based thermal state preparation routines in this translationally invariant setting [86].

The second factor of dimension that could be reduced is the introduction of $\dim$ from the Markov relaxation theorem we used from [63]. This result is general purpose and works for an arbitrary non-reversable Markov chain, but the cost of this flexibility is a high overhead in the number of interactions needed. This could be reduced using knowledge of the distance of the input state to the fixed point, as is typical in many ground state preparation routines [87].

The last factor of dimension that would need to be eliminated is from the remainder bound $\|R_\Phi\|_1$. The factor of dimension appears as we only study thermalization with respect to trace distance from the thermal state. This is the most rigorous distance metric one could use, but may not be the most physically relevant metric. If one has a collection of observables O_i, then measuring the deviation $\max_i \left|\langle O_i\rangle_\beta - \mathrm{tr}(O_i \Phi^{\circ L}(\rho))\right|$ is a more physically relevant error metric. Bounding a relaxed error metric for our thermalization routine for a physically relevant class of observables, such as two-body correlators for example, is a promising avenue of research that could lead to increased performance.

These three areas constitute the main introductions of dimensionful factors into the runtime analysis of our thermalization routine. One last area for potential improvements we would like to mention is the issue of strong-coupling analysis. Numerically, we observed significantly faster thermalization with strong-coupling at the penalty of a higher noise floor. This leads us to conjecture that significant savings could be had with a "cooling schedule", where the thermalization procedure begins with very strong coupling for short times that gradually gets

lowered to small coupling constants and long per-interaction simulation times. Unfortunately our Taylor series expansion does not seem amenable to higher order analysis, and this must be studied numerically. Developing a theoretic understanding of strong-coupling could lead to much better algorithmic guarantees beyond numeric evidence. The previously mentioned avenue of studying specific observable distributions could be a problem better suited for studying a strong-coupling approach.

Lastly, we return to the problem laid out in Chapter 1 of thermal equilibrium. As our channel works for arbitrary Hamiltonians at finite β, this can effectively be viewed as a way for *defining* thermal equilibrium. One way to do so would be to define the fixed point for the joint system-environment dynamics and declare that thermal equilibrium is whatever the fixed point is of this map. One logical way to view this is to take the "environment", which was a single qubit in our work, and instead view it as a probe. By then viewing the system as a reservoir, we can interact the probe with the system and use it's state to *define* what it's temperature is. This gives a simple way of performing interaction agnostic thermometry.

Previous thermal state preparation algorithms fall into one of two camps: they can be "computational" in nature and do not mimic any natural processes or do try to mimic a system-environment interaction (except for ETH inspired algorithms which simulate a large closed system). As the computational algorithms do not have explicit models for an environment they do not provide new perspectives on physical system-environment interactions. Algorithms more inspired by natural processes tend to use the Linbladian formalism, which explicitly does not simulate the state of the environment and instead captures its effects only on the system. Turning to algorithms that do explicitly model a system-environment, many other works have used similar ideas [84,85,88] but tend to use limited interaction models and therefore are only provably correct for a limited subset of systems. What makes our algorithm the first of it's kind is that it provably works for any non-degenerate Hamiltonians at any temperature, even ground states.

Appendix A

Alternative Randomized interactions

For this section I want to show that the following random matrix should satisfy the thermal state prep conditions that I need. The random matrix is a randomly chosen string of Pauli Z operators and a random phase

$$Z = (-1)^{z_0} Z_1^{z_1} \otimes ... \otimes Z_n^{z_n}. \tag{5.237}$$

Where $\Pr[z_i = 0] = \Pr[z_i = 1] = \frac{1}{2}$. Then we need

$$\mathbb{E}_Z Z = 0 \tag{5.238}$$

and

$$\mathbb{E}_Z \langle i|Z|i\rangle\langle j|Z|j\rangle = \delta_{i,j}. \tag{5.239}$$

First the expectation.

$$\mathbb{E}_Z Z = \mathbb{E}_{z_0}(-1)^{z_0} \prod_{i=1}^{n} \mathbb{E}_{z_i} \otimes Z_i^{z_i} = \frac{1}{2} \cdot (+1) \prod_{i=1}^{n} \mathbb{E}_{z_i} \otimes Z_i^{z_i} + \frac{1}{2} \cdot (-1) \prod_{i=1}^{n} \mathbb{E}_{z_i} \otimes Z_i^{z_i} = 0. \tag{5.240}$$

Now the correlation. First we let $z \cdot k = z_0 + z_1 k_1 + ... + z_n k_n$. Then

$$\begin{aligned}\mathbb{E}_Z\langle i|Z|i\rangle\langle j|Z|j\rangle &= \mathbb{E}_Z(-1)^{z\cdot i}(-1)^{z\cdot j}\\ &= \mathbb{E}_Z(-1)^{z\cdot(i+j)}\\ &= \mathbb{E}_{z_0}(-1)^{2z_0}\prod_{k=1}^{n}\mathbb{E}_{z_k}(-1)^{z_k(i_k+j_k)}\\ &= \prod_{k=1}^{n}\mathbb{E}_{z_k}(-1)^{z_k(i_k+j_k)}.\end{aligned} \tag{5.241}$$

Now we just need to compute a single one:

$$\begin{aligned}\mathbb{E}_{z_k}(-1)^{z_k(i_k+j_k)} &= \frac{1}{2}\cdot(1)+\frac{1}{2}\cdot(-1)^{i_k+j_k}\\ &= \begin{cases}1 & \text{if } i_k=0, j_k=0\\ 0 & \text{if } i_k=0, j_k=1\\ 0 & \text{if } i_k=1, j_k=0\\ 1 & \text{if } i_k=1, j_k=1\end{cases}\\ &= \delta_{i_k,j_k}.\end{aligned} \tag{5.242}$$

Then we have that the total product is

$$\mathbb{E}_Z\langle i|Z|i\rangle\langle j|Z|j\rangle = \prod_{k=1}^{n}\delta_{i_k,j_k} = \delta_{i,j}. \tag{5.243}$$

The absolute last thing I would need is that $\mathbb{E}_Z\ |\lambda_k(Z)|^3 = 1$, which is trivial.

- The first moment calculation, Theorem 3.1 but that follows from $\mathbb{E}_R\Lambda_R = 0$.
- The heisenberg evolution Lemma 2.3
- The sandwiched evolution Lemma 2.4
- The remainder bound Theorem 3.1.

Theorem 1.1 (First Order Φ): *Let Φ be the thermalizing quantum channel given by Equation (3.100) and G the randomly chosen interaction term as given by Equation (3.97). The $O(\alpha)$ term in the weak-coupling expansion in Equation (3.109) vanishes*

$$\frac{\partial}{\partial\alpha}\Phi(\rho;\alpha)\ |_{\alpha=0} = 0. \tag{5.244}$$

Proof: We start by moving the α derivative through the linear operations of partial tracing and integrals so that it can act on the fixed interaction map Φ_G

$$\frac{\partial}{\partial\alpha}\Phi(\rho)\mid_{\alpha=0} = \frac{\partial}{\partial\alpha}\,\mathrm{tr}_{\mathcal{H}_E}\left(\int \Phi_G(\rho)dG\right)\mid_{\alpha=0}$$
$$= \mathrm{tr}_{\mathcal{H}_E}\left(\int \frac{\partial}{\partial\alpha}\Phi_{G(\rho)}dG\mid_{\alpha=0}\right). \tag{5.245}$$

Now we use the expression for Φ_G in Eq. Equation (3.101) to compute the derivatives,

$$\frac{\partial}{\partial\alpha}\Phi_G(\rho) = \left(\frac{\partial}{\partial\alpha}e^{+i(H+\alpha G)t}\right)\rho\otimes\rho_E e^{-i(H+\alpha G)t} + e^{+i(H+\alpha G)t}\rho\otimes\rho_E\left(\frac{\partial}{\partial\alpha}e^{-i(H+\alpha G)t}\right)$$
$$= \left(\int_0^1 e^{is(H+\alpha G)t}(itG)e^{i(1-s)(H+\alpha G)t}ds\right)\rho\otimes\rho_E e^{-i(H+\alpha G)t}$$
$$+e^{i(H+\alpha G)t}\rho\otimes\rho_E\left(\int_0^1 e^{-is(H+\alpha G)t}(-itG)e^{-i(1-s)(H+\alpha G)t}ds\right). \tag{5.246}$$

We can set $\alpha = 0$ in the above and introduce the expectation over G that will be required

$$\mathbb{E}_G\left[\frac{\partial}{\partial\alpha}\Phi_{G(\rho)}\mid_{\alpha=0}\right] = it\mathbb{E}_G\int_0^1 e^{isHt}Ge^{-isHt}dse^{iHt}\rho\otimes\rho_E e^{-iHt}$$
$$-ite^{+iHt}\rho\otimes\rho_E\mathbb{E}_G\int_0^1 e^{-isHt}Ge^{-i(1-s)Ht}ds$$
$$= it\int_0^1 e^{isHt}\mathbb{E}_G[G]e^{-isHt}dse^{iHt}\rho\otimes\rho_E e^{-iHt}$$
$$-ite^{+iHt}\rho\otimes\rho_E\int_0^1 e^{-isHt}\mathbb{E}_G[G]e^{-i(1-s)Ht}ds. \tag{5.247}$$

Since our eigenvalues, D_{ii}, are mean zero ($\mathbb{E}_D D = 0$) we can compute $\mathbb{E}_G[G]$ and arrive at the lemma statement

$$\mathbb{E}_G[G] = \mathbb{E}_{\text{haar}}\mathbb{E}_D\left[U_G D U_G^\dagger\right] = \mathbb{E}_{\text{haar}}\left[U_G\mathbb{E}_D[D]U_G^\dagger\right] = 0. \tag{5.248}$$

□

Ok so this proof follows trivially. Now the next one might be harder, but it actually also just follows from the fact that $\mathbb{E}_R\lambda_R(i)\lambda_R(j) = \delta_{i,j}$.

