Quantum Gibbs Sampling

Thesis by
Chi-Fang Chen

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended July 9th, 2024

Chi-Fang Chen
ORCID: 0000-0001-5589-7896

ACKNOWLEDGEMENTS

Graduate school has been full of turns. I am heavily in debt to those who have lent their hands in times of uncertainty and who have fought together in pursuit of truth.

I want to especially thank my advisor, Fernando Brandão, for his incredibly fast and brutally honest advice about research and just anything. He always provides critical input at the right moment. He has also put no constraints on my travel and interests, which allows me to be delocalized among many great places and talents. I feel lucky to have been exposed to the ideas about thermalization and Gibbs sampling in this quantum computing wave. Without Fernando, I also wouldn't have met my fellow Gibbs samplers: András Gilyén, from whom I learned so much about quantum algorithms and mathematical writing, and Michael Kastoryano, from whom I learned about open systems and mixing times.

I also want to thank Joel Tropp, who introduced me to the beauty of randomness and matrices and whom I regarded as my underground co-advisor. Without his generosity and openness to crazy ideas, it would not have been possible to bridge ideas from random matrices to quantum information and back. I also feel privileged to work with Joel's postdoc, Jorge Garza-Vargas, and Ramon van Handel at Princeton, who guided me into the mathematical land of random matrix theory.

I want to thank John Preskill for collaborations and many fun lightning discussions. The wisdom in his words always returns to me. I also wanted to thank him for making his group meeting and IQIM such a thrilling, open, and engaging hub filled with great people. I also wanted to express my gratitude to my committee members at Caltech, Olexei Motrunich and Manuel Endres, for many helpful discussions along the way.

I realized I hadn't had a chance to officially thank those who really set me on this path. During my undergraduate years, I feel privileged to have worked with Andrew Lucas, my long-time collaborator, on earlier projects on Lieb-Robinson bounds since junior year. He was so tolerant, truthful, and open to my random questions about career and science, and he was hospitable during my visits to his group at CUBoulder. I also want to thank my undergrad mentors and advisors, Geoffrey Pennington, Grant Salton, Sean Hartnoll, and Patrick Hayden, who perhaps planted a little seed of quantumness. In retrospect, this all began with the Science Class program at Chien Kuo High School in Taiwan with a bunch of crazy smart peers, where we first tasted

the fun of math and science as we solved the Hydrogen atom wave function from scratch and implemented Fast Fourier transform in C. It was really the vision of our teachers and Prof. Ming-Jui Lin, who created such a stimulating scientific culture.

My travels would not have been so much fun and smooth without enthusiastic hosts: Umesh Vasirani and Lin Lin at Simons' Institute and Berkeley, Aram Harrow and Soonwon Choi at MIT, Anurag Anshu at Harvard, Alex Dalzell, and Sam Mcardle and Mario Berta at AWS, Patrick Hayden and Adam Bouland at Stanford, Minh Tran and Kunal Sharma at IBM, and Nobuyuki Yoshioka at TokyoU, Ying-Jer Gao and Hsi-Sheng Goan at NTU, Kai-Min Chung at Sinica, Han-Shuan Lin and Chung-Yu Mou at NTHU, and Ching-Yi Lai at NCYU. I should also thank Bonnie and Marica for handling much of the logistics and my friends who shared their sofas (for extended periods of time): Pei-Ming Chen, Isaac Yu, Ching-Ting Tsai, Kai-Yuan Cheng, and Vinh Nguyen around the Bay Area; Cyuan-Han Chang, Hau-wei Lin, and Han-Hsin Lin around Caltech; Kai-Feng Chen, and Hsin-Yuan Huang and Chi-Yun Cheng around Boston. Thank you to all my friends. My journey is always filled with random fun stuff; let me just list the activities we did together: board games, video games, bubble tea, hiking, thanksgiving, hotpot, etc.

Lastly, I owe much to my family and especially my parents, Tzu-Chih Chen and Yu-Chu Chiu, for their unbounded kindness and forgiveness, which make my life fault-tolerant to this day.

ABSTRACT

Markov Chain Monte Carlo algorithms are indispensable in classical thermodynamic simulation, perhaps due to their mathematical simplicity, algorithmic efficiency, and physical origin. In particular, Glauber dynamics is a detailed-balanced continuous-time Markov chain that fixes the Gibbs distribution and also serves as a mathematically succinct model of classical thermalization. In this thesis, we proposed a quantum computation analog of Glauber dynamics that is exactly detailed balanced yet algorithmic efficient, inherits the locality of the target Hamiltonian, and resembles Davies'-like generators physically derived from a weak system-bath coupling. We hope our proposal will serve as a quantum algorithm for quantum thermodynamic simulation and a model of open system thermalization where a suitable construction has been lacking for noncommuting Hamiltonians.

PUBLISHED CONTENT AND CONTRIBUTIONS

The following articles are included (with minor edits) as chapter 2, chapter 3, and chapter 4, in order.

[1] Chi-Fang Chen, Michael J Kastoryano, Fernando GSL Brandão, and András Gilyén. Quantum thermal state preparation. *arXiv preprint arXiv:2303.18224*, 2023. URL https://arxiv.org/abs/2303.18224.
CFC participated in the conception of the project, the development of the proofs, and the writing of the manuscript. [In preparation for submission].

[2] Chi-Fang Chen, Michael J Kastoryano, and András Gilyén. An efficient and exact noncommutative quantum gibbs sampler. *Short plenary talk, QIP2024. arXiv preprint arXiv:2311.09207.*, 2023. URL https://arxiv.org/abs/2311.09207.
CFC conceived of the project and participated in the development of the proofs and the writing of the manuscript. [In preparation for submission].

[3] Chi-Fang Chen, Hsin-Yuan Huang, John Preskill, and Leo Zhou. Local minima in quantum systems. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1323–1330, 2024. URL https://doi.org/10.1145/3618260.3649675.
CFC participated in the conception of the project, the proofs, and the writing of the manuscript, particularly in the analysis of energy gradients.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

Davies' generator. A Lindbladian rigorously derived for a system interacting weakly with a thermal bath. It satisfies detailed balance exactly and has played the role of Glauber dynamics in finite-sizes and commuting systems but is unphysical for general noncommuting many-body systems. Today, there are a few modernized proposals of Davies' like generators, but they all violate detailed balance in general.

Detailed balance condition. A key condition in the design and analysis of Markov chains that prescribes a desired stationary state. This condition requires that the mass exchange between two configurations at stationarity cancels each other.

Gibbs state. A quantum state that is a mixture of energy eigenstates with a Boltzmann weight. It is often assumed that a quantum system will eventually arrive at such a state at the same temperature as the bath.

Glauber dynamics. A continuous-time classical Markov that satisfies detailed balance regarding the classical Gibbs distribution. For both numerical and analytic purposes, it is often the go-to dynamical model for classical systems interacting with a thermal bath.

Ground states. The set of minimal energy eigenstates of a Hamiltonian matrix. A Gibbs state at infinitely low temperature would be a mixture of ground states.

Lindbladian. A continuous-time Markovian process acting on quantum states, also known as Liouvillian, or master equation. Such effective description often arises from an open system coupled to a bath but may be considered in its own right.

Markov chain. A discrete-time or continuous-time stochastic process where the future is only correlated to the past through the present. Often they can be implemented efficiently on a classical computer to sample from a desired distribution, and is also called *Markov Chain Monte Carlo* as an algorithmic method.

Phase transition. A phenomenon describing a qualitative change in the physical properties of a system depending on the parameters such as magnetic field or temperature. Such transition can be static in terms of ground or thermal state properties (e.g., long-range correlations) or dynamic in terms of the behavior of an associated Linbladian (e.g., mixing times).

Quantum algorithms. Algorithms that run on quantum computers, often described in terms of quantum gates.

Quantum computational advantage. An advantage of quantum computers over classical computers for certain classes of problems. Currently, the quantum hardware is limited in quality and scale, so most evidence is theoretical.

Quantum computers. Computers that can act quantum gates on qubits given classical instructions. In this thesis, we will assume them to be fault-tolerant so that we can execute quantum algorithms with high precision.

Quantum Gibbs sampling. A quantum algorithm simulating a Quantum Markov chain (continuous or discrete time) whose stationary state is the Quantum Gibbs state.

Quantum many-body physics. Quantum phenomena emerging from systems with many degrees of freedom, contrasting with few-body systems. Often, complex, unexpected behavior can occur, but it is very challenging to solve analytically or numerically.

Quantum simulation. Simulation of a naturally appearing or artificial quantum system with a quantum computer that allows us to extract its physical properties in a programmable, controllable way.

Thermal bath. A large quantum system with a characteristic temperature with Markovian properties.

Thermal state. The state describing a quantum system in equilibrium with a thermal bath. Although broadly defined, this can sometimes mean the quantum Gibbs state.

Chapter 1

INTRODUCTION

We often take for granted that a quantum system without isolation eventually arrives at the *thermal state*, or the *Gibbs state*; at low enough temperatures, the system must be in its *ground state*. This dogma dictates a *static* view we reason about the physical properties of quantum systems, from atoms to materials to chemicals. First, we write down a Hamiltonian operator; second, we solve the eigenvalue problem. This mathematical paradigm has led to the exact solution for the orbitals and spectrum of simple atomic and molecular systems. In the many-body case, while Schrodinger equations become too complex to solve analytically, an arsenal of classical simulation methods has been developed, such as density functional theory [89], tensor network (e.g., [145, 177]), and quantum Monte Carlo [34], that gives a remarkably detailed understanding of phase transitions, ground and thermal state properties, and reaction rates. However, typically, all these classical methods break down in their own ways in the presence of strong quantum correlation, from heavy nuclei, high-T_c superconductors, to catalysts.

Recently, a new type of computer that runs on the laws of quantum mechanics, *quantum computers*, has gathered significant interest among physicists and computer scientists. Hopefully, future quantum computers could offer new lenses into an even broader array of quantum many-body problems that are otherwise classically intractable. Nevertheless, decades of search for particular physical systems where quantum computers can significantly outperform classical methods have been inconclusive. The challenge is that current devices are limited in their sizes and quality, leading the community to rely heavily on a theoretical proof of quantum computational advantage. From computational complexity theory, however, the standard many-body ground-state problem as formulated is generally intractable even for quantum computers (*QMA-hard*, as a quantum analog of NP-hard) [101]. Even though such worst-case pessimism may not apply to physically relevant quantum systems, it foreshadows that the search for provable quantum advantage in ground-state problems might not be clear-cut. Indeed, the efficacy of most existing quantum algorithm proposals for low-temperature physics (e.g., [14, 30, 36, 107]) relies on additional assumptions that are not always easy to verify and have been challenged in numerical studies [108]. Even though quantum simulation is commonly regarded as a

"killer application" where quantum computers may significantly accelerate scientific discovery, still today, there has not been a go-to quantum algorithm for preparing ground states or thermal states. Shor's factoring algorithm from 1994 [159] remains a concrete, stand-alone example of an end-to-end, practical quantum computational advantage, while many other proposals [83] have later found efficient classical solutions [165].

A motivation for this thesis is to offer a fresh perspective on the quantum simulation problem. Conceptually, this requires confronting the fundamental "static" view of thermodynamics and shifting towards a more physically motivated "dynamical" formulation. This draws a parallel to the development of classical Markov chain Monte Carlo (MCMC) algorithms that allow for direct dynamical simulation of thermal fluctuations due to a bath, except that designing a truly quantum MCMC algorithm faced additional conceptual challenges rooted in quantum mechanics. These two physical and computational veins of pursuit and challenges come to a confluence: a new Nature-inspired quantum algorithmic principle for simulation of low-temperature physics—*quantum Gibbs sampling*.

1.1 Nature's cooling process

Take a step back; there is a living example of a large-scale quantum algorithm for low-temperature physics that we encounter every day for a wide range of systems: refrigerators. Our experimental colleagues routinely cool their devices by lowering the temperature of their dilution fridges and laser coolers, through which they observe Bose-Einstein condensation, superconducting phase transitions, and the emergence of nontrivial correlation in graphene. If any of the above quantum mechanical phenomena could be systematically and efficiently reproduced in programmable quantum computers, no one should doubt their practical usefulness as a quantum simulator. Therefore, a quantum algorithm that effectively simulates Nature's cooling process should already be sufficient for quantum simulation purposes, in order to reproduce physical phenomena appearing in natural thermodynamic environments. Indeed, why should we be bothered that quantum computers cannot prepare QMA-hard ground states if these high-complexity states never appear in Nature?

From an algorithmic designer's view, Nature's cooling scheme has features rarely seen in usual quantum algorithms: it follows a macroscopic thermodynamic principle yet works universally for various systems and baths, and it is agnostic to initial state and noises, while standard quantum algorithms are often very structured and prone

to errors. It is tempting to speculate that extracting the essential working principle behind cooling could lead to a new quantum algorithmic paradigm that inherits and even improves its efficacy.

Nevertheless, it has been nontrivial to obtain a mathematically simple but physically adequate description of the cooling process. From first principles, a quantum system interacting with a thermal environment is yet another many-body Schrodinger equation with substantial degrees of freedom; one can certainly simulate or solve the global Hamiltonian evolution [168], but that might be overkill for understanding the system alone, which is often much smaller than the bath. To minimize the complexity of such open system dynamics, physicists often introduce a *master equation*, or a *Lindbladian*, that aims to capture the effect of a bath by a much simpler continuous-time Markovian process acting upon the system (e.g., [29, 150]). However, the seminal *Davies' generator* from 1974 [56, 58], albeit simple, is derived assuming the system to be coupled *infinitely* weakly to an infinitely large Markovian thermal bath. Shockingly, even in modern versions of Davies-like Lindbladians [121, 134] that adhere to realistic system-bath assumptions, there has not been a proof of convergence to the thermal state for general noncommuting Hamiltonians, or provably correct quantum algorithms that precisely simulate such Lindbladians. Even though there has been a great volume of open system literature, quantitative arguments beyond single-qubits appear relatively rare, and we will have to lay out algorithmic and analytic primitives ourselves (see chapter 2).

1.2 Markov Chain Monte Carlo algorithms

Taking another step back, there is an algorithm that simulates *classical* thermodynamics in almost all the ways we wished for quantum: Markov Chain Monte Carlo (MCMC) algorithms, one of the most versatile yet simple algorithmic principles ever devised. For any target distribution — for example the Gibbs distribution of an Ising magnet — the Metropolis-Hastings algorithm [86, 130] prescribes a Markov Chain that *samples* from the distribution and gives estimates for statistical properties without explicitly storing the full probability distribution. The key ingredient of this recipe is the *detailed balance* condition, which simultaneously ensures that the target distribution is stationary and leads to a powerful framework to study their convergence time (the so-called *mixing time*), through a mathematically well-established quantity called the *spectral gap* (see, e.g., [110]). Remarkably, detailed balance admits very efficient algorithmic implementations, as simple as local spin updates in Ising models. Seventy years since its debut, today, Markov Chain Monte Carlo as an algorithmic

principle continues to apply to newly emerging problems beyond physics, from biology to artificial intelligence.

For our purpose, we should not only view MCMC methods as a purely algorithmic construct but also as a succinct mathematical model of Nature's cooling process. Most notably, *Glauber dynamics* [75] is a continuous-time classical Markov chain that can be traced to Davies' generator (for classical Hamiltonians). Even though Glauber dynamics surely does not capture all nuances of open system physics (e.g., quantum mechanical and non-Markovian effects of the bath), the combination of desirable algorithmic and analytic features has nevertheless led to a comprehensive picture of dynamic phase transitions (in thermalization and in complexity) for classical spin systems, including magnets and spin glasses.

If quantum computing will play a likewise revolutionary role in quantum thermodynamics, one naturally speculates that a suitable "quantum MCMC" algorithm will have an equally important and enduring impact as the classical case. Nonetheless, early attempts have been compounded by obstacles rooted in quantum mechanics. To see this, consider the following Metropolis-style procedure [167] on the energy eigenbasis: suppose we start with an energy eigenstate and apply a unitary update. If the energy decreases, we wish to "accept" this move; otherwise, we "reject" this move with a prescribed probability. However, this is in tension with (1) the no-cloning theorem (to return to the previous quantum state without keeping many copies) and (2) the energy-time uncertainty principle (to know confidently whether energy has increased or decreased). The first issue turns out to be a technical problem tied to discrete-time quantum Markov Chains and automatically resolves in continuous-time settings (i.e., Lindbladians) [149, 160, 168]. Nevertheless, the uncertainty in energy has remained a fundamental threat to quantum detailed balance, and existing work has failed to retain both the exactness of quantum detailed balance (the basis for modern Markov chain analysis) and finite energy resolution (inversely proportional to the time for Hamiltonian simulation, the dominant cost in most quantum simulation algorithms). Perhaps due to the lack of a simple "go-to" proposal that matches the features of classical Glauber dynamics, the potential impact of Quantum MCMC algorithms has mainly been under the shadows for the quantum computing community—see chapter 2 for a detailed account for literature.

1.3 Main result

In this thesis, we propose a quantum Markov Chain Monte Carlo algorithm that can also be regarded as a succinct mathematical model of open system thermalization. For any Hamiltonian of interest, we construct a Lindbladian, a continuous-time quantum Markov chain with the following combination of desirable features akin to Glauber's:

- **Exact Quantum Detailed balance.** The Lindbladian satisfies quantum detailed balance exactly. Therefore, the quantum Gibbs is a stationary state of the Lindbladian. This allows for clean mathematical reasoning and systematically extends the formalism and techniques developed in classical and commuting cases for mixing time (e.g., [31, 96, 124]). From a quantum algorithm perspective, detailed balance is also naturally compatible with the quantum walk formalism [164], giving a coherent algorithm for the "purified" Gibbs state.

- **Efficient simulation.** The Lindbladian dynamics can be efficiently simulated on a quantum computer, as long as the Hamiltonian dynamic can be simulated for a time depending on the temperature. We expect the simulation cost to be asymptotically optimal (up to polylogarithms).

- **Quasi-locality.** For lattice Hamiltonians, the Lindbladian decomposes into a sum over (quasi)-local Lindbladians, with locality depending on the temperature (and polylogarithmically on other parameters, such as the precision and system size). Remarkably, each local Lindbladian is defined by local Hamiltonian patches, avoids global Hamiltonian simulation, and may lead to qualitatively better resource estimate.

- **Physical origins.** The Lindbladian takes a similar form as Davies'-like generator and reduces to Davie's generator (and hence Glauber dynamics for classical Hamiltonians) in certain parameter regimes.

That is, our construction circumvents the aforementioned challenges and retains the celebrated features of classical MCMC algorithms that fueled its widespread impact. We expect our new construction to readily signify a multitude of new directions:

- **A provable quantum simulation algorithm.** There has been a flourishing industry to study spectral gap and mixing time of classical Markov chains

(e.g., [110]). In some cases, a sharp theoretical proof of algorithmic efficiency can even be obtained. Our construction gives a well-defined, quasi-local, and detail-balanced map that enables such a study in noncommuting Hamiltonians. Indeed, recently, rapid mixing in lattice Hamiltonians at high temperatures has been proved using our Lindbladian [153]. In chapter 4, we also use the established framework to give implicit evidence of quantum advantage for the "coolable" states through a dedicated mixing time analysis. We hope that new theoretical proof enabled in the quantum case could "unstick" the current inconclusive status of quantum advantage in quantum simulation by unveiling new provably efficient instances.

- **A dynamics angle to quantum Gibbs states.** The mixing time of the Lindbladian is deeply connected to the correlation and complexity of quantum Gibbs states. It was a celebrated result in classical Gibbs sampling that rapid mixing coincides with the decay of Gibbs correlation [124], and only recently has the equivalence been extended to some commuting quantum Hamiltonians [31, 96]. Our newly defined map gives a new dynamic angle to rigorously study intricate quantum correlation, such as the area-law, conditional mutual information, and topological order, in non-commuting Gibbs states.

- **A self-contained model of open system thermodynamics.** As a quantum analog of Glauber dynamics, the physical origin of our construction gives a succinct model of open-system physics. Our Lindbladian enables a precise definition of *dynamical thermal phase transitions* associated with the mixing time. Traditionally, quantum statistical mechanics have largely focused on static properties (e.g., divergence of correlation length as an indicator of phase transitions), but we argue that a dynamical picture might yield new physics that is not visible in the statics. Related concepts include metastable states, the energy landscape, quantum spin glass, and self-correcting quantum memories, whose precise formulation for noncommuting Hamiltonian has also been lacking (see chapter 4). The explicit form of our Lindbladian also enables direct numerical studies regarding the above notions.

To reiterate, given the celebrated theoretical and empirical triumph of Markov chain Monte Carlo methods and its successors over the past 70 years, quantum Gibbs sampling could serve similar roles in quantum computing. Especially given the current skepticism on the practical applicability of quantum computers, we aspire to

bring hope to the community by initiating a new wave of directions covering theory, experiment, numerics, and application. The remaining chapters of the thesis are organized as follows.

In chapter 2, we introduce the algorithmic and theoretic foundation for the simulation and analysis of a family of Davies-like Lindbladians (not necessarily detailed-balanced). The main algorithmic result is a black-box simulation algorithm for Davies-like Lindbladians (in fact any Lindbladians with substantially large Kraus rank) and the *operator Fourier transform*, a subroutine that decomposes any operator according to the changes in energy (akin to how phase estimation decomposes the states by its energies). The main analytic result is that Nature's cooling process, modeled by the existing Davies'-like Lindbladian, indeed converges to approximate Gibbs states; we introduce and prove a notion of approximate detailed balance that depends on the temperature, Hamiltonian simulation time, and the mixing time for the Lindbladian.

In chapter 3, we give the advertised detailed balanced Lindbladian as a particular instance of the Davies'-like family introduced in chapter 2, whose analytic and algorithmic properties are largely inherited and improved. The new key insight is a recipe for a *coherent* term in addition to the dissipation part that perfectly cancels out deviations from quantum detailed balance. With exact detailed balance, the proof of Gibbs stationarity is immediate, and the Lindbladian naturally corresponds to a parent Hamiltonian (akin to the quantum walk operator for classical Markov chains) that allows for preparation for the purified Gibbs state. We do not know where there is a qualitative distinction from Davies'-like Lindbladians introduced in chapter 2.

In chapter 4, we provide implicit evidence of quantum advantage in cooling by showing that Nature's cooling process is just as powerful as universal quantum computation. In particular, we construct a family of 2D Hamiltonians whose ground states encode arbitrary polynomial-time quantum computations, and the associated Davies'-like Lindbladian converges to the ground state in polynomial time. Therefore, there are instances of ground state problems that are quantumly easy and classically hard unless $BQP = BPP$ (roughly, unless classical computers can efficiently simulate all decision problems solvable with quantum computations). To study the mixing times, we introduce the idea of *quantum local minima*, which attempts to capture an energy landscape associated with the cooling process. Technically, the results are proven for physically derived Davies-like Lindbladian building upon the framework from chapter 2, but we expect an even faster convergence to hold with better

parameters for the Lindbladian with exact detailed balance from chapter 3.

Chapter 2

ALGORITHMIC PRIMITIVES FOR QUANTUM THERMAL STATE PREPARATION

2.1 Introduction

How do we prepare quantum Gibbs or ground states on a quantum computer? This initial state preparation problem appears as the obstacle for simulating quantum systems [68, 116]– a popular candidate for practical quantum advantage. This mystery has its roots in the seemingly contradictory teachings of computer science and physics: computational complexity theory tells us that few-body Hamiltonian ground states are generally QMA-hard to prepare [4, 76, 101] and thus are likely intractable in general even for quantum computers; on the contrary, thermodynamics asserts that physical systems interacting with a thermal bath are naturally in the thermal states or ground states. How do we draw an appropriate boundary between the two cases?

Practically, recent end-to-end industry resource estimates (e.g., [14, 30, 37, 107]) of quantum simulation rely on initial state preparation assumptions[1], exposing our ignorance of the complexity of practically relevant states. Often, practitioners turn to heuristic algorithms such as the Variational Quantum Eigensolver (see, e.g., [169]) or the adiabatic algorithm (see, e.g., [7, 64]), yet each with concerns for practicality. The former suffers from the so-called Barren Plateau phenomena [128], and its scalability has been debated; the latter requires a gapped adiabatic path, which appears nontrivial in recent large-scale numerical studies for quantum chemistry applications [108]. So far, there is a thin consensus on a 'go-to' ground state or thermal state quantum algorithm that could work in practice.