Lemma 1.2: *Let $G(t)$ denote the Heisenberg evolved random interaction $G(t) = e^{iHt}Ge^{-iHt}$ for a total Hamiltonian H. After averaging over the interaction measure the product $G(x)G(y)$ can be computed as*

$$\int G(x)G(y)dG = \frac{1}{\dim+1}\left(\sum_{(i,j),(k,l)} e^{\Delta(i,j\mid k,l)(x-y)}\ |i,j\rangle\langle i,j| + \mathbb{1}\right). \tag{5.249}$$

Proof: The overall structure of this proof is to evaluate the product in the Hamiltonian eigenbasis and split the product into three factors: a phase contribution from the time evolution, a Haar integral from the eigenvalues of the random interaction, and the eigenvalue contribution of the random interaction. Since this involves the use of multiple indices, it will greatly simplify the proof to use a single index over the total Hilbert space $\mathcal{H}$ as opposed to two indices over $\mathcal{H}_S \otimes \mathcal{H}_E$. For example, the index a should be thought of as a pair (a_s, a_e), and functions $\lambda(a)$ should be thought of as $\lambda(a_s, a_e)$. Once the final form of the expression is reached we will substitute in pairs of indices for easier use of the lemma in other places.

$$\begin{aligned}
\int G(x)G(y)dG &= \int e^{+iHx}U_G D U_G^\dagger e^{-iHx} e^{+iHy} U_G D U_G^\dagger e^{-iHy} dU_G dD \\
&= \int \Bigg[\sum_a e^{+i\lambda(a)x}|a\rangle\langle a|U_G \sum_b D(b)|b\rangle\langle b|U_G^\dagger \\
&\quad \sum_c e^{-i\lambda(c)(x-y)}|c\rangle\langle c|U_G \sum_d D(d)|d\rangle\langle d|U_G^\dagger \sum_e e^{-i\lambda(e)y}|e\rangle\langle e|\Bigg] dU_G dD \\
&= \sum_{a,b,c,d,e} |a\rangle\langle e| e^{-i(\lambda(c)-\lambda(a))x} e^{-i(\lambda(e)-\lambda(c))y} \\
&\quad \times \int \langle a|U_G|b\rangle\langle c|U_G|d\rangle\langle b|U_G^\dagger|c\rangle\langle d|U_G^\dagger|e\rangle dU_G \int D(b)D(d)dD \\
&= \sum_{a,b,c,d,e} \delta_{bd}|a\rangle\langle e| e^{-i(\lambda(c)-\lambda(a))x} e^{-i(\lambda(e)-\lambda(c))y} \\
&\quad \times \int \langle a|U_G|b\rangle\langle c|U_G|d\rangle\langle b|U_G^\dagger|c\rangle\langle d|U_G^\dagger|e\rangle dU_G.
\end{aligned} \tag{5.250}$$

Now the summation over d fixes $d = b$ and we use Lemma 2.2 to compute the Haar integral, which simplifies greatly due to the repeated b index. Plugging the result into the above yields the following

$$= \frac{1}{\dim^2 - 1} \sum_{a,b,c,e} |a\rangle\langle e| e^{-i(\lambda(c)-\lambda(a))x} e^{-i(\lambda(e)-\lambda(c))y} \left(\delta_{ac}\delta_{ce} + \delta_{ae} - \frac{1}{\dim}(\delta_{ac}\delta_{ce} + \delta_{ae}) \right)$$

$$= \frac{1}{\dim^2 - 1} \left(1 - \frac{1}{\dim} \right) \sum_{a,b,c,e} |a\rangle\langle e| e^{-i(\lambda(c)-\lambda(a))x} e^{-i(\lambda(e)-\lambda(c))y} \delta_{ae}(1 + \delta_{ac})$$

$$= \frac{1}{\dim^2 - 1} \left(1 - \frac{1}{\dim} \right) \sum_{a,b,c} |a\rangle\langle a| e^{i(\lambda(a)-\lambda(c))(x-y)} (1 + \delta_{ac})$$

$$= \frac{1(\dim - 1)}{\dim^2 - 1} \sum_{a,c} |a\rangle\langle a| e^{i(\lambda(a)-\lambda(c))(x-y)} (1 + \delta_{ac})$$

$$= \frac{1}{\dim + 1} \left(\sum_{a,c} e^{i(\lambda(a)-\lambda(c))(x-y)} |a\rangle\langle a| + \mathbb{1} \right). \tag{5.251}$$

Reindexing by $a \mapsto i, j$, $c \mapsto k, l$, and plugging in the definition of Δ yields the statement of the lemma. □

this is the second to last one

Lemma 1.3: *Given two Heisenberg evolved random interactions $G(x)$ and $G(y)$ we can compute their action on the outer product $|i,j\rangle\langle k,l|$ as*

$$\int G(x)\ |i,j\rangle\langle k,l|\ G(y) dG$$

$$= \frac{1}{\dim + 1} \left(|i,j\rangle\langle k,l| + \langle i,j|k,l\rangle \sum_{m,n} e^{\Delta(m,n\ |\ i,j)(x-y)}\ |m,n\rangle\langle m,n| \right). \tag{5.252}$$

Proof: This proof is structured the same as Lemma 2.3 and similarly we will use a single index of the total Hilbert space $\mathcal{H}$ and switch to two indices to match the rest of the exposition.

$$\begin{aligned}
\int G(x)|a\rangle\langle b|G(y)dG &= \int e^{iHx}U_G D U_G^\dagger e^{-iHx}|a\rangle\langle b|e^{iHy}U_G D U_G^\dagger e^{-iHy}dG \\
&= \sum_{c,d,e,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y} \\
&\quad \times \int |c\rangle\langle c|U_G D(d)|d\rangle\langle d|U_G^\dagger|a\rangle\langle b|U_G D(e)|e\rangle\langle e|U_G^\dagger|f\rangle\langle f|dG \\
&= \sum_{c,d,e,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y}|c\rangle\langle f| \\
&\quad \times \int D(d)D(e)dD \int \langle c|U_G|d\rangle\langle b|U_G|e\rangle\langle d|U_G^\dagger|a\rangle\langle e|U_G^\dagger|f\rangle dU_G \\
&= \sum_{c,d,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y}|c\rangle\langle f| \\
&\quad \times \int \langle c|U_G|d\rangle\langle b|U_G|d\rangle\langle a|\overline{U_G}|d\rangle\langle f|\overline{U_G}|d\rangle dU_G \\
&= \frac{1}{\dim^2-1}\sum_{c,d,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y}|c\rangle\langle f|\big(\delta_{ca}\delta_{bf}+\delta_{cf}\delta_{ab}\big)\Big(1-\frac{1}{\dim}\Big) \\
&= \frac{1}{\dim+1}\sum_{c,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y}|c\rangle\langle f|\big(\delta_{ca}\delta_{bf}+\delta_{cf}\delta_{ab}\big) \\
&= \frac{1}{\dim+1}\left(|a\rangle\langle b| + \delta_{ab}\sum_c e^{i(\lambda(c)-\lambda(a))(x-y)}|c\rangle\langle c|\right). \qquad (5.253)
\end{aligned}$$

Now re-indexing by $a \mapsto (i,j)$, $b \mapsto (k,l)$ and $c \mapsto (m,n)$ results in the expression given in the statement of the lemma. □

So all four of these results follow trivially from I.I.D and that the absolute value of the third moment of the Rademacher distribution is bounded.

Now I can get to the really difficult thing, which is showing that random Z rotations can approximate the rotations.

Need to come up with a circuit for $e^{iG\alpha t}$. Call $\theta = \alpha t$ for fun right now. Then we do the following

$$\begin{aligned}
e^{iU\Lambda_R U^\dagger\theta} &= U e^{i\Lambda\theta}U^\dagger \\
&= U\left(\sum_{i=1}^{\dim} e^{i\lambda_R(i)\theta}\right)U^\dagger \qquad (5.254)
\end{aligned}$$

Actually at the end of the day I need to approximate the channel $\mathbb{E}_G e^{i(H+\alpha G)t}\rho e^{-i(H+\alpha G)t}$ using a composite simulation. We let H be simulated via a Trotter formula and G a QDrift expression. But that would yield:

$$\begin{aligned}\mathbb{E}_G e^{i(H+\alpha G)t}\rho e^{-i(H+\alpha G)t} &= \mathbb{E}_G e^{iG\theta}\rho e^{-iG\theta} \\ &= \mathbb{E}_U \mathbb{E}_R U e^{i\Lambda_R\theta}U^\dagger \rho U e^{-i\Lambda_R\theta}U^\dagger. \end{aligned} \tag{5.255}$$

Letting C denote the Clifford distribution and let Z denote a randomly chosen string of Z pauli strings then I think

$$\mathbb{E}_U \mathbb{E}_R U e^{i\Lambda_R\theta}U^\dagger \rho U e^{-i\Lambda_R\theta}U^\dagger = \mathbb{E}_C \mathbb{E}_Z C e^{iZ\theta}C^\dagger \rho C e^{-iZ\theta}C^\dagger \tag{5.256}$$

The Clifford equality follows from the 3-design property. I will need a new proof for the Z argument. I'll try the following

$$\begin{aligned}\mathbb{E}_R \langle i|e^{i\Lambda_R\theta}\ |j\rangle\langle k|\ e^{-i\Lambda_R\theta}|l\rangle &= \mathbb{E}_R \sum_{a,b} e^{i(\lambda_R(a)-\lambda_R(b)\theta)}\langle i|a\rangle\langle a|j\rangle\langle k|b\rangle\langle b|l\rangle \\ &= \mathbb{E}_R \delta_{i,j}\delta_{k,l} e^{i(\lambda_R(j)-\lambda_R(k))\theta} \\ &= \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot e^{i\theta} + \frac{1}{4}\cdot e^{-i\theta} \\ &= \frac{1}{2}(1+\cos(\theta)). \end{aligned} \tag{5.257}$$

Actually the above only holds for $j \neq k$, for $j = k$ then it is just 1. Now to show the Z expectation. For this let $w(k)$ denote the Hamming weight of the computational state k. No we actually need the agreement between the bitstring k and the bitstring z, which has a 0 if the sampled value for Z at that site is I and a 1 if the sampled value is Z. AKA $Z = Z_1^{z_1} \otimes Z_2^{z_2} \otimes ... \otimes Z_n^{z_n}$, where each z_i is a Bernoulli random variable.

Wait, why not just use $G \sim U_G Z U_G^\dagger$?