This work approaches the state preparation problem via *Quantum Gibbs samplers*. In physics language, this is closely related to open system dynamics where the system of interest is coupled to a thermal bath (see, e.g., [152]). Here, the conceptual boundary is blurred between the underlying physical process and the algorithm [168]. If a system thermalizes in nature and our physical model is accurate, we expect the associated quantum Gibbs sampler to converge quickly (i.e., the mixing time or the

[1] More precisely, they assume the existence of *trial states* with good overlap with the ground state so that running phase estimation provably works [114]. See also [70].

inverse-spectral gap is small); conversely, proving the latter also gives a rigorous formulation of open-system thermodynamics. This complements the mainstream formulation of closed-system thermodynamics via the Eigenstate Thermalization Hypothesis (see, e.g., [62]), where theoretical progress has been elusive. Practically, our general analysis for open system thermalization could be relevant to analog quantum simulators for Gibbs sampling, although our presentation mainly focuses on fault-tolerant quantum computers.

In computer science language, quantum Gibbs samplers are the quantum analogs of classical Markov chain Monte Carlo (MCMC) algorithms, most notably Metropolis sampling (see, e.g., [110]). They proved to be an indispensable pillar in classical computer science, both theoretically and practically, for computational physics and, more recently, optimization problems and machine learning. In a nutshell, the simple yet general idea is a (discrete or continuous) Markov chain whose unique fixed point yields the target distribution π (a vector with positive entries); given the energy E_s as a function of the configuration s, the Markov chain's transition matrix $\boldsymbol{M}$ satisfies

$$\boldsymbol{M}\pi = \pi \quad \text{where} \quad \pi_s := \frac{\mathrm{e}^{-\beta E_s}}{\sum_s \mathrm{e}^{-\beta E_s}} \quad \text{for each} \quad s \quad \text{at temperature} \quad \frac{1}{\beta}.$$

The algorithmic cost for preparing a sample from the Gibbs distribution π scales directly with the *mixing time*, the number of iterations such that any initial conjugation converges to the stationary distribution π. The mixing time can be unpredictable and vary wildly depending on the specific problems (e.g., [110]). Theoretically, rapid mixing can sometimes be proven under suitable assumptions, most notably in lattice Ising models assuming exponential decay of Gibbs state correlation (see, e.g., [124]). Practically, even when mixing time estimates are elusive, MCMC algorithms often serve as a starting point for more sophisticated algorithms. Given the triumphant impact of classical Gibbs sampling, we argue that Quantum Gibbs samplers have been thus far underexplored in the community and will likely play a central role when more robust quantum computers become available. Indeed, in addition to quantum simulation, quantum Gibbs sampling has been identified as a key subroutine in solving semidefinite programs (SDPs) [11, 27] and quantum machine learning [9]. To clarify, we will focus on quantum Hamiltonians; quantum algorithms for classical Gibbs states are not in the scope of this work.[2]

[2]Quantum Gibbs sampler for fast-forwardable Hamiltonian (including commuting Hamiltonians) is already well-defined since one can effectively apply phase estimation to exponential accuracy [167, 180]. The challenges we confront in this work are rooted in the noncommutativity.

To set the stage for quantum Gibbs sampling, we shall review the classical cousins, which we consider the seminal Metropolis-Hastings algorithm (see, e.g., [110]) as a representative. This algorithm iterates a *discrete-time Markov chain* as follows: apply a random "jump" (or "update," "move") $\boldsymbol{A}^a$ with probability $p(a)$ (Figure 2.1). If the energy decreases, *accept* the move, otherwise accept only with probability $\mathrm{e}^{-\beta\omega}$ (i.e., rejecting the move with probability $1-\mathrm{e}^{-\beta\omega}$), where ω being the energy gain. This can be described as a stochastic matrix over pairs of configurations $s's$

$$\boldsymbol{M}_{s's} := \underbrace{\sum_{a\in A} p(a)\gamma(E_{s'}-E_s)\boldsymbol{A}^a_{s's}}_{\text{"Accept"}} + \underbrace{\boldsymbol{R}_{s's}\delta_{s's}}_{\text{"Reject"}} \quad \text{where} \quad \gamma(\omega) := \min(1, \mathrm{e}^{-\beta\omega}), \tag{2.1}$$

and $\boldsymbol{A}^a_{s's}$ are stochastic matrices corresponding to each move (e.g., flipping one of the spins). The matrix elements are weighted by the Metropolis factor $\gamma(\omega)$ depending on the energy change. Importantly, the particular function satisfies a particular symmetry (Figure 2.1), known as the *detailed balance* condition

$$\gamma(\omega)/\gamma(-\omega) = \mathrm{e}^{-\beta\omega} \quad \text{such that} \quad \boldsymbol{M}_{s's}\pi_s = \pi_{s'}\boldsymbol{M}_{s's} \quad \text{for each} \quad s, s'. \tag{2.2}$$

Detailed balance ensures that the Gibbs state π is a fixed point of the Markov chain $\boldsymbol{M}\pi = \pi$. The rejection part $\boldsymbol{R}_{s's}$ is a diagonal matrix determined by the probability preserving constraints. Similarly, one may define a *continuous-time Markov chain generator*

$$\boldsymbol{L}_{s's} := \sum_{a\in A} p(a)\left(\underbrace{\gamma(E_{s'}-E_s)\boldsymbol{A}^a_{s's}}_{\text{"transition"}} - \underbrace{\delta_{s's}\sum_{s''}\gamma(E_{s''}-E_s)\boldsymbol{A}^a_{s''s}}_{\text{"decay"}}\right) \quad \text{for each} \quad s, s'. \tag{2.3}$$

The second term ensures that the generated semi-group $\mathrm{e}^{\boldsymbol{L}t}$ preserves probability. The operators $\boldsymbol{A}^a$ can be arbitrary nonnegative matrices and need not be stochastic.

While classical Markov chain Monte Carlo methods have been theoretically and practically mature, the quantum analogs are still in their infancy. The study of Quantum Gibbs sampling currently faces fundamental challenges; surprisingly, even a satisfactory map has not been appropriately defined for general noncommutative Hamiltonians. As the very first step, we need to algorithmically design a quantum

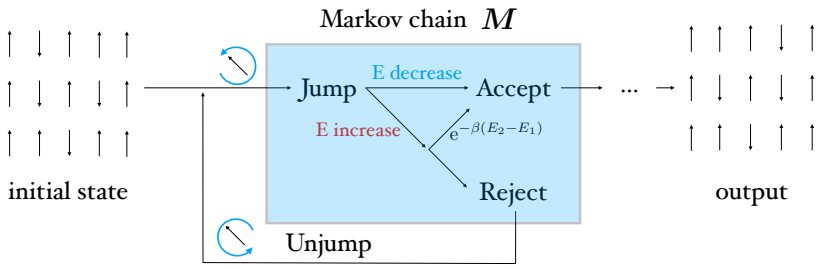


Figure 2.1: The Metropolis-Hastings algorithm iterates a Markov chain to sample from the Gibbs distribution. Each step begins with a (random) jump: if the energy decreases, accept; if the energy increases, accept with a carefully chosen probability. Otherwise, reject the move. Remarkably, detailed balance can be enforced in a "lazy" manner via rejection sampling without storing the whole matrix.

analog of Markov chain generator[3] $\mathcal{L}_\beta$, a *Lindbladian*, [181] whose (unique) fixed point is the *quantum Gibbs state* $\boldsymbol{\rho}_\beta$. More precisely, given a Hamiltonian $\boldsymbol{H}$ and an inverse temperature β

$$\text{design} \quad \mathcal{L}_\beta \quad \text{such that} \quad e^{\mathcal{L}_\beta t}[\boldsymbol{\rho}_\beta] = \boldsymbol{\rho}_\beta \quad \text{where} \quad \boldsymbol{\rho}_\beta := e^{-\beta \boldsymbol{H}}/\mathrm{Tr}(e^{-\beta \boldsymbol{H}}) \tag{2.4}$$

for any $t > 0$. Subsequently, we may begin studying the properties of the proposed Lindbladian, especially the mixing time. This work aims to lay the foundation for the first challenge. The second challenge was partially addressed in Ref. [38] using more primitive Gibbs samplers. Unlike the classical case, the construction of quantum Gibbs samplers is nontrivial due to imprecise energy estimates for noncommuting Hamiltonians (i.e., the energy-time uncertainty principle); the fixed point would not be exactly the Gibbs state (2.4). Previous attempts [38, 167, 168, 180, 186] have their shortcomings, which we discuss in more detail in section 2.1 and Table 2.1. Our work, in parallel with the recent paper [149], provides the first implementable Lindbladian for Gibbs sampling, with provable guarantees and without unrealistic assumptions. To do so, we introduce a robust analytic framework, which additionally applies to physical Lindbladians derived in open systems and to *coherent* Gibbs samplers with Szegedy-type speedups.

Our particular construction draws inspiration from thermalization in nature. As the starting point, a system in thermal contact with a bath can be effectively described by the so-called *Davies generator* in the Schrödinger Picture in a specific (weak-

[3] We focus on the infinitesimal generators for simplicity. One may alternatively consider discrete quantum channels, also known as completely-positive-trace-preserving (CPTP) maps.

coupling/infinite-time) limit ([56, 58, 152], and see [134] for a modern discussion)

$$\mathcal{L}_{\mathrm{Davies}}(\boldsymbol{\rho}) = \sum_{a\in A}\sum_{\nu\in B}\gamma(\nu)\left(\underbrace{\boldsymbol{A}^a_\nu \boldsymbol{\rho}(\boldsymbol{A}^a_\nu)^\dagger}_{\text{“transition”}} - \underbrace{\frac{1}{2}((\boldsymbol{A}^a_\nu)^\dagger \boldsymbol{A}^a_\nu \boldsymbol{\rho} + \boldsymbol{\rho}(\boldsymbol{A}^a_\nu)^\dagger \boldsymbol{A}^a_\nu)}_{\text{“decay”}}\right), \tag{2.5}$$

where $\{\boldsymbol{A}^a\}_{a\in A}$ are the set of “quantum” jumps and $\nu \in B := \mathrm{spec}(\boldsymbol{H}) - \mathrm{spec}(\boldsymbol{H})$ are the *Bohr frequencies*, the set of energy differences of the Hamiltonian. This resembles its classical Markov chain cousin (2.3), featuring two terms: transition and decay rates. Since we work with density operators $\boldsymbol{\rho}$ instead of probability vectors, the input $\boldsymbol{\rho}$ must be formally sandwiched by operators on the left and right. However, if the input states are diagonal in the energy basis and the energy levels are nondegenerate, then the Davies' generator can be faithfully represented as a continuous-time Markov chain (2.3) on the energy eigenstates by *literally* replacing

$$\begin{aligned} s &\to |\psi_i\rangle\langle\psi_i|, \\ \boldsymbol{A}^a_{s's} &\to \left|\langle\psi_i|\boldsymbol{A}^a|\psi_j\rangle\right|^2 \quad \text{where} \quad \boldsymbol{H} = \sum_i E_i|\psi_i\rangle\langle\psi_i|. \end{aligned} \tag{2.6}$$

A concrete physical example is a geometrically local Hamiltonian on a lattice. The jump operators $\boldsymbol{A}^a$ can be one-body Pauli operators on each lattice site, and the cardinality of jumps scales with the system size $|A| \propto n$. Of course, Davies' generator is merely an instance of many possible Lindbladian one can write down (see section 2.2), which generally may not come from thermodynamics.

In (2.5), the quantum mechanical transition rate $\boldsymbol{A}^a_\nu$ is defined as

$$\boldsymbol{A}^a_\nu := \sum_{E_i - E_j = \nu} \boldsymbol{P}_{E_i} \boldsymbol{A}^a \boldsymbol{P}_{E_j} \quad \text{for each Bohr frequency} \quad \nu \in B,$$

where $\boldsymbol{P}_E$ denotes energy eigenspace projectors associated with energy E (Figure 2.2). In general, the dynamics can be inherently quantum-mechanical when the energy subspaces are degenerate; the quantum transition can include *coherent* rotations within the subspaces $\boldsymbol{P}_E$.

The function $\gamma(\omega)$ depends on the physical model of the bath. Under physical assumptions (thermal bath and Markovianity), the heating transitions are penalized by a Boltzmann factor relative to the cooling transitions[4] $\gamma(\omega) = \mathrm{e}^{-\omega\beta}\gamma(-\omega)$.

[4]The sign convention we use (here and also for Fourier Transforms) might differ from that of other works in the open systems literature.

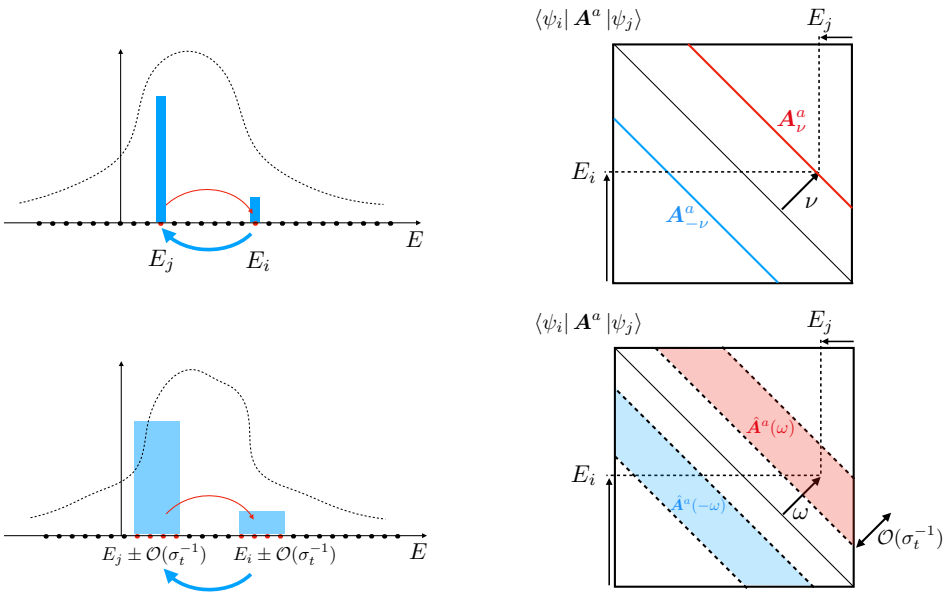


Figure 2.2: (Up) Davies' generator gives a continuous-time Markov generator on the energy spectrum (assuming the Hamiltonian is nondegenerate and the input state is diagonal in the energy basis.). The transitions are weighted by $\gamma(\omega)$: the heating transitions (red) are suppressed by a Boltzmann factor relative to the cooling transitions (blue), entailing detailed balance. The operator $\boldsymbol{A}^a_\nu$ contains the transitions with energy difference ν, which requires an infinite-time Fourier Transform. (Down) Our Lindbladian Gibbs sampler can be considered a "semi-classical" random walk where nearby Bohr-frequencies $\omega \pm \mathcal{O}(\sigma_t^{-1})$ cannot be distinguished. The operator Fourier Transform $\hat{\boldsymbol{A}}^a(\omega)$ contains a band of transitions. This breaks the detailed balance condition; the fixed point deviates from the Gibbs state.

Mathematically, this ensures that the Davies' generator $\mathcal{L}_{Davies}$ satisfies the *quantum detailed balance condition* (section 2.2) with respect to the Gibbs state $\boldsymbol{\rho}_\beta$, implying that $\boldsymbol{\rho}_\beta$ is a fixed point (2.4). At first glance, the above properties of Davies generators (2.5) seem to qualify for a Quantum Gibbs sampler (2.4) - if we were able to simulate it efficiently.

Unfortunately, the generator (2.5) is generally nonphysical because isolating an exact energy transition ν requires an *infinite-time* Fourier Transform[5] over Heisenberg evolution

$$\boldsymbol{A}^a_\nu \propto \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} \underbrace{\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^a \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}}_{\boldsymbol{A}^a(t) :=} \mathrm{d}t \quad \text{for each} \quad a \in A \quad \text{and} \quad \nu \in B.$$

[5]With the exception when the Hamiltonian spectrum takes well-separated discrete values with (roughly) known locations.

This allows the Davies generator to decohere arbitrary close Bohr frequencies ω, ω'. This contradicts the energy-time uncertainty principle, requiring the runtime to scale inversely with the level spacing, which generally can be *exponentially* small in the system size n. Unfortunately, at shorter times, the resulting fuzziness of energy resolution breaks detailed balance, which has been central to the analysis of (both classical and quantum) Markov chains. Many finite-time versions of the Davies generator have been proposed to capture more realistic physical settings [38, 121, 134, 150]. Yet, to the best of our knowledge, no Lindbladian arising from a finite-time Fourier Transform has been proven to have a fixed point close to the Gibbs state.[6]

Quantum Gibbs sampling algorithms face analogous technical challenges due to a limited algorithmic runtime. Here, the energy-time uncertainty principle incarnates as the statistical uncertainty of the energy measurement via quantum phase estimation. Several works [148, 180] evade this issue by imposing a convenient *rounding promise* on the Hamiltonian, requiring its spectrum to be disjoint from certain specific ranges of energy. Such a rounding promise is not physically motivated and does not hold in general but allows for rigorous performance guarantees for Gibbs samplers using boosted phase estimation. Very recently, Ref. [149] circumvents the need for a rounding promise by randomly alternating the phase estimation mesh, but this comes at a high additional algorithmic cost with potentially worsened mixing time due to forbidding certain transitions and seems distant from the physical origins of quantum Gibbs samplers.

In this work, we present quantum Gibbs sampling algorithms inspired by thermalization in Nature. In particular, our construction is a "smoothed" version of the Davies' generator (2.5). The first algorithm simulates a Lindbladian whose fixed point is approximately a quantum Gibbs state; the second algorithm further "quantizes" the Lindbladian to prepare the purified Gibbs state and features a quadratic Szegedy-type speedup. Our algorithms are efficient and comparatively simple to implement while having a provable performance guarantee. The key ingredient in our algorithmic design is to use the weighted *operator Fourier Transform* for the Lindblad operators

$$\hat{\boldsymbol{A}}^a(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} f(t) \boldsymbol{A}^a(t) \mathrm{d}t \quad \text{for each} \quad a \in A \quad \text{and} \quad \omega \in \mathbb{R}. \tag{2.7}$$

In practice, we use a discrete Fourier Transform (which will be denoted by $\hat{\boldsymbol{A}}^a(\bar{\omega})$ for discrete frequency label $\bar{\omega}$). Still, for conceptual simplicity, we focus on

[6] Ref. [38] gives a *nonCPTP* generator that does have approximately a Gibbs fixed point.

the continuous case in the introduction. Unlike in ordinary phase estimation where boosting usually adopts median-of-means tricks [138], we weigh the Fourier Transform by a Gaussian distribution $f(t)$ with a tunable width σ_t. Consequently, the Fourier Transform remains a Gaussian, achieving an analog of a boosted phase estimation with uncertainty $\omega \pm O(\sigma_t^{-1})$.

To give performance guarantees for our construction, our main analytic insight is to define a notion of *approximate detailed balance* (see section 2.7)

$$\boldsymbol{\rho}_\beta^{1/4} \mathcal{L}^\dagger[\boldsymbol{\rho}_\beta^{-1/4} \boldsymbol{O} \boldsymbol{\rho}_\beta^{-1/4}] \boldsymbol{\rho}_\beta^{1/4} \approx \boldsymbol{\rho}_\beta^{-1/4} \mathcal{L}[\boldsymbol{\rho}_\beta^{1/4} \boldsymbol{O} \boldsymbol{\rho}_\beta^{1/4}] \boldsymbol{\rho}_\beta^{-1/4} \quad \text{for each operator} \quad \boldsymbol{O}$$

to handle energy uncertainty. In the exact case, this is the quantum generalization of classical detailed balance, where the distribution becomes an operator, and the Markov transition matrix becomes a superoperator.

Our construction and analysis of the Gibbs sampler are physically inspired insofar as it closely resembles the Davies generator of Eqn. (2.5), but we do not know whether it can be derived in some physical limit from a weak system-bath coupling. Incidentally, starting from a microscopic system-bath interaction, a recent proposal [121, 134] specifically derives from first principles a Lindbladian with Lindblad operators

$$\hat{\boldsymbol{A}}^a(\omega) :\propto \int_{-T/2}^{T/2} \mathrm{e}^{-\mathrm{i}\omega t} \boldsymbol{A}^a(t) \mathrm{d}t \quad \text{for each} \quad a \in A \quad \text{and} \quad \omega \in \mathbb{R}.$$

Here, the Fourier Transform time-scale T sets the energy uncertainty $\omega \pm O(T^{-1})$; the fixed point will not be exactly the Gibbs state. Using our new analytic framework, we show that this Lindbladian derived in Refs. [121, 134] has a stationary state close to the Gibbs state. To the best of our knowledge, our work completes the first general proof of many-body Gibbs states in the open system setting (assuming a reasonably short mixing time).

The runtime of both our algorithms has simple dependence on the mixing time or spectral gap of the Lindbladian. In general, the gap will depend sensitively on the details of the physical system, and its calculation for specific Hamiltonians is beyond the scope of this work. We refer to Refs. [31, 38, 96, 97] and references therein for a more detailed discussion of mixing times (and spectral gaps) for various Hamiltonian and Lindbladians. At an intuitive level, we expect the Lindbladian for lattice systems with jump operators on each site to have a constant local Lindbladian gap[7] independent of the volume at high enough temperatures or within the same

[7]The local gap, in our normalization, is defined as $n \cdot \lambda_{gap}(\mathcal{L})$, where $\lambda_{gap}(\mathcal{L})$ is the Lindbladian

phase. In practice, we believe that quantum Gibbs sampling algorithms will be employed on a case-by-case basis in combination with various heuristics, as is the case with classical Monte Carlo algorithms.

Existing work

The first attempt at designing a quantum algorithm for Gibbs sampling with per-update efficiency guarantees is the quantum Metropolis algorithm [167], as a quantum analog of (2.1). The main guiding principle is to "do Metropolis sampling over the energy spectrum" in the spirit of (2.6). To do so, we need two quantitative changes due to quantum mechanics, one algorithmic and one analytic: algorithmically, one needs a subroutine to "reject" a quantum state back to the same energy. Classically, one can clone the configuration s before the update. Quantumly, however, we cannot clone the (unknown) quantum state and must be careful not to collapse the quantum state due to energy measurement. Ref. [167] handles this issue using the Mariott-Watrous [123] algorithm (or the "rewinding" technique in quantum Cryptography). Crucially, this algorithmic subroutine preserves probability, drawing a distinction from imaginary time evolution or post-selection, but this comes with its limitation and significantly complicates the algorithm. Second, energy measurements based on quantum phase estimation have a finite resolution inversely proportional to the runtime $\delta E \sim 1/T$. Consequently, the detailed balance condition may not generally hold, and one needs to prove that the fixed point remains approximately the Gibbs state.

The technical results of [167] contain three approaches based on different phase estimation subroutines: (i) assuming perfect phase estimation (with performance guarantee but with an exponential Hamiltonian simulation time); (ii) un-boosted phase estimation (without performance guarantees); (iii) boosted shift-invariant phase estimation (with performance guarantees).[8] Unfortunately, we recently realized that such a boosted, shift-invariant phase estimation (see section 2.13) is provably impossible;[9] we do not know whether Quantum Metropolis sampling [167], as stated explicitly, actually works in practice. The quantum metropolis is regarded as an important theoretical milestone, but due to its complicated form (especially due to the rejection subroutine), the particular algorithm largely serves as a high-level inspiration.

eigenvalue gap and n is the system size. A parallel version of the algorithm could, in principle, gain this n factor in the circuit depth. Also, the Lindbladian gap should not be confused with the energy gap of the Hamiltonian, which is not directly relevant to Gibbs sampling at nonvanishing temperatures.

[8] The approximate detailed balance argument was later completed in [38].

[9] The authors [167] communicated with us that there might be ways to fix their algorithm.

From a physical point of view, one may implement Nature's quantum algorithm [168] by emulating the global system-bath interaction. However, this black-box approach is a double-edged sword: indeed, this method should work as well as Nature, but as we know from open system thermodynamics, nonasymptotic results are extremely challenging without liberal use of approximations [152, 168], rendering the result qualitative but not quantitative. (For example, it is elusive how big of a bath is needed for the desired accuracy.) Recently, Ref. [38, 134, 160] took the physics inspirations seriously and quantitatively studied a system-bath interaction from scratch. Ref. [134] revisits the text-book open system derivation and extracts a *nonasymptotic* version of Davies' generator with explicit error bounds and without unphysical limits. However, it was not known whether the derived Lindbladian has the Gibbs state as the stationary state; ref. [38] was the first provably polynomial-time algorithm for Gibbs state assuming a reasonably short mixing time, although it assumes good control of the bath and its error bounds are large polynomials and impractical to apply; ref. [160] is conceptually similar to [38] but focuses more on near-term feasibility. Technically, its accuracy guarantees require the Eigenstate Thermalization Hypothesis, which significantly simplifies the analysis. Unfortunately, both cases [38, 160] failed to extract a Lindbladian (the generators are not completely positive). In some sense, this motivates us to give a unifying conceptual and analytic perspective on this subject.