Appendix B

Haar Integrals

Lemma 2.1 (Sinc Function Bounds): *For* $\text{sinc}^2(x\frac{t}{2})$ *and* $\delta_{\min}$ *as defined in Equation (3.96), we will make significant use of the following Bounds*

$$|x| \geq \delta_{\min} \Longrightarrow \text{sinc}^2\left(\frac{xt}{2}\right) \leq \frac{4}{\delta_{\min}^2 t^2}$$
$$|x| \leq \frac{\sqrt{2}}{t} \Longrightarrow \text{sinc}^2\left(\frac{xt}{2}\right) \geq 1 - \frac{|x|^2\ t^2}{2}. \tag{6.258}$$

Proof: The first inequality is rather trivial

$$\text{sinc}^2\left(\frac{xt}{2}\right) = \frac{\sin^2\left(\frac{xt}{2}\right)}{\left(\frac{xt}{2}\right)^2} \leq \frac{4}{x^2t^2} \leq \frac{4}{\delta_{\min}^2 t^2}. \tag{6.259}$$

The second involves a Taylor Series for sinc^2, which we compute using the expression of sinc as $\text{sinc}(\frac{xt}{2}) = \frac{\sin\frac{xt}{2}}{\frac{xt}{2}} = \int_0^1 \cos(s\frac{xt}{2})ds$. The first two derivatives can then be computed easily

$$\frac{d\text{sinc}^2\left(\frac{xt}{2}\right)}{dx} = -t\int_0^1 \sin(sx)s\ ds \int_0^1 \cos(sx)ds$$
$$\frac{d^2\text{sinc}^2\left(\frac{xt}{2}\right)}{dx^2} = -\frac{t^2}{2}\int_0^1 \cos(sx)s^2 ds \int_0^1 \cos(sx)ds + \frac{t^2}{2}\int_0^1 \sin(sx)sds \int_0^1 \sin(sx)sds \tag{6.260}$$

We can evaluate each of these derivatives about the origin using continuity of the derivatives along with the limits $\lim_{x\to 0}\cos(sx) = 1$ and $\lim_{x\to 0}\sin(sx) = 0$. We can now compute the mean-value version Taylor series as

$$\text{sinc}^2\left(\frac{xt}{2}\right) = \text{sinc}^2(0) + x\frac{d}{dx}\text{sinc}^2\left(\frac{xt}{2}\right)\,|_{x=0} + \frac{x^2}{2!}\frac{d^2}{dx^2}\text{sinc}^2\left(\frac{xt}{2}\right)\,|_{x=x_\star}, \tag{6.261}$$

where $x_\star \in [0,1]$. Plugging in $\text{sinc}^2(0) = 1$ and $\frac{d\text{sinc}^2(x\frac{t}{2})}{dx}|_{x=0} = 0$ then yields

$$\left|\text{sinc}^2\left(\frac{xt}{2}\right) - 1\right| = \frac{|x|^2}{2}\left|\frac{d^2\text{sinc}^2(x\frac{t}{2})}{dx^2}\,|_{x=x_\star}\right|. \tag{6.262}$$

We make use of the rather simplistic bound

$$\begin{aligned}
\left|\frac{d^2\text{sinc}^2(sx\frac{t}{2})}{dx^2}|_{x=x_\star}\right| &\leq \frac{t^2}{2}\left|\int_0^1 \cos\left(sx_\star\frac{t}{2}\right)s^2 ds \int_0^1 \cos\left(sx_\star\frac{t}{2}\right)ds\right| + \frac{t^2}{2}\left|\int_0^1 \sin\left(sx_\star\frac{t}{2}\right)sds \int_0^1 \sin\left(sx_\star\frac{t}{2}\right)sds\right| \\
&\leq \frac{t^2}{2}\int_0^1 \left|\cos\left(sx_\star\frac{t}{2}\right)\right| s^2 ds \int_0^1 \left|\cos\left(sx_\star\frac{t}{2}\right)\right| ds + \frac{t^2}{2}\left(\int_0^1 \left|\sin\left(sx_\star\frac{t}{2}\right)\right|\,|s|\,ds\right)^2 \\
&\leq \frac{t^2}{2}\int_0^1 s^2 ds + \frac{t^2}{2}\left(\int_0^1 sds\right)^2 \\
&\leq t^2.
\end{aligned} \tag{6.263}$$

This yields the final inequality $|\text{sinc}^2(\frac{xt}{2}) - 1| \leq \frac{|x|^2\,t^2}{2}$ which yields Equation (6.258). □

B.1 Haar Integral Proofs

In this section we present the more technical work needed to state our results in Section 3.2. Lemma 2.3 and Lemma 2.4 are used to compute the effects of the randomized interactions in a form that are usable in the main result of Lemma 3.2. Lemma 2.2 can be derived from Appendix C in [89].

Lemma 2.2: *Let $\mathbb{E}_U$ denote the expectation over the Haar measure over the set of unitary matrices acting on a* dim *dimensional Hilbert space. Then for $|i_1\rangle, |i_2\rangle, ..., |k_2\rangle$ drawn from an orthonormal basis*

$$\begin{aligned}
&\mathbb{E}_U\left[\langle i_1|U|j_1\rangle\langle i_2|U|j_2\rangle\langle k_1|U^\dagger|l_1\rangle\langle k_{12}|U^\dagger|l_2\rangle\right] \\
&= \frac{1}{\text{dim}^2-1}\left(\delta_{i_1,l_1}\delta_{j_1,k_1}\delta_{i_2,l_2}\delta_{j_2,k_2} + \delta_{i_1,l_2}\delta_{j_1,k_2}\delta_{i_2,l_1}\delta_{j_2,k_1}\right) \\
&\quad - \frac{1}{\text{dim}(\text{dim}^2-1)}\left(\delta_{i_1,l_2}\delta_{j_1,k_1}\delta_{i_2,l_1}\delta_{j_2,k_2} + \delta_{i_1,l_1}\delta_{j_1,k_2}\delta_{i_2,l_2}\delta_{j_2,k_1}\right).
\end{aligned} \tag{6.264}$$

Lemma 2.3: *Let $G(t)$ denote the Heisenberg evolved random interaction $G(t) = e^{iHt}Ge^{-iHt}$ for a total Hamiltonian H. After averaging over the interaction measure the product $G(x)G(y)$ can be computed as*

$$\mathbb{E}_G[G(x)G(y)] = \frac{1}{\dim + 1}\left(\sum_{(i,j),(k,l)} e^{\Delta(i,j\mid k,l)(x-y)}\ |i,j\rangle\langle i,j| + \mathbb{1}\right). \tag{6.265}$$

Proof: The overall structure of this proof is to evaluate the product in the Hamiltonian eigenbasis and split the product into three factors: a phase contribution from the time evolution, a Haar expectation from the eigenvectors of the random interaction, and the eigenvalue expectation of the random interaction. Since this involves the use of multiple indices, it will greatly simplify the proof to use a single index over the total Hilbert space $\mathcal{H}$ as opposed to two indices over $\mathcal{H}_S \otimes \mathcal{H}_E$. For example, the index a should be thought of as a pair (a_s, a_e), and functions $\lambda(a)$ should be thought of as $\lambda(a_s, a_e)$. Once the final form of the expression is reached we will substitute in pairs of indices for easier use of the lemma in other places.

$$\begin{aligned}
\mathbb{E}_G[G(x)G(y)] &= \mathbb{E}_{\Lambda_G}\mathbb{E}_{U_G} e^{+iHx}U_G\Lambda_G U_G^\dagger e^{-iHx}e^{+iHy}U_G\Lambda_G U_G^\dagger e^{-iHy} \\
&= \mathbb{E}_{\Lambda_G}\mathbb{E}_{U_G}\Bigg[\sum_a e^{+i\lambda(a)x}|a\rangle\langle a|U_G\sum_b \Lambda_G(b)|b\rangle\langle b|U_G^\dagger \\
&\quad \sum_c e^{-i\lambda(c)(x-y)}|c\rangle\langle c|U_G\sum_d \Lambda_G(d)|d\rangle\langle d|U_G^\dagger\sum_e e^{-i\lambda(e)y}|e\rangle\langle e|\Bigg] \\
&= \sum_{a,b,c,d,e}|a\rangle\langle e|e^{-i(\lambda(c)-\lambda(a))x}e^{-i(\lambda(e)-\lambda(c))y} \\
&\quad \times \mathbb{E}_{U_G}\left[\langle a|U_G|b\rangle\langle c|U_G|d\rangle\langle b|U_G^\dagger|c\rangle\langle d|U_G^\dagger|e\rangle\right]\mathbb{E}_{\Lambda_G}[\Lambda_G(b)\Lambda_G(d)] \\
&= \sum_{a,b,c,d,e}\delta_{bd}|a\rangle\langle e|e^{-i(\lambda(c)-\lambda(a))x}e^{-i(\lambda(e)-\lambda(c))y} \\
&\quad \times \mathbb{E}_{U_G}\left[\langle a|U_G|b\rangle\langle c|U_G|d\rangle\langle b|U_G^\dagger|c\rangle\langle d|U_G^\dagger|e\rangle\right].
\end{aligned} \tag{6.266}$$

We used the fact that the eigenvalues of G are I.I.D with variance 1 to make the substitution $\mathbb{E}_{\Lambda_G}[\Lambda_G(b)\Lambda_G(d)] = \delta_{a,b}$. This allows us to reduce the sum over d to the condition when $d = b$, which greatly simplifies the Haar expectation we have to take. Using Lemma 2.2 we have

$$\mathbb{E}_{U_G}\left[\langle a|U_G|b\rangle\langle c|U_G|b\rangle\langle b|U_G^\dagger|c\rangle\langle b|U_G^\dagger|e\rangle\right] = \frac{1}{\dim^2 - 1}\left(\delta_{ac}\delta_{ce} + \delta_{ae} - \frac{1}{\dim}(\delta_{ac}\delta_{ce} + \delta_{ae})\right). \tag{6.267}$$

and we use Lemma 2.2 to compute the Haar integral, which simplifies greatly due to the repeated b index. Plugging the result into the above yields the following

$$
\begin{aligned}
&= \frac{1}{\dim^2 - 1} \sum_{a,b,c,e} |a\rangle\langle e| e^{-i(\lambda(c)-\lambda(a))x} e^{-i(\lambda(e)-\lambda(c))y} \left(\delta_{ac}\delta_{ce} + \delta_{ae} - \frac{1}{\dim}(\delta_{ac}\delta_{ce} + \delta_{ae}) \right) \\
&= \frac{1}{\dim^2 - 1} \left(1 - \frac{1}{\dim} \right) \sum_{a,b,c,e} |a\rangle\langle e| e^{-i(\lambda(c)-\lambda(a))x} e^{-i(\lambda(e)-\lambda(c))y} \delta_{ae}(1 + \delta_{ac}) \\
&= \frac{1}{\dim^2 - 1} \left(1 - \frac{1}{\dim} \right) \sum_{a,b,c} |a\rangle\langle a| e^{i(\lambda(a)-\lambda(c))(x-y)} (1 + \delta_{ac}) \\
&= \frac{1(\dim - 1)}{\dim^2 - 1} \sum_{a,c} |a\rangle\langle a| e^{i(\lambda(a)-\lambda(c))(x-y)} (1 + \delta_{ac}) \\
&= \frac{1}{\dim + 1} \left(\sum_{a,c} e^{i(\lambda(a)-\lambda(c))(x-y)} |a\rangle\langle a| + \mathbb{1} \right).
\end{aligned} \tag{6.268}
$$