Coherent quantum Metropolis sampling [186] is a natural generalization of quantum Metropolis sampling [167] that further gives a quadratic runtime speedup by invoking Szegedy's quantum walk strategy [164]. Since the dissipative map is quantized as a Hermitian operator, one cannot evolve a semi-group but requires an additional *quantum simulated annealing* step [161] (a particular adiabatic state preparation along temperatures) to prepare the *purified Gibbs state; see section 2.12*. Unfortunately, Ref. [186] assumes perfect phase estimation and a nondegenerate Hamiltonian spectrum, and it was unclear how one incorporates imperfect phase estimation in such a coherent algorithm. Ref. [180] improves and generalizes their result but still makes an unphysical *rounding promise* assumption: the Hamiltonian spectrum has periodic gaps to amplify phase estimation.

Perhaps inspired by the rounding promise, the recent related work [149] proposes an algorithm that implements a Lindbladian that also provably has approximately Gibbs fixed point using randomized rounding. Their approach is quite different in nature from ours, and they are not known to enjoy the quadratic speedup. Randomized rounding seems to incur large resource overhead (See Table 2.1[7]) and substantially

departs from the physical origin of these ideas.

Our Lindbladian Gibbs samplers build upon the literature for open system simulation [49] as well as the coherent Gibbs sampler of [180]. In both cases, we remove the need for any unphysical assumption (e.g., especially the rounding promise) yet still maintain a simple error bound. This is made possible by identifying the right choice of jump operators in terms of discrete Fourier Transform and refining the analytic technical tool introduced in [38] (nonasymptotic bounds for secular approximation)[10]. In a nutshell, it seems the "right" approach to quantum Gibbs sampling is to simulate a (continuous-time) Lindbladian, which Nature implements by default, instead of a (discrete-time) Metropolis-Hastings style quantum channel. The rejection step is handled automatically for any Lindbladians. We leave for future work to simplify the rejection step in quantum Metropolis sampling [167] or to design a discrete-time channel with provable guarantees.[11] We further note a general distinction between coherent Gibbs samplers and Lindbladian Gibbs samplers regarding obtaining the fixed point. The former relies on the gap of the Lindbladian staying open along the entire adiabatic path from high to low temperature $0 \to \beta$, while the latter (e.g., our construction or [149]) does not. Thus, the two algorithmic costs are not directly comparable. Empirical intuition from classical Gibbs sampling suggests that for "simple problems" where the gaps remain largely open throughout the phase of interest, adiabatic and direct sampling methods perform similarly. However, for strongly frustrated systems like spin glasses, adiabatic heuristics are the go-to Monte Carlo method. It is tempting to speculate that the same will be true in the quantum Gibbs sampling case, which lends itself well to our approach.

In addition to Monte Carlo style algorithms, other thermal state preparation algorithms based on quite different principles also exist; we briefly summarize their gate complexities in Table 2.1. We only discuss methods where quantitative arguments are possible and pay less attention to heuristic approaches such as variational circuits [51, 125, 158, 184, 188]), energy filtering assuming good initial states, and heuristic quantum assisted Monte Carlo [119, 157].

[10] Some preliminary version of approximate detailed balance was discussed in an earlier version of [38] regarding quantum Metropolis sampling. That part was completely removed after the authors realized the phase estimation assumption was impossible as stated (section 2.13).

[11] For Lindbladians, the designer has the freedom to choose *arbitrary* Lindblad operators, and the decay part automatically guarantees trace-preserving. However, it is more challenging to design quantum channels as the trace-preserving condition seems less flexible. Indeed, Quantum Metropolis Sampling [167] had to invoke Mariot-Wattrous-style rewinding [123] multiple times to ensure probability is preserved, which unfortunately increases the complexity of the algorithm.

Outline and main results

Our discussion features the following Lindbladian in the Schrödinger picture

$$\mathcal{L}_\beta[\boldsymbol{\rho}] := \sum_{a\in A}\int_{-\infty}^{\infty}\gamma(\omega)\left(\hat{\boldsymbol{A}}^a(\omega)\boldsymbol{\rho}\hat{\boldsymbol{A}}^a(\omega)^\dagger - \frac{1}{2}\{\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega),\boldsymbol{\rho}\}\right)\mathrm{d}\omega \tag{2.8}$$

with the anti-commutator $\{\boldsymbol{A},\boldsymbol{B}\} = \boldsymbol{A}\boldsymbol{B} + \boldsymbol{B}\boldsymbol{A}$. We can read out the set of *Lindblad operators*

$$\{\sqrt{\gamma(\omega)}\hat{\boldsymbol{A}}^a(\omega)\}_{a\in A,\omega\in\mathbb{R}} \quad \text{where} \quad \hat{\boldsymbol{A}}^a(\omega) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\omega t}f(t)\boldsymbol{A}^a(t)\mathrm{d}t. \tag{2.9}$$

In particular, all Lindbladians we consider in this work, natural or algorithmic, satisfies the following symmetry and normalization conditions:

- The set of *jump operators* $\boldsymbol{A}^a$, which "drives" the transition, can be arbitrary (and often depends on the Hamiltonian) as long as the set contains their adjoints

$$\{\boldsymbol{A}^a : a\in A\} = \{\boldsymbol{A}^{a\dagger} : a\in A\} \quad \text{and} \quad \|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \le 1. \tag{2.10}$$

  Indeed, classical Metropolis sampling often starts with a *reversible* Markov chain to algorithmic impose the detailed balance condition; the quantum analog is the adjoint condition. Single-site Pauli operators (which are individually self-adjoint) are handy choices, but few-body operators with arbitrary connectivity are certainly permissible[12]. The normalization is natural for block-encoding the set of jump operators (Definition 2.1.2 and (2.27)). For example, choosing single-site Paulis as jump operators requires dividing them by $\sqrt{|A|}$ (where $|A|$ is the cardinality of the set) to fulfill the normalization requirement.[13]

- The Fourier Transform in the time domain is weighted by a *filter* function $f(t)$ that is real and ℓ_2-normalized

$$f^*(t) = f(t) \quad \text{for} \quad t\in\mathbb{R} \quad \text{and} \quad \|f\|_2^2 := \int_{-\infty}^{\infty}|f(t)|^2\mathrm{d}t = 1. \tag{2.11}$$

  Sometimes, we drop the subscript by $\|f\| = \|f\|_2$. The real constraint serves a purpose similar to reversibility in classical Gibbs sampling. When considering discrete Fourier Transforms (which is necessary for implementation), we adapt the corresponding (discrete) normalization $\sum_{\bar{t}\in S_{t_0}}|f(\bar{t})|^2 = 1$.

[12] In fact, the ability to perform carefully chosen (often not natural) jumps is what empowers classical Gibbs sampling algorithms, e.g., cluster updates.

[13] This is slightly different from the physical setting where each jump has operator norm $\|\boldsymbol{A}^a\| = 1$. There, the "strength" of the Lindbladian (2.13) scales with the number of jumps $|A|$.

- The *transition weight* $\gamma(\omega)$ can be any function satisfying the *KMS condition* and the bound

$$\gamma(\omega)/\gamma(-\omega) = e^{-\beta\omega} \quad \text{and} \quad 0 \leq \gamma(\omega) \leq 1 \quad \text{for fixed} \quad \beta \quad \text{and each} \quad \omega \in \mathbb{R}. \tag{2.12}$$

  This coincides with the classical recipe for detailed balance (2.2). Natural choices include the Metropolis weight $\gamma(\omega) = \min(1, e^{-\beta\omega})$ or the (smoother) Glauber dynamics weight $\gamma(\omega) = (e^{\beta\omega} + 1)^{-1}$.

To summarize, the above list of symmetry conditions is the key to ensuring (approximate) detailed balance; the above normalization choices are not only natural for implementation but also conveniently ensures that the "strength" of the Lindbladian is normalized

$$\|\mathcal{L}_\beta\|_{1-1} \leq 2 \tag{2.13}$$

in the superoperator 1-1 norm.

Lindbladians from Nature

Before we discuss Gibbs sampling algorithms, we first address the fundamental question: why do Gibbs states faithfully capture physical systems in thermal equilibrium? In physics, the quantum Gibbs state is often imposed without rigorous justification. As a mathematical physics result, we complete the first proof of open system thermodynamics: the Gibbs state is indeed *approximately* the fixed point of Lindbladians governing open system dynamics.

Of course, this further requires a rigorous derivation of Lindbladian from reasonable open system assumptions; this is not the intention of this work, but thankfully, this has been worked out under a Markovian, weak-coupling assumption [134]. All we need as a black box is that it indeed satisfies the constraints we imposed (2.10),(2.11), and (2.12). For simplicity, we have omitted the Hamiltonian part of the Lindbladian and focus only on the dissipative part; see section 2.9 for the complete results.

Theorem 2.1.1 (Gibbs state is thermodynamic)**.** *Any $\mathcal{L}_\beta$ satisfying the symmetry and normalization conditions* (2.10),(2.11)*, and* (2.12) *with the particular weight function*

$$f(t) = \frac{\mathbb{1}(t \leq |T|/2)}{\sqrt{T}}$$

has an approximate Gibbs fixed point

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}_\beta) - \boldsymbol{\rho}_\beta\|_1 \leq O\left(\sqrt{\frac{\beta}{T}} t_{mix}(\mathcal{L}_\beta)\right). \tag{2.14}$$

In particular, (dropping the Hamiltonian term and under suitable normalization) such a Lindbladian can arise from a system (with Hamiltonian $\boldsymbol{H}$) interacting weakly with a Markovian bath (with inverse temperature β) through jump operators $\boldsymbol{A}^a$.

The above introduces the notion of *mixing time* for Linbladians: the time scale for which any pair of initial states become indistinguishable. Of course, the physical interpretation of this time scale depends on how the Lindbladian is normalized; for our cases, we conveniently have that $\|\mathcal{L}_\beta\|_{1-1} \leq 2$ (2.13).

Definition 2.1.1 (Lindbladian mixing time)**.** *For any Lindbladian $\mathcal{L}$, we define the mixing time $t_{mix}(\mathcal{L})$ in the Schrödinger picture to be the shortest time for which*

$$\left\|e^{\mathcal{L}t_{mix}}[\boldsymbol{\rho} - \boldsymbol{\rho}']\right\|_1 \leq \frac{1}{2}\|\boldsymbol{\rho} - \boldsymbol{\rho}'\|_1 \quad \textit{for any states} \quad \boldsymbol{\rho}, \boldsymbol{\rho}'.$$

Theorem 2.1.1 states that the approximation of Gibbs state degrades at a low temperature, a poor energy resolution (i.e., a short Fourier Transform time T), or a long mixing time. While the parameters β, T are tunable parameters; the mixing time is generally Hamiltonian dependent. Still, one may obtain a bound using additional assumptions (such as the decay of correlation of commuting Hamiltonian Gibbs states [16, 96] or the Eigenstate Thermalization Hypothesis [38, 160]), empirical intuition, or conversion from a numerically-obtained spectral gap (Proposition 2.2.3).

Assuming some grasp of the mixing time, how large should the time scale T get to obtain a Gibbs sample? Roughly, according to the error bound (2.14), the time T should scale with the mixing time by

$$T \sim \beta t_{mix}^2/\epsilon^2.$$

More carefully, the RHS (2.14) does not obviously permit a solution for the desired accuracy ϵ (similarly for Theorem 2.1.3 and Theorem 2.1.4). Indeed, the mixing time t_{mix} can generally depend on other parameters, especially the width T; heuristically, one may guess that the mixing time t_{mix} depends mildly on the width T, but we leave a careful analysis for future works. An optimistic instance is when the Eigenstate Thermalization Hypothesis holds, and the mixing time at finite energy resolution can be related to the infinite resolution ($T \to \infty$) case [38, 160], which can be calculated.