Reindexing by $a \mapsto i, j$, $c \mapsto k, l$, and plugging in the definition of Δ yields the statement of the lemma. □

Lemma 2.4: *Given two Heisenberg evolved random interactions $G(x)$ and $G(y)$ we can compute their action on the outer product $|i,j\rangle\langle k,l|$ as*

$$
\begin{aligned}
&\mathbb{E}_G[G(x)\ |i,j\rangle\langle k,l|\ G(y) dG] \\
&= \frac{1}{\dim + 1} \left(|i,j\rangle\langle k,l| + \langle i,j|k,l\rangle \sum_{m,n} e^{\Delta(m,n\ |\ i,j)(x-y)}\ |m,n\rangle\langle m,n| \right).
\end{aligned} \tag{6.269}
$$

Proof: This proof is structured the same as Lemma 2.3 and similarly we will use a single index of the total Hilbert space $\mathcal{H}$ and switch to two indices to match the rest of the exposition.

$$
\begin{aligned}
\mathbb{E}_G[G(x)|a\rangle\langle b|G(y)] &= \mathbb{E}_G\left[e^{iHx}U_G D U_G^\dagger e^{-iHx}|a\rangle\langle b|e^{iHy}U_G D U_G^\dagger e^{-iHy}\right] \\
&= \sum_{c,d,e,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y} \\
&\quad \times \mathbb{E}_G\left[|c\rangle\langle c|U_G D(d)|d\rangle\langle d|U_G^\dagger|a\rangle\langle b|U_G D(e)|e\rangle\langle e|U_G^\dagger|f\rangle\langle f|\right] \\
&= \sum_{c,d,e,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y}|c\rangle\langle f| \\
&\quad \times \mathbb{E}_{\Lambda_G}[\Lambda_G(d)\Lambda_G(e)]\mathbb{E}_{U_G}\left[\langle c|U_G|d\rangle\langle b|U_G|e\rangle\langle d|U_G^\dagger|a\rangle\langle e|U_G^\dagger|f\rangle\right] \\
&= \sum_{c,d,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y}|c\rangle\langle f| \\
&\quad \times \mathbb{E}_{U_G}\left[\langle c|U_G|d\rangle\langle b|U_G|d\rangle\langle a|\overline{U_G}|d\rangle\langle f|\overline{U_G}|d\rangle\right] \\
&= \frac{1}{\dim^2-1}\sum_{c,d,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y}|c\rangle\langle f|\left(\delta_{ca}\delta_{bf}+\delta_{cf}\delta_{ab}\right)\left(1-\frac{1}{\dim}\right) \\
&= \frac{1}{\dim+1}\sum_{c,f} e^{i(\lambda(c)-\lambda(a))x}e^{i(\lambda(b)-\lambda(f))y}|c\rangle\langle f|\left(\delta_{ca}\delta_{bf}+\delta_{cf}\delta_{ab}\right) \\
&= \frac{1}{\dim+1}\left(|a\rangle\langle b| + \delta_{ab}\sum_c e^{i(\lambda(c)-\lambda(a))(x-y)}|c\rangle\langle c|\right).
\end{aligned} \tag{6.270}
$$

We used the fact that $\mathbb{E}_{\Lambda_G}[\Lambda_G(d)\Lambda_G(e)] = \delta_{d,e}$ to eliminate the sum over e. Re-indexing by $a \mapsto (i,j)$, $b \mapsto (k,l)$ and $c \mapsto (m,n)$ results in the expression given in the statement of the lemma.□

TODO: Need to update the lemma number for this to essentially be a manual "restatable."

Lemma 2.5: *Given a system Hamiltonian H_S, an environment Hamiltonian H_E, a simulation time t, and coupling coefficient α, let Φ_G denote the time evolution channel under a fixed interaction term G as given in Equation (3.101), let χ denote the following coherence prefactor*

$$\chi(i,j) := \sum_{a,b:\Delta(i,j,|a,b)\neq 0} \frac{1 - i\Delta(i,j|a,b)t - e^{-i\Delta(i,j|a,b)t}}{\Delta(i,j|a,b)^2}, \tag{6.271}$$

and let $\eta(i,j)$ denote the degeneracy of the $(i,j)^{\text{th}}$ eigenvalue of $H = H_S + H_E$. Then the $O(\alpha^2)$ term of Φ_G in a weak-coupling expansion is given by

$$\begin{aligned}
&\frac{\alpha^2}{2}\mathbb{E}_G\left[\frac{\partial^2}{\partial\alpha^2}\Phi_G(|i,j\rangle\langle k,l|)|_{\alpha=0}\right] \\
&= -\frac{\alpha^2 e^{i\Delta(i,j|k,l)t}}{\dim + 1}\left(\chi(i,j) + \chi(k,l)^* + \frac{t^2}{2}(\eta(i,j)+\eta(k,l))\right)|i,j\rangle\langle k,l| \\
&\quad + \langle i,j|k,l\rangle\frac{\alpha^2 t^2}{\dim+1}\sum_{a,b}\operatorname{sinc}^2\left(\Delta(i,j|a,b)\frac{t}{2}\right)|a,b\rangle\langle a,b|
\end{aligned} \tag{6.272}$$

For $|i,j\rangle = |k,l\rangle$ the above expression simplifies to

$$\begin{aligned}
&\frac{\alpha^2}{2}\mathbb{E}_G\left[\frac{\partial^2}{\partial\alpha^2}\Phi_G(|i,j\rangle\langle i,j|)|_{\alpha=0}\right] \\
&= \tilde{\alpha}^2\sum_{(a,b)\neq(i,j)}\operatorname{sinc}^2\left(\Delta(i,j|a,b)\frac{t}{2}\right)(|a,b\rangle\langle a,b| - |i,j\rangle\langle i,j|)
\end{aligned} \tag{6.273}$$

which also demonstrates that $\operatorname{tr}\mathcal{T}(\rho) = 0$ for ρ such that $[\rho, H_S] = 0$.

Proof: To start we would like to note that we will use a single index notation to refer to the joint system-environment eigenbasis during this proof to help shorten the already lengthy expressions. We will convert back to a double index notation to match the statement of the theorem. We start from the expression for the first derivative of the channel $\frac{\partial}{\partial\alpha}\Phi_G(\rho_S)$ given by Equation (3.114). To take the second derivative there are six factors involving α, so we will end up with six terms. We repeat Equation (3.114) below, add a derivative, and label each factor containing an α for easier computation

$$\frac{\partial^2}{\partial\alpha^2}\Phi_G(\rho_S) = \frac{\partial}{\partial\alpha}\left(\int_0^1 \underbrace{e^{is(H+\alpha G)t}}_{\text{(A)}}(itG)\underbrace{e^{i(1-s)(H+\alpha G)t}}_{\text{(B)}}ds\ \rho\ \underbrace{e^{-i(H+\alpha G)t}}_{\text{(C)}}\right)$$
$$+\frac{\partial}{\partial\alpha}\left(\underbrace{e^{i(H+\alpha G)t}}_{\text{(D)}}\ \rho\ \int_0^1 \underbrace{e^{-is(H+\alpha G)t}}_{\text{(E)}}(-itG)\underbrace{e^{-i(1-s)(H+\alpha G)t}}_{\text{(F)}}ds\right). \quad (6.274)$$

Our goal is to get each of these terms in a form in which we can use either Lemma 2.3 or Lemma 2.4.

$$(A) = it\int_0^1\left(\frac{\partial}{\partial\alpha}e^{is_1(H+\alpha G)t}\right)Ge^{i(1-s_1)(H+\alpha G)t}ds_1\rho e^{-i(H+\alpha G)t}\ |_{\alpha=0}$$
$$= (it)^2\int_0^1\left(\int_0^1 e^{is_1s_2(H+\alpha G)t}s_1Ge^{is_1(1-s_2)(H+\alpha G)t}ds_2\right)Ge^{i(1-s_1)(H+\alpha G)t}ds_1\rho e^{-i(H+\alpha G)t}\ |_{\alpha=0}$$
$$= -t^2\int_0^1\int_0^1 e^{is_1s_2Ht}Ge^{-is_1s_2Ht}e^{is_1Ht}Ge^{-is_1Ht}s_1ds_1ds_2e^{iHt}\rho e^{-iHt}$$
$$= -t^2\int_0^1\int_0^1 G(s_1s_2t)G(s_1t)s_1ds_1ds_2\rho(t). \quad (6.275)$$

$$(B) = it\int_0^1 e^{is_1(H+\alpha G)t}G\frac{\partial}{\partial\alpha}\left(e^{i(1-s_1)(H+\alpha G)t}\right)ds_1\rho e^{-i(H+\alpha G)t}\ |_{\alpha=0}$$
$$= (it)^2\int_0^1 e^{is_1(H+\alpha G)t}G\left(\int_0^1 e^{i(1-s_1)s_2(H+\alpha G)t}(1-s_1)Ge^{i(1-s_1)(1-s_2)(H+\alpha G)t}ds_2\right)ds_1\ \rho e^{-i(H+\alpha G)t}\ |_{\alpha=0}$$
$$= -t^2\int_0^1\int_0^1 e^{is_1Ht}Ge^{i(1-s_1)s_2Ht}Ge^{i(1-s_1)(1-s_2)Ht}(1-s_1)ds_1ds_2\ \rho e^{-iHt}$$
$$= -t^2\int_0^1\int_0^1 e^{is_1Ht}Ge^{-is_1Ht}e^{i(s_1+s_2-s_1s_2)Ht}Ge^{-i(s_1+s_2-s_1s_2)Ht}(1-s_1)ds_1ds_2\ \rho(t)$$
$$= -t^2\int_0^1\int_0^1 G(s_1t)G((s_1+s_2-s_1s_2)t)(1-s_1)ds_1ds_2\ \rho(t) \quad (6.276)$$

$$(C) = it\int_0^1 e^{is(H+\alpha G)t}Ge^{i(1-s)(H+\alpha G)t}ds\ \rho\ \frac{\partial}{\partial\alpha}\left(e^{-i(H+\alpha G)t}\right)\ |_{\alpha=0}$$
$$= (it)(-it)\int_0^1 e^{is(H+\alpha G)t}Ge^{i(1-s)(H+\alpha G)t}ds\ \rho\ \left(\int_0^1 e^{-is(H+\alpha G)t}Ge^{-i(1-s)(H+\alpha G)t}ds\right)|_{\alpha=0}$$
$$= +t^2\left(\int_0^1 e^{isHt}Ge^{-isHt}ds\right)e^{iHt}\rho e^{-iHt}\left(\int_0^1 e^{i(1-s)Ht}Ge^{-i(1-s)Ht}ds\right)$$
$$= +t^2\int_0^1 G(st)ds\ \rho(t)\int_0^1 G((1-s)t)ds \quad (6.277)$$

$$
\begin{aligned}
(D) &= (-it)\frac{\partial}{\partial\alpha}\big(e^{i(H+\alpha G)t}\big)\rho\int_0^1 e^{-is(H+\alpha G)t}Ge^{-i(1-s)(H+\alpha G)t}ds\ |_{\alpha=0}\\
&= t^2\left(\int_0^1 e^{is(H+\alpha G)t}Ge^{i(1-s)(H+\alpha G)t}ds\right)\rho\int_0^1 e^{-is(H+\alpha G)t}Ge^{-i(1-s)(H+\alpha G)t}ds\ |_{\alpha=0}\\
&= t^2\int_0^1 e^{isHt}Ge^{-isHt}ds\ \rho(t)\int_0^1 e^{i(1-s)Ht}Ge^{-i(1-s)Ht}ds\\
&= t^2\int_0^1 G(st)ds\ \rho(t)\ \int_0^1 G((1-s)t)ds
\end{aligned}
\tag{6.278}
$$

$$
\begin{aligned}
(E) &= (-it)e^{i(H+\alpha G)t}\ \rho\ \int_0^1\frac{\partial}{\partial\alpha}\big(e^{-is_1(H+\alpha G)t}\big)Ge^{-i(1-s_1)(H+\alpha G)t}ds_1\ |_{\alpha=0}\\
&= -t^2e^{i(H+\alpha G)t}\ \rho\ \int_0^1\left(\int_0^1 e^{-is_1s_2(H+\alpha G)t}(s_1G)e^{-is_1(1-s_2)(H+\alpha G)t}ds_2\right)Ge^{-i(1-s_1)(H+\alpha G)t}ds_1\ |_{\alpha=0}\\
&= -t^2e^{iHt}\rho e^{-iHt}\int_0^1\int_0^1 e^{i(1-s_1s_2)Ht}Ge^{-i(s_1-s_1s_2)Ht}Ge^{-i(1-s_1)Ht}s_1ds_1ds_2\\
&= -t^2\rho(t)\int_0^1\int_0^1 G((1-s_1s_2)t)G((1-s_1)t)s_1ds_1ds_2
\end{aligned}
\tag{6.279}
$$

$$
\begin{aligned}
(F) &= (-it)e^{i(H+\alpha G)t}\rho\int_0^1 e^{-is_1(H+\alpha G)t}G\frac{\partial}{\partial\alpha}\big(e^{-i(1-s_1)(H+\alpha G)t}\big)ds_1\ |_{\alpha=0}\\
&= (-it)^2e^{i(H+\alpha G)t}\rho\int_0^1 e^{-is_1(H+\alpha G)t}G\left(\int_0^1 e^{-i(1-s_1)s_2(H+\alpha G)t}(1-s_1)Ge^{-i(1-s_1)(1-s_2)(H+\alpha G)t}ds_2\right)ds_1\ |_{\alpha=0}\\
&= -t^2e^{-iHt}\rho e^{-iHt}\int_0^1\int_0^1 e^{i(1-s_1)Ht}Ge^{-i(1-s_1)Ht}e^{i(1-s_1)(1-s_2)Ht}Ge^{-i(1-s_1)(1-s_2)Ht}(1-s_1)ds_1ds_2\\
&= -t^2\rho(t)\int_0^1\int_0^1 G((1-s_1)t)G((1-s_1)(1-s_2)t)(1-s_1)ds_1ds_2
\end{aligned}
\tag{6.280}
$$

Now our goal is to compute the effects of averaging over the interaction G on the above terms, starting with (A). As this involves a lot of index manipulations, similarly to the proofs of Lemmas Lemma 2.3 and Lemma 2.4 we will use a single index for the total system-environment Hilbert space and switch back to a double index to state the results. We will make heavy use of Lemma Lemma 2.3.

$$\begin{aligned}
\mathbb{E}_G(A) &= -t^2 \int_0^1 \int_0^1 \mathbb{E}_G[G(s_1 s_2 t) G(s_1 t)] s_1 ds_1 ds_2 \rho(t) \\
&= \frac{-t^2}{\dim + 1} \int_0^1 \int_0^1 \left(\sum_{i,j} e^{i(\lambda(i)-\lambda(j))(s_1 s_2 t - s_1 t)} |i\rangle\langle i| + \mathbb{1} \right) s_1 ds_1 ds_2 \rho(t) \\
&= \frac{-t^2}{\dim + 1} \left(\sum_i \sum_{j:\lambda(i)\neq\lambda(j)} \int_0^1 \int_0^1 e^{i(\lambda(i)-\lambda(j))t(s_1 s_2 - s_1)} s_1 ds_1 ds_2 |i\rangle\langle i| + \sum_i \sum_{j:\lambda(i)=\lambda(j)} \frac{1}{2}|i\rangle\langle i| + \frac{1}{2}\mathbb{1} \right) \rho(t) \\
&= \frac{-t^2}{\dim + 1} \left(\sum_i \sum_{j:\lambda(i)\neq\lambda(j)} \frac{1 - i(\lambda(i)-\lambda(j))t - e^{-i(\lambda(i)-\lambda(j))t}}{t^2(\lambda(i)-\lambda(j))^2} |i\rangle\langle i| + \frac{1}{2}\sum_i (\eta(i)+1)|i\rangle\langle i| \right) \rho(t) \\
&= \frac{-1}{\dim + 1} \left(\sum_i \sum_{j:\Delta_{ij}\neq 0} \frac{1 - i\Delta_{ij}t - e^{-i\Delta_{ij}t}}{\Delta_{ij}^2} |i\rangle\langle i| + \frac{t^2}{2}\sum_i (\eta(i)+1)|i\rangle\langle i| \right) \rho(t) \qquad (6.281)
\end{aligned}$$

We can similarly compute the averaged (B) term:

$$\begin{aligned}
\mathbb{E}_G(B) &= -t^2 \int_0^1 \int_0^1 \mathbb{E}_G[G(s_1 t) G((s_1 + s_2 - s_1 s_2)t)](1 - s_1) ds_1 ds_2 \ \rho(t) \\
&= \frac{-t^2}{\dim + 1} \int_0^1 \int_0^1 \left(\sum_{i,j} e^{i(\lambda(i)-\lambda(j))(s_1 s_2 - s_2)t} |i\rangle\langle i| + \mathbb{1} \right) (1 - s_1) ds_1 ds_2 \ \rho \\
&= \frac{-t^2}{\dim + 1} \left(\sum_i \sum_{j:\lambda(i)\neq\lambda(j)} \int_0^1 \int_0^1 e^{(i(\lambda(i)-\lambda(j))t(s_1 s_2 - s_2)\}} (1 - s_1) ds_1 ds_2 |i\rangle\langle i| + \sum_i \sum_{j:\lambda(i)=\lambda(j)} \frac{1}{2}|i\rangle\langle i| + \frac{1}{2}\mathbb{1} \right) \rho(t) \\
&= \frac{-t^2}{\dim + 1} \left(\sum_i \sum_{j:\lambda(i)\neq\lambda(j)} \frac{1 - i(\lambda(i)-\lambda(j))t - e^{-i(\lambda(i)-\lambda(j))t}}{t^2(\lambda(i)-\lambda(j))^2} |i\rangle\langle i| + \frac{1}{2}\sum_i (\eta(i)+1)|i\rangle\langle i| \right) \rho(t) \\
&= \frac{-1}{\dim + 1} \left(\sum_i \sum_{j:\Delta_{ij}\neq 0} \frac{1 - i\Delta_{ij}t - e^{-i\Delta_{ij}t}}{\Delta_{ij}^2} |i\rangle\langle i| + \frac{t^2}{2}\sum_i (\eta(i)+1)|i\rangle\langle i| \right) \rho(t), \qquad (6.282)
\end{aligned}$$

which we note is identical to $\mathbb{E}_G(A)$. As terms (C) and (D) involve a different method of computation we skip them for now and compute (E) and (F).

$$\begin{aligned}
\mathbb{E}_G(E) &= -t^2\rho(t)\int_0^1\int_0^1 \mathbb{E}_G[G((1-s_1s_2)t)G((1-s_1)t)]s_1ds_1ds_2 \\
&= \frac{-t^2}{\dim+1}\rho(t)\int_0^1\int_0^1\left(\sum_{i,j} e^{i(\lambda(i)-\lambda(j))t(s_1-s_1s_2)}|i\rangle\langle i| + \mathbb{1}\right)s_1ds_1ds_2 \\
&= \frac{-t^2}{\dim+1}\rho(t)\left(\sum_i\sum_{j:\lambda(i)\neq\lambda(j)}\frac{1+i(\lambda(i)-\lambda(j))t-e^{i(\lambda(i)-\lambda(j))t}}{t^2(\lambda(i)-\lambda(j))^2}|i\rangle\langle i| + \frac{1}{2}\sum_i(\eta(i)+1)|i\rangle\langle i|\right) \\
&= \frac{-1}{\dim+1}\rho(t)\left(\sum_i\sum_{j:(\Delta_{ij}\neq 0)}\frac{1+i\Delta_{ij}t-e^{i\Delta_{ij}t}}{\Delta_{ij}^2}|i\rangle\langle i| + \frac{t^2}{2}\sum_i(\eta(i)+1)|i\rangle\langle i|\right). \qquad (6.283)
\end{aligned}$$

Computing (F) yields

$$\begin{aligned}
\mathbb{E}_G(F) &= -t^2\rho(t)\int_0^1\int_0^1 \mathbb{E}_G[G((1-s_1)t)G((1-s_1)(1-s_2)t)](1-s_1)ds_1ds_2 \\
&= \frac{-t^2\sigma^2}{\dim+1}\rho(t)\int_0^1\int_0^1\left(\sum_{i,j} e^{i(\lambda(i)-\lambda(j))t(s_2-s_1s_2)}|i\rangle\langle i| + \mathbb{1}\right)(1-s_1)ds_1ds_2 \\
&= \frac{-t^2}{\dim+1}\rho(t)\left(\sum_i\sum_{j:\lambda(i)\neq\lambda(j)}\frac{1+i(\lambda(i)-\lambda(j))t-e^{i(\lambda(i)-\lambda(j))t}}{t^2(\lambda(i)-\lambda(j))^2}|i\rangle\langle i| + \frac{1}{2}\sum_i(\eta(i)+1)|i\rangle\langle i|\right) \\
&= \frac{-1}{\dim+1}\rho(t)\left(\sum_i\sum_{j:(\Delta_{ij}\neq 0)}\frac{1+i\Delta_{ij}t-e^{i\Delta_{ij}t}}{\Delta_{ij}^2}|i\rangle\langle i| + \frac{t^2}{2}\sum_i(\eta(i)+1)|i\rangle\langle i|\right) \qquad (6.284)
\end{aligned}$$

which is identical to $\int(E)dG$.