The main analytic challenge to prove Theorem 2.1.1 is how to control the convergence and fixed point of Lindbladians without exact detailed balance. Our technical contribution is to formulate an *approximate detailed balance* condition using the appropriate superoperator norm

$$\left\| \boldsymbol{\rho}_\beta^{1/4} \mathcal{L}^\dagger [\boldsymbol{\rho}_\beta^{-1/4} \cdot \boldsymbol{\rho}_\beta^{-1/4}] \boldsymbol{\rho}_\beta^{1/4} - \boldsymbol{\rho}_\beta^{-1/4} \mathcal{L} [\boldsymbol{\rho}_\beta^{1/4} \cdot \boldsymbol{\rho}_\beta^{1/4}] \boldsymbol{\rho}_\beta^{-1/4} \right\|_{2-2}. \tag{2.15}$$

The above two superoperators are each other's adjoints, and thus, the above quantifies the magnitude of certain *anti-Hermitian* component of the Lindbladian under similarity transformation. Indeed, traditionally, the detailed balance condition is convenient as it effectively reduces the mixing time of superoperators to the spectral theory of Hermitian operators, which is conceptually and technically more transparent. Our observation is that the consequences of detailed balance, including Gibbs fixed point and spectral bounds on mixing time, are remarkably stable against perturbation. Much ink is devoted to bound (2.15), which is yet another technical challenge. Indeed, we are *inverting* the Gibbs state, which has exponentially small weights. The energy uncertainty in the Fourier Transforms $\hat{\boldsymbol{A}}^a(\omega)$ could potentially blow up (2.15). In response, we further introduce an intermediate Lindbladian by applying a rigorous *secular approximation* (related to the *rotating wave approximation*), such that transitions with large energy deviation are truncated.

$$\hat{\boldsymbol{A}}^a(\omega) \approx \hat{\boldsymbol{S}}^a(\omega) \quad \text{such that} \quad \langle \psi_i | \hat{\boldsymbol{S}}^a(\omega) | \psi_j \rangle = 0 \quad \text{whenever} \quad \left| (E_i - E_j) - \omega \right| \gg 0.$$

The secular approximation interplays nicely with the operator Fourier Transform and should be widely applicable in the rigorous, nonasymptotic analysis of open-system Lindbladians.

Conceptually, there are two *opposite* ways to understand Theorem 2.1.1: pessimistically, the Gibbs state may not be physical if the mixing time is too long, and we might have to simulate the natural Lindbladian to understand its fixed point; optimistically, if the Gibbs state is indeed physical, we might ignore its physical origin and take a short-cut to design even more efficient Gibbs sampling algorithms. The two perspectives are individually addressed in the following sections.

Simulating Nature

Taking a step back from Gibbs sampling, how do we simulate open system dynamics in nature? This boils down to the task of Lindbladian simulation, which has been

studied largely restricted to the black-box setting [49]. However, we have an *explicit* Lindbladian in mind to simulate. As a result, we had to modify existing black-box input models to capture our Lindbladian (2.9); this also inspires us to design *even more* efficient Lindbladian simulation algorithm for our access model. First, we define how we want our Lindbladian to be block-encoded.

Definition 2.1.2 (Block-encoding of a Lindbladian)**.** *Given a purely irreversible Lindbladian*

$$\mathcal{L}[\rho] := \sum_{j\in J}\left(\boldsymbol{L}_j\rho\boldsymbol{L}_j^\dagger - \frac{1}{2}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\rho - \frac{1}{2}\rho\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\right),$$

we say that a matrix[14] $\boldsymbol{U}$ *is a block-encoding of the Lindblad operators* $\{\boldsymbol{L}_j\}_{j\in J}$ *if* [15]

$$(\langle 0^b| \otimes \boldsymbol{I}) \cdot \boldsymbol{U} \cdot (|0^c\rangle \otimes \boldsymbol{I}) = \sum_{j\in J} |j\rangle \otimes \boldsymbol{L}_j \quad \textit{for} \quad b, c \in \mathbb{N}.$$

Indeed, because of the many Lindblad operators $j \in J$, other access models are certainly valid (e.g., given block-encoding for each Lindblad operator $\boldsymbol{L}_j$ [49, 112]). Nevertheless, Definition 2.1.2 interplays nicely with our Lindbladian (especially the operator Fourier Transform) and the following efficient Lindbladian simulation algorithm.

Theorem 2.1.2 (Linear-time Lindbladian simulation, simplified)**.** *Suppose* $\boldsymbol{U}$ *is a unitary block-encoding of the Lindbladian* $\mathcal{L}$ *as in Definition 2.1.2. Let* $t > 1$ *and* $\epsilon \le 1/2$*, then we can simulate the map* $e^{t\mathcal{L}}$ *to error* ϵ *in diamond norm using*

$$\begin{aligned} & O\left(\left(c + \log(\frac{t}{\epsilon})\right)\log(\frac{t}{\epsilon})\right) && \textit{(resettable) ancilla qubits,} \\ & \widetilde{O}(t) && \textit{(controlled) uses of } \boldsymbol{U} \textit{ and } \boldsymbol{U}^\dagger, \\ \textit{and} \quad & \widetilde{O}(t(c+1)) && \textit{other two-qubit gates.} \end{aligned}$$

See Theorem 2.3.2 for the complete result and Theorem 2.3.1 for a simpler algorithm with suboptimal asymptotic scaling. Compared with the best existing results as sum-of-norm $\sum_{j\in J}\|\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\|$, we achieve a strictly better scaling with the norm-of-sum $\|\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\|$ when the Lindblad operators are altogether block-encoded as

[14] For implementation purposes $\boldsymbol{U}$ will be a unitary quantum circuit, but we also consider nonunitary block-encodings for the sake of analysis.

[15] In the first register, we could use any orthonormal basis, sticking to computational basis elements $|j\rangle$ is just for ease of presentation. Intuitively, one can think about b as the number of ancilla qubits used for implementing the operators $\boldsymbol{L}_j$, while typically $a - b \approx \log|J|$.

in Definition 2.1.2; if we are only given block-encodings for each Lindblad operators, we can always convert them to our input model (Definition 2.1.2) and recover the existing scaling $\sum_{j\in J} \|\boldsymbol{L}_j^\dagger \boldsymbol{L}_j\|$.

It remains to create a unitary block-encoding (Definition 2.1.2) for our particular Lindblad operators (2.9). Since our algorithms run on discrete qubits, our implementation requires discretizing the operator Fourier Transform, with a change of the notation

$$\hat{\boldsymbol{A}}^a(\omega) \to \hat{\boldsymbol{A}}^a(\bar{\omega}) := \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} f(\bar{t}) \boldsymbol{A}^a(\bar{t}) \quad \text{for each} \quad a \in A \quad \text{and} \quad \bar{\omega} \in S_{\omega_0}. \tag{2.16}$$

The discretized frequency and time labels $\omega \to \bar{\omega} \in S_{\omega_0} = \{0, \pm\omega_0, \cdots\}$ and $t \to \bar{t} \in S_{t_0} = \{0, \pm t_0, \cdots\}$ corresponds to the discrete Fourier Transform (section 2.5) using a finite grid of size $N = \left|S_{\omega_0}\right| = \left|S_{t_0}\right|$, which can be stored using $\lceil \log(N) \rceil$ additional ancillas.[16] While the discretization parameters are needed for explicit algorithmic implementation, conceptually, they merely incur *logarithmic* overhead in the runtime and ancillas (section 2.8). We may now concretely present our algorithmic goal: efficiently construct a block-encoding in the form

$$(\langle 0^b| \otimes I)\boldsymbol{U}(|0^c\rangle \otimes I) = \sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})}|\bar{\omega}, a\rangle \otimes \hat{\boldsymbol{A}}^a(\bar{\omega}). \tag{2.17}$$

Lemma 2.1.1 (Efficient block-encoding)**.** *In the setting of Theorem 2.1.1, a unitary block-encoding* $\boldsymbol{U}$ *for the (discretized) Lindblad operators (2.17) can be created using one query of* $\boldsymbol{Prep}$, $\boldsymbol{W}$, $\boldsymbol{V}_{jump}$, $\boldsymbol{QFT}$, *and*

$\mathcal{O}(T)$	*(controlled) Ham. sim. time for* $\boldsymbol{H}$
$n + 1 + \lceil \log_2(\lvert A\rvert) \rceil + \lceil \log_2(N) \rceil$	*(resettable) qubits.*

See Lemma 2.3.1 for the explicit construction. In perceivable usage, we expect the number of Fourier labels registers N to scale polynomially with all other parameters (section 2.8)

$$N \sim \mathrm{Poly}(n, \beta, \|\boldsymbol{H}\|, T, \epsilon^{-1}, t, |A|)$$

for a good approximation for the continuous Fourier Transform (2.9). Morally, our algorithm extracts the essential functionality of a Markovian bath (which naively

[16] We require $N\omega_0 \geq 4\|\boldsymbol{H}\| + 2/\beta$ to store all possible energy transitions.

may require a substantial number of qubits to implement [38, 160]) by merely polylogarithmic resettable ancillas.

To make the simulation cost transparent, we list the main circuit components required for implementation: the controlled Hamiltonian simulation

$$\sum_{\bar{t}\in S_{t_0}} |\bar{t}\rangle\langle\bar{t}| \otimes e^{\pm i\bar{t}\boldsymbol{H}},$$

the unitary for preparing the filter function in superposition

$$\boldsymbol{Prep} : |\bar{0}\rangle \to |f\rangle := \sum_{\bar{t}\in S_{t_0}} f(\bar{t})|\bar{t}\rangle,$$

the controlled rotation for transition weights

$$\boldsymbol{W} := \sum_{\bar{\omega}\in S_{\omega_0}} \begin{pmatrix} \sqrt{\gamma(\bar{\omega})} & -\sqrt{1-\gamma(\bar{\omega})} \\ \sqrt{1-\gamma(\bar{\omega})} & \sqrt{\gamma(\bar{\omega})} \end{pmatrix} \otimes |\bar{\omega}\rangle\langle\bar{\omega}|,$$

the quantum Fourier Transform $\boldsymbol{QFT}$, and the block-encoding $\boldsymbol{V}_{jump}$ of the jump operators $\sum_{a\in A}|a\rangle \otimes \boldsymbol{A}^a$. In practice, synthesizing the above incurs additional overhead but should be treated as an independent subroutine to study (see section 2.3); we expect Hamiltonian simulation to be the dominant source of cost, which we present[17] by the accumulated time for the (controlled) unitaries $e^{i\boldsymbol{H}\bar{t}}$.

The key idea behind implement the unitary block-encoding (2.17) is the *operator Fourier Transform* (section 2.3) as an alternative to phase estimation (Figure 2.5)

$$\mathcal{F}[\cdot] : |f\rangle \otimes \boldsymbol{O} \to \sum_{\bar{\omega}} |\bar{\omega}\rangle \otimes \boldsymbol{O}_{\bar{\omega}}$$

which is physically motivated, compatible with our analytic framework, and leads to simple explicit circuits.

Improving Nature

Suppose our goal is to prepare the Gibbs state, then according to Theorem 2.1.1 and Lemma 2.1.1 we may algorithmically simulate the physical Lindbladian till the mixing time. However, the Fourier Transform occurring in Nature (Theorem 2.1.1), in fact, has a "heavy tail" in the frequency domain; this uncertainty in energy may significantly contribute to the Gibbs state error. With full algorithmic freedom, can

[17] To obtain the end-to-end gate complexity, one should specify a Hamiltonian simulation subroutine (e.g.,[44, 116, 117]).

we do better? In this section, we simply tweak the Lindbladian by considering a nicer Fourier Transform weight

$$f(t) \propto \mathrm{e}^{-\frac{t^2}{4\sigma_t^2}} \quad \text{with Gaussian width} \quad \sigma_t.$$

The width sets the Hamiltonian simulation time scale $\sim \sigma_t$ and the energy resolution $\sim \sigma_t^{-1}$. The Gaussian distribution is particularly nice as it enjoys sharp concentration in *both* time and frequency domains; in principle, other normalized functions are also permissible, such as Kaiser-window functions [20, 127], as long as they can be efficiently generated in superposition $\sum_{\bar{t}\in S_{t_0}} |f(\bar{t})\rangle$, but we will stick to Gaussians for simplicity.

Theorem 2.1.3 (Approximate Gibbs fixed point)**.** *Any Lindbladian* $\mathcal{L}_\beta$ (2.8) *satisfying the symmetry and normalization conditions* (2.10),(2.11)*, and* (2.12) *with the normalized Gaussian weight*

$$f(t) \propto \mathrm{e}^{-\frac{t^2}{4\sigma_t^2}}$$

has an approximate Gibbs fixed point

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}_\beta) - \boldsymbol{\rho}_\beta\|_1 = \widetilde{O}\left(\frac{\beta}{\sigma_t} t_{mix}(\mathcal{L}_\beta)\right).$$

See section 2.2 for the proof. We briefly present the analogous block-encoding costs; the Gaussian width σ_t plays a similar role as the time scale T, and Gaussians exhibit a better scaling than (2.14) due to its sharp concentration.

Lemma 2.1.2 (Efficient block-encoding)**.** *In the setting of Theorem 2.1.3, a unitary block-encoding* $\boldsymbol{U}$ *for the (suitably discretized) Lindblad operators* (2.17) *can be constructed using one query each to* $\boldsymbol{Prep}$, $\boldsymbol{W}$, $\boldsymbol{V}_{jump}$, $\boldsymbol{QFT}$, *and*

$$\begin{aligned} &\tilde{O}(\sigma_t) && \textit{(controlled) Ham. sim. time for } \boldsymbol{H} \\ &n + 1 + \lceil \log_2(|A|) \rceil + \lceil \log_2(N) \rceil && \textit{(resettable) qubits.} \end{aligned}$$

See Lemma 2.3.1 for the explicit construction (which is essentially the same circuit leading to Lemma 2.1.1) and the required N to ensure a good discretization error (section 2.8). Therefore, the Gaussian width merely needs to scale as

$$\sigma_t \sim \beta t_{mix}/\epsilon$$

to prepare a Gibbs sample; see Table 2.1 for the total cost.

Quantum-walk speedup

With full algorithmic freedom, we may further depart from physics and seek a Szegedy-type speedup by considering a coherent representation of the Lindbladian [164, 180].[18] Inheriting the notation of (2.8), we consider the following *discriminant proxy*:

$$\mathcal{D}_\beta := \sum_{a\in A} \int \sqrt{\gamma(\omega)\gamma(-\omega)}\hat{\boldsymbol{A}}^a(\omega)\otimes\hat{\boldsymbol{A}}^{a*}(\omega) - \frac{\gamma(\omega)}{2}\left(\boldsymbol{A}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega)\otimes\boldsymbol{I} + \boldsymbol{I}\otimes\hat{\boldsymbol{A}}^a(\omega)^{\dagger *}\hat{\boldsymbol{A}}^a(\omega)^*\right)\mathrm{d}\omega. \tag{2.18}$$

The superscript $\boldsymbol{O}^*$ denotes the entry-wise complex conjugation. Indeed, as required by the quantum walk formalism, this operator is Hermitian (Corollary 2.3.2)

$$(\mathcal{D}_\beta)^\dagger = (\mathcal{D}_\beta) \quad \text{assuming symmetries} \quad (2.10), (2.11).$$

Analogously to the discriminant of classical Markov chains, the discriminant proxy is approximately the Davies-type Lindbladian (2.8) superoperator conjugated by the Gibbs state (section 2.3) but *vectorized* into an operator (section 2.3). This construction comes at the cost of duplicating the Hilbert space but, as a bonus, provides access to the following canonical purification of the Gibbs state

$$\left|\sqrt{\boldsymbol{\rho}_\beta}\right\rangle \propto \sum_i \mathrm{e}^{-\beta E_i/2}|\psi_i\rangle\otimes\left|\psi_i^*\right\rangle \quad \text{where} \quad \boldsymbol{H} = \sum_i E_i|\psi_i\rangle\langle\psi_i|$$

as the (approximate) top eigenvector of $\mathcal{D}_\beta$. The superscript $\left|\psi_i^*\right\rangle$ denotes entrywise complex conjugate in the computational basis[19]. If the Hamiltonian is diagonal in the computation basis, this is essentially equivalent to the purified classical distribution. For general Hamiltonians, the state $\left|\sqrt{\boldsymbol{\rho}_\beta}\right\rangle$ is also known as the thermofield double state in quantum gravity (see, e.g., [122]). Taking a partial trace recovers the Gibbs state, but the purification can sometimes be more useful, e.g., for efficient verification by a swap test or faster evaluation of observables (see, e.g., [102]).

Theorem 2.1.4 (Approximate purified Gibbs state)**.** *Instantiate the Lindbladian parameters of Theorem 2.1.3 with Gaussian width σ_t for the corresponding discriminant proxy $\mathcal{D}_\beta$ (2.18). Then, the leading eigenvector $\left|\lambda_1(\mathcal{D}_\beta)\right\rangle$ well-approximates the Gibbs state*

$$\left\|\left|\lambda_1(\mathcal{D}_\beta)\right\rangle - \left|\boldsymbol{\rho}_\beta\right\rangle\right\| = \widetilde{O}\left(\frac{\beta}{\sigma_t}\frac{1}{\lambda_{gap}(\mathcal{D}_\beta)}\right).$$

The quantity $\lambda_{gap}(\mathcal{D}_\beta) := \lambda_1(\mathcal{D}_\beta) - \lambda_2(\mathcal{D}_\beta)$ is the spectral gap.

[18] For convenience, we define the discriminant such that it is shifted by the identity matrix compared to definitions in earlier work.

[19] The above purified state is independent of which basis one applies complex conjugation to.

See section 2.3 for the proof of Theorem 2.1.4. The top-eigenvector error resembles the Lindbladian case (Theorem 2.1.3) with the mixing time $t_{mix}(\mathcal{L}_\beta)$ replaced by the inverse spectral gap $\lambda_{gap}^{-1}(\mathcal{D}_\beta)$. Though, unlike its Lindbladian cousin $\mathcal{L}_\beta$, the discriminant proxy $\mathcal{D}_\beta$ does not generate a semi-group and does not by itself prepare its gapped eigenvector; we need to additionally perform *quantum simulated annealing* (section 2.12), which is basically adiabatic state preparation supplemented with a natural adiabatic path from high to low temperature $0 < \beta' < \beta$. This additional step uses

$$\sim \text{(adiabatic path length)} \quad \beta\|\boldsymbol{H}\| \times \text{(worst root-inverse-gap)} \quad \sqrt{\lambda_{gap}^{-1}}$$

queries to block-encodings of $\boldsymbol{I} + \mathcal{D}_{\beta'}$ across values of β'. In other words, the quantum-walk speedup against Lindbladians boils down to replacing the mixing time by $t_{mix} \to \beta\|\boldsymbol{H}\|\sqrt{\lambda_{gap}^{-1}}$; this speedup comes with the cost of doubling the number of qubits $n \to 2n$. Of course, to perform adiabatic state preparation, we must algorithmically construct the block-encoding for $\boldsymbol{I} + \mathcal{D}_\beta$.

Lemma 2.1.3 (Efficient block-encoding)**.** *In the setting of Theorem 2.1.4, a unitary block-encoding of $\boldsymbol{I} + \mathcal{D}_\beta$ can be constructed up to ϵ spectral norm error using $O(1)$ query each to $\boldsymbol{Prep}$, $\boldsymbol{W}$, $\boldsymbol{V}_{jump}$, $\boldsymbol{QFT}$, $\boldsymbol{F}$, $\boldsymbol{P}$, and*

$\tilde{O}(\sigma_t)$	*(controlled) Ham. sim. time for* $\boldsymbol{H}$
$2n + \lceil \log_2(\lvert A\rvert) \rceil + \tilde{O}(1)$	*(resettable) qubits.*

See Proposition 2.3.5 for the circuit for Lemma 2.1.3. In the above, we have implicitly chosen the appropriate discretization N (section 2.8). Note that to obtain a quantum-walk speedup, we made use of two additional (low-cost) circuit components (see section 2.3): the reflection on energy

$$\boldsymbol{F} := \sum_{\bar{\omega} \in S_{\omega_0}} |-\bar{\omega}\rangle\langle\bar{\omega}|$$

and a permutation of the jump operator labels

$$\boldsymbol{P} := \sum_{a \in A} |a'\rangle\langle a| \quad \text{where} \quad \boldsymbol{A}^{a'} = (\boldsymbol{A}^a)^\dagger \quad \text{for each} \quad a \in A.$$

To grasp the algorithmic cost, we roughly expect the width along the adiabatic path to scale as

$$\sigma_t(\beta') \sim \frac{\beta'}{\epsilon' \lambda_{gap}(\mathcal{D}_{\beta'})} \quad \text{for} \quad \epsilon'\text{-approximated} \quad \left|\sqrt{\rho_{\beta'}}\right\rangle;$$

see Table 2.1 and section 2.12 for an quantitative accumulated cost for the adiabatic algorithm.

The remainder of the main text is organized by the analytic (section 2.2) and the algorithmic parts (section 2.3). We begin the analytic exposition by reviewing basic facts circling detailed balance and mixing time, and then introduce consequences of approximate detailed balance. The algorithmic arguments include black-box Lindbladian simulation and a general recipe to quantize a Lindbladian. These abstract algorithms can be understood assuming merely block-encodings for the Lindblad operator, whose explicit circuit construction is laid out in section 2.3. We conclude the main text by highlighting plausible future directions in section 2.4.

The appendices are organized as follows. We begin with the supporting details for our key analytic and algorithmic argument: section 2.5 discusses properties of the operator Fourier Transform and the secular approximation; section 2.7 proves approximate detailed balance for the constructed Lindbladians $\mathcal{L}_\beta$ and discriminant proxies $\mathcal{D}_\beta$.

The rest of the appendices consist of isolated topics. section 2.8 discusses the relation between continuous Fourier Transforms, which is conceptually simple, and the discrete Fourier Transform, which we implement. Fortunately, the rule of thumb is that the Fourier Transform register merely needs to be poly-logarithmic for a small discretization error. section 2.9 discusses Lindbladians arising from a microscopic open system derivation and prove their fixed point accuracy; this requires a moderate generalization of the main analytic framework. section 2.10 is devoted to supporting approximate detailed balance (section 2.2), especially on perturbation theory for nonHermitian matrices; these facts tend to be intuitively akin to the Hermitian case but we include the (nonstandard) proofs for completeness. section 2.12 reviews quantum simulated annealing in a modern quantum algorithm language, which we largely employ as a black box.

2.2 Approximate stationarity of the Gibbs state

We begin our analysis of the generator in Eqn. (2.8) by recalling some general properties of detailed balance Lindbladians before introducing the key notion of *approximate detailed balance*. At the heart of classical Markov chain Monte Carlo algorithms is a rapid mixing Markov chain whose fixed point yields the desired distribution. In the quantum setting, central to our discussion is the generator of a quantum dynamical semi-group [29, 182], the *Lindbladian* in the Schrödinger

Picture

$$\mathcal{L}[\boldsymbol{\rho}] = -\mathrm{i}[\boldsymbol{H}, \boldsymbol{\rho}] + \sum_{j\in J}\left(\boldsymbol{L}_j \boldsymbol{\rho} \boldsymbol{L}_j^\dagger - \frac{1}{2}\{\boldsymbol{L}_j^\dagger \boldsymbol{L}_j, \boldsymbol{\rho}\}\right)$$

parameterized by a set of *Lindblad operators* $\{\boldsymbol{L}_j\}_{j\in J}$ and a Hermitian matrix $\boldsymbol{H}$. Mathematically, the above elegant form encompasses all possible Lindbladians, including, but not restricted to, those arising from a microscopic system-bath derivation. In particular, from an algorithmic perspective, we enjoy the additional freedom of choosing favorable Lindblad operators $\boldsymbol{L}_j$ with the hope that

1. the Lindbladian can be implemented efficiently,
2. the fixed point is unique and yields the desired state $\boldsymbol{\rho}_\beta$, and
3. the Lindbladian converges rapidly.

The above summarizes the desirable criteria for a quantum Gibbs sampler.[20] The convergence depends on the particular Hamiltonian of interest and is generally nontrivial to analyze. Fortunately, the *detailed balance condition* enables systematic analysis of quantum dynamical semi-groups, similarly to how detailed balance is central in analyzing classical Markov chains (see, e.g., [110]).