The last two terms $(C) = (D)$ are computed as follows:

$$\begin{aligned}
\mathbb{E}_G(C) &= t^2\int_0^1\int_0^1 \mathbb{E}_G[G(s_1t)\rho(t)G((1-s_2)t)]\ ds_1ds_2 \qquad (6.285) \\
&= t^2\sum_{i,j}\rho_{ij}e^{i(\lambda(i)-\lambda(j))t}\int_0^1\int_0^1 \mathbb{E}_G[G(s_1t)|i\rangle\langle j|G((1-s_2)t)]\ ds_1ds_2 \\
&= \frac{t^2}{\dim+1}\sum_{i,j}\rho_{ij}e^{i(\lambda(i)-\lambda(j))t}\left(|i\rangle\langle j| + \delta_{ij}\sum_a\int_0^1\int_0^1 e^{i(\lambda(a)-\lambda(i))(s_1+s_2-1)t}ds_1ds_2|a\rangle\langle a|\right) \\
&= \frac{t^2}{\dim+1}\sum_{i,j}\rho_{ij}e^{i\Delta_{ij}t}\left(|i\rangle\langle j| + \delta_{ij}\sum_{a:\Delta_{ai}\neq 0}\frac{2(1-\cos(\Delta_{ai}t))}{\Delta_{ai}^2t^2}|a\rangle\langle a| + \delta_{ij}\sum_{a:\Delta_{ai}=0}|a\rangle\langle a|\right)
\end{aligned}$$

We can now combine each of these terms to offer the full picture of the output of the channel to second order. We make two modifications to the results from each sum: first, we will switch to double index notation to make for easier use in other areas, and secondly we let $\rho = |i,j\rangle\langle k,l|$.

We note that the first term in the following equation is provided by $(A)+(B)$, the second through $(E)+(F)$, and the last two through $(C)+(D)$.

$$\mathbb{E}_G\left[\frac{\partial^2}{\partial\alpha^2}\Phi_{G(|i,j\rangle\langle k,l|)}\big|_{\alpha=0}\right] \tag{6.286}$$

$$= -\frac{2e^{i\Delta(i,j|k,l)t}}{\dim+1}\left(\sum_{(a,b):\Delta(i,j|a,b)\neq 0}\frac{1-i\Delta(i,j|a,b)t-e^{-i\Delta(i,j|a,b)t}}{\Delta(i,j|a,b)^2}\right.$$

$$\left.+\sum_{(a,b):\Delta(k,l|a,b)\neq 0}\frac{1+i\Delta(k,l|a,b)t-e^{i\Delta(k,l|a,b)t}}{\Delta(k,l|a,b)^2}+\frac{t^2}{2}(\eta(i,j)+\eta(k,l))\right)|i,j\rangle\langle k,l|$$

$$+\delta_{i,k}\delta_{j,l}\frac{2e^{i\Delta(i,j|k,l)t}}{\dim+1}\left(\sum_{(a,b):\Delta(i,j|a,b)\neq 0}\frac{2(1-\cos(\Delta(i,j|a,b)t))}{\Delta(i,j|a,b)^2}|a,b\rangle\langle a,b|+t^2\sum_{(a,b):\Delta(i,j|a,b)=0}|a,b\rangle\langle a,b|\right)$$

The last step we need is to use the half angle formula to change the cosine to a sine

$$\frac{2(1-\cos(\Delta(i,j|\ a,b)t))}{\Delta(i,j|a,b)^2}=\frac{2\left(1-\left(1-2\sin^2\left(\frac{\Delta(i,j|a,b)t}{2}\right)\right)\right)}{\Delta(i,j|a,b)^2}=t^2\,\mathrm{sinc}^2\left(\frac{\Delta(i,j|a,b)t}{2}\right), \tag{6.287}$$

which yields the statement.

We can compute these by plugging in to Eq. Equation (3.117) again, which yields

$$\mathbb{E}_G[\langle i',j'|\mathcal{T}(|i,j\rangle\langle i,j|)|i',j'\rangle]=\begin{cases}\tilde{\alpha}^2\,\mathrm{sinc}^2(\Delta(i,j\mid i',j')\frac{t}{2}) & (i,j)\neq(i',j')\\ -\tilde{\alpha}^2\sum_{(a,b)\neq(i,j)}\mathrm{sinc}^2(\Delta(a,b|i,j)\frac{t}{2}) & (i,j)=(i',j')\end{cases}. \tag{6.288}$$

The $(i,j)\neq(i',j')$ case should be apparent, the first term with the coherence factors χ are zero and the second term is what remains. The $(i,j)=(i',j')$ case can be seen as follows. For the first term we have

$$-\frac{\alpha^2 e^{i\Delta(i,j|\ i,j)t}}{\dim+1}\left(\chi(i,j)+\chi(i,j)^*+\frac{t^2}{2}(\eta(i,j)+\eta(i,j))\right)|i,j\rangle\langle i,j|. \tag{6.289}$$

We first compute the sum $\chi(i,j)+\chi(i,j)^*$ as

$$\chi(i,j)+\chi(i,j)^* = \sum_{a,b:\Delta(i,j,|a,b)\neq 0} \frac{1-i\Delta(i,j|a,b)t-e^{-i\Delta(i,j|a,b)t}}{\Delta(i,j|a,b)^2} + \sum_{a,b:\Delta(i,j,|a,b)\neq 0} \frac{1+i\Delta(i,j|a,b)t-e^{+i\Delta(i,j|a,b)t}}{\Delta(i,j|a,b)^2}$$
$$= \sum_{a,b:\Delta(i,j|\ a,b)\neq 0} \frac{2-e^{-i\Delta(i,j|\ a,b)t}-e^{+i\Delta(i,j|\ a,b)t}}{\Delta(i,j|a,b)^2}$$
$$= \sum_{a,b:\Delta(i,j|\ a,b)\neq 0} t^2\,\mathrm{sinc}^2\left(\frac{\Delta(i,j|\ a,b)t}{2}\right), \tag{6.290}$$

where the last step follows from a trigonometric identity (see Equation (6.287) in Appendix B.1 for details). Since $\mathrm{sinc}(0)=1$ the $\eta(i,j)$ term can be expressed as $\eta(i,j) = \sum_{a,b:\Delta(i,j|a,b)=0}\mathrm{sinc}^2\left(\frac{\Delta(i,j|\ a,b)t}{2}\right)$. Plugging this into Eq. Equation (3.117) gives

$$\mathbb{E}_G[\langle i,j|\mathcal{T}(|i,j\rangle\langle i,j|)|i,j\rangle]$$
$$= \langle i,j|\left(-\frac{\alpha^2t^2}{\dim+1}\sum_{a,b}\mathrm{sinc}^2\left(\frac{\Delta(i,j|\ a,b)t}{2}\right)|i,j\rangle\langle i,j| + \sum_{a,b}\mathrm{sinc}^2\left(\frac{\Delta(i,j\mid a,b)t}{2}\right)|a,b\rangle\langle a,b|\right)|i,j\rangle$$
$$= -\frac{\alpha^2t^2}{\dim+1}\sum_{(a,b)\neq(i,j)}\mathrm{sinc}^2\left(\frac{\Delta(i,j|\ a,b)t}{2}\right). \tag{6.291}$$

As a by-product of this computation we have also shown that $\mathrm{tr}(\mathcal{T}(\rho))=0$ and that our mapping is trace preserving to $O(\alpha^2)$.

□

Proof of Theorem 3.1: First we note that although $R_\Phi(\rho) = \frac{\alpha^3}{6}\partial_\alpha^3\Phi(\rho)\ |_{\alpha=\alpha_\star}$ for a specific value $\alpha_\star>0$ our proof will actually hold for any value of $\alpha_\star>0$. To compute the trace norm we will use the triangle inequality, unitary invariance of the Schätten norms, and submultiplicativity. To start,

$$\left\|\partial_\alpha^3\Phi(\rho)\right\|_1 = \left\|\frac{\partial^3}{\partial\alpha^3}\mathbb{E}_G\,\mathrm{tr}_E\,e^{i(H+\alpha G)t}\rho\otimes\rho_E e^{-i(H+\alpha G)t}\right\|_1$$
$$\leq \mathbb{E}_G\left\|\frac{\partial^3}{\partial\alpha^3}e^{i(H+\alpha G)t}\rho\otimes\rho_E e^{-i(H+\alpha G)t}\right\|_1, \tag{6.292}$$

where we can take $\mathbb{E}_G$ out of the norm via the triangle inequality and we can remove the trace via Proposition 1 of [90], which proves $\left\|\mathrm{tr}_E[X]\right\|_1 \leq \|X\|_{\dim_E} \leq \|X\|_1$. To proceed we use the decomposition of the second derivatives from the proof of Lemma 3.2, specifically Equation (6.274). This gives the following

$$\begin{aligned}\|R_\Phi\|_1 &\leq \frac{\alpha^3}{6}\mathbb{E}_G\Big\|\partial_\alpha((A)+(B)+(C)+(D)+(E)+(F))|_{\alpha=\alpha_\star}\Big\|_1 \\ &\leq \frac{\alpha^3}{6}\Big(\mathbb{E}_G\Big\|\partial_\alpha(A)|_{\alpha=\alpha_\star}\Big\|_1 + ... + \mathbb{E}_G\Big\|\partial_\alpha(F)|_{\alpha=\alpha_\star}\Big\|_1\Big). \end{aligned} \tag{6.293}$$

We will demonstrate how this can be computed for the first term $\partial_\alpha(A)$. Using Equation (6.275) and letting $H_\alpha = H + \alpha G$ for brevity we can write

$$\partial_\alpha(A) = -t^2\partial_\alpha \int_0^1\int_0^1 \underbrace{e^{is_1s_2H_\alpha t}}_{\text{(A.1)}} G \underbrace{e^{is_1(1-s_2)H_\alpha t}}_{\text{(A.2)}} G \underbrace{e^{i(1-s_1)H_\alpha t}}_{\text{(A.3)}} \rho \underbrace{e^{-iH_\alpha t}}_{\text{(A.4)}} s_1 ds_1 ds_2, \tag{6.294}$$

where there are four spots for the derivative to act via Duhamel's formula. We will show only one of these terms, starting with (A.1)

$$\begin{aligned}\text{(A.1)} &= -t^2\int_0^1\int_0^1 \partial_\alpha(e^{is_1s_2H_\alpha t})Ge^{is_1(1-s_2)H_\alpha t}Ge^{i(1-s_1)H_\alpha t}\rho e^{-iH_\alpha t}s_1 ds_1 ds_2 \\ &= (it)^3\int_0^1\int_0^1\int_0^1 e^{is_1s_2s_3H_\alpha t}Ge^{is_1s_2(1-s_3)H_\alpha t}Ge^{is_1(1-s_2)H_\alpha t}Ge^{i(1-s_1)H_\alpha t}\rho e^{-iH_\alpha t}s_1^2 s_2 ds_1 ds_2 ds_3.\end{aligned} \tag{6.295}$$

Our goal is to compute the 1-norm of the above expression at $\alpha = \alpha_\star$. We can do so using the triangle inequality to move the norms into the integrand and then use submultiplicativity and unitary invariance to achieve

$$\begin{aligned}&\Big\|e^{is_1s_2s_3H_{\alpha_\star}t}Ge^{is_1s_2(1-s_3)H_{\alpha_\star}t}Ge^{is_1(1-s_2)H_{\alpha_\star}t}Ge^{i(1-s_1)H_{\alpha_\star}t}\rho e^{-iH_{\alpha_\star}t}\Big\|_1 \\ &\leq \Big\|e^{is_1s_2s_3H_{\alpha_\star}t}G\Big\|_1\Big\|e^{is_1s_2(1-s_3)H_{\alpha_\star}t}G\Big\|_1\Big\|e^{is_1(1-s_2)H_{\alpha_\star}t}G\Big\|_1\Big\|e^{i(1-s_1)H_{\alpha_\star}t}\rho e^{-iH_{\alpha_\star}t}\Big\|_1 \\ &\leq \|G\|_1^3\|\rho\|_1 = \|G\|_1^3.\end{aligned} \tag{6.296}$$

Similar computations can be carried out for the other three terms (A.2) - (A.4). In total these yield the inequality

$$\begin{aligned}\frac{\alpha^3}{6}\int\|\partial_\alpha(A)\|_1 dG &\leq \frac{(\alpha t)^3}{6}\mathbb{E}_G\int_0^1\int_0^1\int_0^1 \|G\|_1^3(s_1^2s_2 + s_1^2(1-s_2) + s_1(2-s_1))ds_1ds_2ds_3 \\ &\leq \frac{4}{6}(\alpha t)^3\mathbb{E}_G\|G\|_1^3.\end{aligned} \tag{6.297}$$

Now that we have computed a bound for the norm of the derivative acting on (A) we only have terms (B) through (F) to compute. These can all be checked to satisfy the same bound on (A) from Equation (6.297), and as there are six terms in total we have the inequality

$$\|R_\Phi\|_1 \leq 4(\alpha t)^3\mathbb{E}_G\|G\|_1^3, \tag{6.298}$$

which holds for all inputs ρ.

Our last remaining problem is to compute the expected norm of G. Using the decomposition of our interaction $G = U_G \Lambda_G U_G^\dagger$ to get

$$\mathbb{E}_G \|G\|_1^3 = \mathbb{E}_{\Lambda_G} \mathbb{E}_{U_G} \left\| U_G \Lambda_G U_G^\dagger \right\|_1^3 = \mathbb{E}_{\Lambda_G} \|\Lambda_G\|_1^3 = \mathbb{E}_{\Lambda_G} \sum_{i=1}^{\dim} |\Lambda_G(i)|^3 = \dim, \tag{6.299}$$

Since Λ_G is just ± 1 times a Pauli Z string each eigenvalue has norm 1. This gives the final inequality

$$\|R_\Phi\|_1 \leq 4 \dim (\alpha t)^3. \tag{6.300}$$

□

Appendix C

Scratch

C.1 Template Thermal State Prep Proof

I'm thinking of including a "template" theorem that can be used to simplify the 4 proofs contained in the following section. Let ρ_{fix} denote the unique fixed point for a channel $\mathbb{E}_\gamma\left[\mathbb{1} + \mathcal{T}_{\text{on}}^{(\gamma)}\right]$ and $\tilde{\lambda_\star}$ the spectral gap of the scaled transition matrix, so $\tilde{\alpha}^2\tilde{\lambda_\star} = \lambda_\star$. Then we have

$$\begin{aligned}
\left\|\rho_{\text{fix}} - \left(\mathbb{E}_\gamma \Phi_\gamma\right)^{\circ L}(\rho)\right\|_1 &\leq \left\|\rho_{\text{fix}} - \left(\mathbb{1} + \mathbb{E}_\gamma \mathcal{T}_{\text{on}}^{(\gamma)}\right)^{\circ L}\right\|_1 + L\left\|\mathbb{E}_\gamma \mathcal{T}_{\text{off}}^{(\gamma)} + R_\Phi\right\|_1 \\
&\leq \left\|\rho_{\text{fix}} - \left(\mathbb{1} + \mathbb{E}_\gamma \mathcal{T}_{\text{on}}^{(\gamma)}\right)^{\circ L}\right\|_1 + L\left(\left\|\mathbb{E}_\gamma \mathcal{T}_{\text{off}}^{(\gamma)}\right\|_1 + \|R_\Phi\|_1\right) \\
&\leq \left\|\rho_{\text{fix}} - \left(\mathbb{1} + \mathbb{E}_\gamma \mathcal{T}_{\text{on}}^{(\gamma)}\right)^{\circ L}\right\|_1 + L\left(\frac{8\alpha^2}{\delta_{\min}^2} + \frac{16\sqrt{2}}{\sqrt{\pi}}\dim_S (\alpha t)^3\right) \quad (7.301)
\end{aligned}$$

So now in order to balance these terms we can set $\alpha = 1/(\dim_S \delta_{\min}^2 t^3)$ and the expression on the right becomes $L\frac{\alpha^2}{\delta_{\min}^2}\left(8 + 16\sqrt{\frac{2}{\pi}}\right)$. Now using Jerison's theorem we can argue that

$$L \geq \frac{\dim^2}{\alpha^2 t^2 \tilde{\lambda_\star}} J \quad (7.302)$$

is sufficient to guarantee that the distance to the fixed point is $\tilde{O}(\varepsilon)$. Now we note that the right hand side forces us to require $L\frac{\alpha^2}{\delta_{\min}^2} \in \tilde{O}(\varepsilon)$ holds only if

$$\frac{\dim^2}{\delta_{\min}^2 t^2 \tilde{\lambda_\star}} \in \tilde{O}(\varepsilon) \quad (7.303)$$

can be satisfied if $t = \frac{\dim}{\delta_{\min}\sqrt{\varepsilon\tilde{\lambda_\star}}}$. and then we are done.

Bibliography

[1] R. M. Neal, Probabilistic inference using Markov chain Monte Carlo methods, (1993)

[2] J. M. Deutsch, Eigenstate Thermalization Hypothesis, Reports on Progress in Physics **81**, 82001 (2018)

[3] E. B. Davies, Markovian master equations, Communications in Mathematical Physics **39**, 91 (1974)

[4] C.-F. Chen, M. J. Kastoryano, F. G. S. L. Brandão, and A. Gilyén, Quantum Thermal State Preparation, (2023)

[5] A. Gilyén, C.-F. Chen, J. F. Doriguello, and M. J. Kastoryano, Quantum generalizations of Glauber and Metropolis dynamics, (2024)

[6] Z. Ding, B. Li, and L. Lin, Efficient quantum Gibbs samplers with Kubo–Martin–Schwinger detailed balance condition, (2024)

[7] T. S. Cubitt, Dissipative ground state preparation and the Dissipative Quantum Eigensolver, (2023)

[8] M. Hagan and N. Wiebe, Composite Quantum Simulations, Arxiv Preprint Arxiv:2206.06409 (2022)

[9] M. Hagan and N. Wiebe, The Thermodynamic Cost of Ignorance: Thermal State Preparation with One Ancilla Qubit, Arxiv Preprint Arxiv:2502.03410 (2025)

[10] M. Pocrnic, M. Hagan, J. Carrasquilla, D. Segal, and N. Wiebe, Composite Qdrift-product formulas for quantum and classical simulations in real and imaginary time, Physical Review Research **6**, 13224 (2024)

[11] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Simulation of electronic structure Hamiltonians using quantum computers, Molecular Physics **109**, 735 (2011)

[12] S. P. Jordan, K. S. Lee, and J. Preskill, Quantum algorithms for quantum field theories, Science **336**, 1130 (2012)

[13] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proceedings of the National Academy of Sciences **114**, 7555 (2017)

[14] R. Babbush, D. W. Berry, and H. Neven, Quantum simulation of the Sachdev-Ye-Kitaev model by asymmetric qubitization, Phys. Rev. A **99**, 40301 (2019)

[15] Y. Su, D. W. Berry, N. Wiebe, N. Rubin, and R. Babbush, Fault-Tolerant Quantum Simulations of Chemistry in First Quantization, PRX Quantum **2**, 40332 (2021)

[16] T. E. O'Brien et al., Efficient quantum computation of molecular forces and other energy gradients, Phys. Rev. Res. **4**, 43210 (2022)

[17] D. Aharonov and A. Ta-Shma, *Adiabatic Quantum State Generation and Statistical Zero Knowledge*, in *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing* (2003), pp. 20–29

[18] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, Efficient quantum algorithms for simulating sparse Hamiltonians, Communications in Mathematical Physics **270**, 359 (2007)

[19] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Simulating Hamiltonian Dynamics with a Truncated Taylor Series, Phys. Rev. Lett. **114**, 90502 (2015)

[20] A. M. Childs, A. Ostrander, and Y. Su, Faster quantum simulation by randomization, Quantum **3**, 182 (2019)

[21] G. H. Low and I. L. Chuang, Hamiltonian Simulation by Qubitization, Quantum **3**, 163 (2019)

[22] G. H. Low, V. Kliuchnikov, and N. Wiebe, Well-conditioned multiproduct Hamiltonian simulation, Arxiv Preprint Arxiv:1907.11679 (2019)

[23] G. H. Low and N. Wiebe, Hamiltonian Simulation in the Interaction Picture, Arxiv Preprint Arxiv:1897.10070 (2019)

[24] E. Campbell, Random Compiler for Fast Hamiltonian Simulation, Phys. Rev. Lett. **123**, 70503 (2019)

[25] N. Wiebe, D. Berry, P. Høyer, and B. C. Sanders, Higher order decompositions of ordered operator exponentials, Journal of Physics A: Mathematical and Theoretical **43**, 65203 (2010)

[26] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of Trotter Error with Commutator Scaling, Phys. Rev. X **11**, 11020 (2021)

[27] D. W. Berry, A. M. Childs, Y. Su, X. Wang, and N. Wiebe, Time-dependent Hamiltonian simulation with L^1-norm scaling, Quantum **4**, 254 (2020)

[28] D. Wecker, B. Bauer, B. K. Clark, M. B. Hastings, and M. Troyer, Gate-count estimates for performing quantum chemistry on small quantum computers, Physical Review a **90**, (2014)

[29] D. Poulin, M. B. Hastings, D. Wecker, N. Wiebe, A. C. Doherty, and M. Troyer, The Trotter step size required for accurate quantum simulation of quantum chemistry, Arxiv Preprint Arxiv:1406.4920 (2014)

[30] I. D. Kivlichan, C. E. Granade, and N. Wiebe, Phase estimation with randomized Hamiltonians, Arxiv Preprint Arxiv:1907.10070 (2019)

[31] A. Rajput, A. Roggero, and N. Wiebe, Hybridized Methods for Quantum Simulation in the Interaction Picture, Quantum **6**, 780 (2022)

[32] Y. Ouyang, D. R. White, and E. T. Campbell, Compilation by stochastic Hamiltonian sparsification, Quantum **4**, 235 (2020)

[33] S. Jin and X. Li, A Partially Random Trotter Algorithm for Quantum Hamiltonian Simulations, Arxiv Preprint Arxiv:2109.07987 (2021)

[34] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996)

[35] M. Hagan and N. Wiebe, Composite quantum simulations, Quantum **7**, 1181 (2023)

[36] M. Pocrnic and M. Hagan, Trotter-QDrift-Simulation, (n.d.)

[37] F. Pedregosa et al., Scikit-learn: Machine learning in Python, The Journal of Machine Learning Research **12**, 2825 (2011)

[38] J. R. McClean et al., OpenFermion: the electronic structure package for quantum computers, Quantum Science and Technology **5**, 34014 (2020)

[39] Q. Sun et al., PySCF: the Python-based simulations of chemistry framework, Wires Computational Molecular Science **8**, e1340 (2018)

[40] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, Low-Depth Quantum Simulation of Materials, Phys. Rev. X **8**, 11044 (2018)

[41] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Reviews of Modern Physics **73**, 33 (2001)

[42] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, Science **309**, 1704 (2005)

[43] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction, PRX Quantum **2**, 30305 (2021)

[44] F. G. S. L. Brandão, A. Kalev, T. Li, C. Y.-Y. Lin, K. M. Svore, and X. Wu, Quantum SDP Solvers: Large Speed-ups, Optimality, and Applications to Quantum Learning, (2019)

[45] A. Anshu, S. Arunachalam, T. Kuwahara, and M. Soleimanifar, Sample-efficient learning of quantum many-body systems, Nature Physics **17**, 931 (2021)

[46] J. Kempe, A. Kitaev, and O. Regev, The Complexity of the Local Hamiltonian Problem, (2005)

[47] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, The Journal of Chemical Physics **21**, 1087 (1953)

[48] D. Roth, On the hardness of approximate reasoning, Artificial Intelligence **82**, 273 (1996)

[49] E. Loh Jr., J. Gubernatis, R. Scalettar, S. White, D. Scalapino, and R. Sugar, Sign problem in the numerical simulation of many-electron systems, Prb **41**, 9301 (1990)

[50] M. Troyer and U.-J. Wiese, Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations, Physical Review Letters **94**, (2005)

[51] A. Beskos, N. S. Pillai, G. O. Roberts, J. M. Sanz-Serna, and A. M. Stuart, Optimal tuning of the Hybrid Monte-Carlo Algorithm, (2010)

[52] M. Betancourt, A Conceptual Introduction to Hamiltonian Monte Carlo, (2018)

[53] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, Quantum Metropolis sampling, Nature **471**, 87 (2011)

[54] C. Marriott and J. Watrous, Quantum arthur–merlin games, Computational Complexity **14**, 122 (2005)

[55] D. Motlagh, M. S. Zini, J. M. Arrazola, and N. Wiebe, Ground State Preparation via Dynamical Cooling, Arxiv Preprint Arxiv:2404.05810 (2024)

[56] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nature Physics **16**, 205 (2019)

[57] Z. Ding, C.-F. Chen, and L. Lin, Single-ancilla ground state preparation via Lindbladians, Physical Review Research **6**, 33147 (2024)

[58] R. Kubo, Statistical-Mechanical Theory of Irreversible Processes. I, Journal of the Physical Society of Japan **12**, 570 (1957)

[59] P. C. Martin and J. Schwinger, Theory of Many-Particle Systems. I, Physical Review **115**, 1342 (1959)

[60] A. M. Dalzell et al., Quantum algorithms: A survey of applications and end-to-end complexities, (2023)

[61] C. Rouzé, D. S. França, and Á. M. Alhambra, Efficient thermalization and universal quantum computing with quantum Gibbs samplers, (2024)

[62] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, Vol. 107 (American Mathematical Soc., 2017)

[63] D. Jerison, General mixing time bounds for finite Markov chains via the absolute spectral gap, Arxiv Preprint Arxiv:1310.8021 (2013)

[64] D. Auerbach, Supercooling and the Mpemba effect: When hot water freezes quicker than cold, American Journal of Physics **63**, 882 (1995)

[65] E. B. Mpemba and D. G. Osborne, Cool?, Physics Education **4**, 172 (1969)

[66] F. Ivander, N. Anto-Sztrikacs, and D. Segal, Hyperacceleration of quantum thermalization dynamics by bypassing long-lived coherences: An analytical treatment, Phys. Rev. E **108**, 14130 (2023)

[67] Z. Lu and O. Raz, Nonequilibrium thermodynamics of the Markovian Mpemba effect and its inverse, Proceedings of the National Academy of Sciences **114**, 5083 (2017)

[68] J. Zhang et al., Observation of quantum strong Mpemba effect, Nature Communications **16**, 301 (2025)

[69] M. D. Hoffman and A. Gelman, The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, (2011)

[70] A. Prositto, M. Forbes, and D. Segal, Equilibrium and nonequilibrium steady states with the repeated interaction protocol: Relaxation dynamics and energetic cost, Arxiv Preprint Arxiv:2501.05392 (2025)

[71] J. Choi et al., Preparing random states and benchmarking with many-body quantum chaos, Nature **613**, 468 (2023)

[72] G. Alagic et al., Status Report on the First Round of the Additional Digital Signature Schemes for the NIST Post-Quantum Cryptography Standardization Process, NIST IR **8528**, (2024)

[73] A. Schmidhuber, R. O'Donnell, R. Kothari, and R. Babbush, Quartic quantum speedups for planted inference, (2024)

[74] D. W. Berry, Y. Su, C. Gyurik, R. King, J. Basso, A. D. T. Barba, A. Rajput, N. Wiebe, V. Dunjko, and R. Babbush, Analyzing prospects for quantum advantage in topological data analysis, PRX Quantum **5**, 10319 (2024)

[75] S. P. Jordan, N. Shutty, M. Wootters, A. Zalcman, A. Schmidhuber, R. King, S. V. Isakov, and R. Babbush, Optimization by decoded quantum interferometry, Arxiv Preprint Arxiv:2408.08292 (2024)

[76] S. Lee et al., Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, Nature Communications **14**, 1952 (2023)

[77] S. R. White, Density matrix formulation for quantum renormalization groups, Physical Review Letters **69**, 2863 (1992)

[78] R. Orús, Tensor networks for complex quantum systems, Nature Reviews Physics **1**, 538 (2019)

[79] L. J. &. C. G. K.-L. Lee S., Collection of papers referring exponential quantum advantage in quantum chemistry., (n.d.)

[80] A. Kan and B. Symons, Resource-optimized fault-tolerant simulation of the Fermi-Hubbard model and high-temperature superconductor models, Arxiv Preprint Arxiv:2411.02160 (2024)

[81] I. D. Kivlichan et al., Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via trotterization, Quantum **4**, 296 (2020)

[82] J. P. LeBlanc et al., Solutions of the two-dimensional Hubbard model: Benchmarks and results from a wide range of numerical algorithms, Physical Review X **5**, 41041 (2015)

[83] Q. Zhao, Y. Zhou, and A. M. Childs, Entanglement accelerates quantum simulation, Arxiv Preprint Arxiv:2406.02379 (2024)

[84] M. Metcalf, E. Stone, K. Klymko, A. F. Kemper, M. Sarovar, and W. A. de Jong, Quantum Markov Chain Monte Carlo with Digital Dissipative Dynamics on Quantum Computers, (2021)

[85] K. C. Young, M. Sarovar, J. Aytac, C. M. Herdman, and K. B. Whaley, Finite temperature quantum simulation of stabilizer Hamiltonians, Journal of Physics B: Atomic, Molecular and Optical Physics **45**, 154012 (2012)

[86] Y. Zhan, Z. Ding, J. Huhn, J. Gray, J. Preskill, G. K.-L. Chan, and L. Lin, Rapid quantum ground state preparation via dissipative dynamics, (2025)

[87] L. Lin and Y. Tong, Near-optimal ground state preparation, Quantum **4**, 372 (2020)

[88] O. Shtanko and R. Movassagh, Preparing thermal states on noiseless and noisy programmable quantum processors, (2023)

[89] F. G. Brandão, W. Chemissany, N. Hunter-Jones, R. Kueng, and J. Preskill, Models of Quantum Complexity Growth, PRX Quantum **2**, 30316 (2021)

[90] A. E. Rastegin, Relations for certain symmetric norms and anti-norms before and after partial trace, Journal of Statistical Physics **148**, 1040 (2012)