Definition 2.2.1 (Detailed balance condition)**.** *For a normalized, full-rank state $\boldsymbol{\rho} \succ 0$, we say that an endomorphism $\mathcal{L}$ satisfies $\boldsymbol{\rho}$-detailed balance whenever the associated discriminant is self-adjoint with respect to $\boldsymbol{\rho}$, i.e.,*

$$\begin{aligned}\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}) &:= \boldsymbol{\rho}^{-1/4}\mathcal{L}[\boldsymbol{\rho}^{1/4} \cdot \boldsymbol{\rho}^{1/4}]\boldsymbol{\rho}^{-1/4}. \\ &= \boldsymbol{\rho}^{1/4}\mathcal{L}^\dagger[\boldsymbol{\rho}^{-1/4} \cdot \boldsymbol{\rho}^{-1/4}]\boldsymbol{\rho}^{1/4} = \mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger.\end{aligned}$$

In the above definition (and in the rest of the paper), we define the adjoint of a superoperator with respect to the Hilbert-Schmidt inner product $\langle \boldsymbol{X}, \boldsymbol{Y}\rangle_{HS} = \mathrm{Tr}(\boldsymbol{X}^\dagger \boldsymbol{Y})$. Explicitly,

$$C[\cdot] = \sum_j \alpha_j \boldsymbol{A}_j[\cdot]\boldsymbol{B}_j \quad \text{implies} \quad C^\dagger[\cdot] = \sum_j \alpha_j^* \boldsymbol{A}_j^\dagger[\cdot]\boldsymbol{B}_j^\dagger \quad \text{for any} \quad \boldsymbol{A}_j, \boldsymbol{B}_j$$

$$\text{since} \quad \mathrm{Tr}(\boldsymbol{X}^\dagger C[\boldsymbol{Y}]) = \mathrm{Tr}(\boldsymbol{X}^\dagger \sum_j \alpha_j \boldsymbol{A}_j \boldsymbol{Y} \boldsymbol{B}_j) = \mathrm{Tr}((\sum_j \alpha_j^* \boldsymbol{A}_j^\dagger \boldsymbol{X} \boldsymbol{B}_j^\dagger)^\dagger \boldsymbol{Y}) = \mathrm{Tr}[\left(C^\dagger(\boldsymbol{X})\right)^\dagger \boldsymbol{Y}]. \tag{2.19}$$

[20] One can certainly consider discrete-time quantum channels [167] as Gibbs sampler candidates. However, the continuous-time Lindbladian, inspired by physics, appears technically nicer for our purposes.

This will be revisited when defining the vectorization (section 2.3). In particular, the superoperator adjoint for Lindbladians coincides with converting between the Heisenberg and Schrödinger pictures

$$\mathrm{Tr}[\boldsymbol{\rho}\mathcal{L}^\dagger[\boldsymbol{O}]] = \mathrm{Tr}[\mathcal{L}[\boldsymbol{\rho}]\boldsymbol{O}] \quad \text{for each} \quad \boldsymbol{\rho} \geq 0 \quad \text{and} \quad \boldsymbol{O}$$

using that $\boldsymbol{\rho}^\dagger = \boldsymbol{\rho}$ and that Lindbladians $\mathcal{L}[\boldsymbol{\rho}] = (\mathcal{L}[\boldsymbol{\rho}])^\dagger$ preserves Hermiticity.

Analogously to the classical case, the detailed balance condition considers a similarity transformation according to the target distribution weights $\boldsymbol{\rho}$.[21] The detailed balance condition brings about two desirable properties. First, it ensures that the state $\boldsymbol{\rho}$ is a fixed point (Point 2).

Proposition 2.2.1 (Gibbs fixed point [166])**.** *If a superoperator $\mathcal{L}$ generates a trace-preserving map and satisfies $\boldsymbol{\rho}$-detailed balance, then it annihilates the state $\mathcal{L}[\boldsymbol{\rho}] = 0$, or equivalently, $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})[\sqrt{\boldsymbol{\rho}}] = 0$.*

Proof. We know that the infiniestimal exponential map $\mathrm{e}^{\varepsilon\mathcal{L}}[\cdot] = [\cdot] + \varepsilon\mathcal{L}[\cdot] + \mathcal{O}(\varepsilon^2)[\cdot]$ is trace-preserving, thus the leading order term must satisfy $\mathcal{L}^\dagger[\boldsymbol{I}] = 0$. Therefore,

$$0 = \boldsymbol{\rho}^{1/4}\mathcal{L}^\dagger[\boldsymbol{I}]\boldsymbol{\rho}^{1/4} = \mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger[\sqrt{\boldsymbol{\rho}}] = \mathcal{D}(\boldsymbol{\rho}, \mathcal{L})[\sqrt{\boldsymbol{\rho}}]$$

using $\boldsymbol{\rho}$-detailed balance in the last equality. ∎

Second, it relates the Lindbladian mixing time to the spectral gap (Point 3). We only state the following result here but later prove a qualitatively more robust statement in Proposition 2.2.3 applicable to the approximate case.

Proposition 2.2.2 (Mixing time from spectral gap [97])**.** *If a Lindbladian $\mathcal{L}$ satisfies $\boldsymbol{\rho}$-detailed balance, then*

$$t_{mix}(\mathcal{L}) \leq \frac{\ln(2\|\boldsymbol{\rho}^{-1/2}\|)}{\lambda_{gap}(\mathcal{L})},$$

where $\lambda_{gap}(\mathcal{L})$ is the eigenvalue gap of the Lindbladian, and the mixing time t_{mix} is the smallest time for which

$$\left\|\mathrm{e}^{\mathcal{L}t_{mix}}[\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2]\right\|_1 \leq \frac{1}{2}\|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\|_1 \quad \textit{for any states} \quad \boldsymbol{\rho}_1, \boldsymbol{\rho}_2.$$

[21] Technically, an alternative definition of detailed balance may distribute the power somewhat arbitrarily $\boldsymbol{\rho}^s[\cdot]\boldsymbol{\rho}^{1/2-s}$, but we stick to the symmetric case $\boldsymbol{\rho}^{1/4}[\cdot]\boldsymbol{\rho}^{1/4}$ for simplicity.

The analysis of a superoperator gap $\lambda_{gap}(\mathcal{L})$ is perhaps more tractable than the mixing time t_{mix} but still nontrivial and instance specific.[22] Otherwise, we see that the detailed balance condition readily addresses two criteria (Point 2 and Point 3) for Gibbs samplers.

Approximate detailed balance

Unfortunately, we do not know of general *efficient* constructions of quantum Gibbs samplers satisfying the detailed balance condition *exactly* (Point 1)[23]; this is rooted in the energy-time uncertainty principle where quantum algorithms only access the energies of a quantum system approximately. As our main technical contribution, we formulate the *ϵ-approximate detailed balance condition* that addresses all three requirements for a quantum Gibbs sampler.

Definition 2.2.2 (Approximate detailed balance condition)**.** *For any Lindbladian $\mathcal{L}$ and full-rank state $\boldsymbol{\rho}$, take a similarity transformation and decompose into the Hermitian and the anti-Hermitian parts*

$$\mathcal{D}(\boldsymbol{\rho},\mathcal{L}) = \boldsymbol{\rho}^{-1/4}\mathcal{L}[\boldsymbol{\rho}^{1/4}\cdot\boldsymbol{\rho}^{1/4}]\boldsymbol{\rho}^{-1/4} = \mathcal{H}(\boldsymbol{\rho},\mathcal{L}) + \mathcal{A}(\boldsymbol{\rho},\mathcal{L}).$$
$$\mathcal{D}(\boldsymbol{\rho},\mathcal{L})^\dagger = \boldsymbol{\rho}^{1/4}\mathcal{L}^\dagger[\boldsymbol{\rho}^{-1/4}\cdot\boldsymbol{\rho}^{-1/4}]\boldsymbol{\rho}^{1/4} = \mathcal{H}(\boldsymbol{\rho},\mathcal{L}) - \mathcal{A}(\boldsymbol{\rho},\mathcal{L})$$

We say the Lindbladian $\mathcal{L}$ satisfies the ϵ-approximate $\boldsymbol{\rho}$-detailed balance condition if the anti-Hermitian part $\mathcal{A}$ is small

$$\frac{1}{2}\left\|\mathcal{D}(\boldsymbol{\rho},\mathcal{L}) - \mathcal{D}(\boldsymbol{\rho},\mathcal{L})^\dagger\right\|_{2-2} = \|\mathcal{A}(\boldsymbol{\rho},\mathcal{L})\|_{2-2} \le \epsilon.$$

If the anti-Hermitian part vanishes, we recover the exact detailed balance condition $\mathcal{D}(\boldsymbol{\rho},\mathcal{L})^\dagger = \mathcal{D}(\boldsymbol{\rho},\mathcal{L})$. If not, we show that the fixed point still approximates the state $\boldsymbol{\rho}$ (Point 2).

Corollary 2.2.1 (Fixed point accuracy)**.** *If a Lindbladian $\mathcal{L}$ satisfies the ϵ-approximate $\boldsymbol{\rho}$-detailed balance condition, then its fixed point $\boldsymbol{\rho}_{fix}(\mathcal{L})$ deviates from $\boldsymbol{\rho}$ by at most*

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\|_1 \le 20 t_{mix}(\mathcal{L})\epsilon.$$

See section 2.10 for the proof. We see that the error bound deteriorates if the map has a large anti-Hermitian component $\mathcal{A}$ or if the Lindbladian mixes slowly.[24] The

[22] For the experts, the gap may not give the tightest possible mixing time bounds; techniques beyond gap-based bounds typically require proving a Log-Sobolev inequality, which can be very challenging in the noncommuting cases.

[23] This problem is resolved in a follow-up work [42].

[24] We actually prove a stronger statement in section 2.10 that gives a bound in terms of the gap $\lambda_{gap}(\mathcal{H})$ of the Hermitian part.

anti-Hermitian component involves the inverse Gibbs state $\boldsymbol{\rho}^{-1}$, and might be difficult to bound directly. As a remedy, it is helpful to introduce an intermediate Lindbladian $\mathcal{L}'$ for which approximate detailed balance is easier to show. In that case, we can write

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\|_1 \leq \|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}_{fix}(\mathcal{L}')\|_1 + \|\boldsymbol{\rho}_{fix}(\mathcal{L}') - \boldsymbol{\rho}\|_1.$$

The first term on the RHS does not directly involve the inverse $\boldsymbol{\rho}^{-1}$ and can be controlled by a Lindbladian perturbation bound as follows.

Lemma 2.2.1 (Fixed point difference)**.** *For any two Lindbladians $\mathcal{L}_1$ and $\mathcal{L}_2$, the difference of their fixed points (in the Schrödinger picture) is bounded by*

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}_1) - \boldsymbol{\rho}_{fix}(\mathcal{L}_2)\|_1 \leq 4\|\mathcal{L}_1 - \mathcal{L}_2\|_{1-1} \cdot t_{mix}(\mathcal{L}_1).$$

See section 2.10 for the proof. Conveniently, even without detailed balance, the mixing time t_{mix} remains controlled by spectral properties of the Hermitian part (addressing Point 3):

Proposition 2.2.3 (Mixing time from Hermitian gap)**.** *For any Lindbladian $\mathcal{L}$ and a full-rank state $\boldsymbol{\rho}$, suppose the self-adjoint component $\mathcal{H} = \mathcal{H}(\boldsymbol{\rho}, \mathcal{L})$ satisfies*

$$\frac{\lambda_1(\mathcal{H})}{\lambda_{gap}(\mathcal{H})} \leq \frac{1}{100}, \quad \textit{then} \quad t_{mix}(\mathcal{L}) \leq 3\frac{\ln(3\|\boldsymbol{\rho}^{-1/2}\|)}{\lambda_{gap}(\mathcal{H})}.$$

See section 2.10 for the proof. In particular, the top eigenvalue can be bounded by the anti-Hermitian part $\lambda_1(\mathcal{H}) \leq \|\mathcal{A}\|_{2-2}$ for any Lindbladian (2.76); therefore, it remains to provide an efficient construction of the Lindbladian (section 2.3) and prove approximate detailed balance (section 2.7).

Proof of fixed point correctness (Theorem 2.1.3)

We are now in a position to prove our first main theorem: the proximity of the stationary state $\boldsymbol{\rho}_{fix}$ to the Gibbs state $\boldsymbol{\rho}_\beta$. Most of the technical definitions and lemmata are relegated to Appendix 2.6. Here, we address the essential features of the proof together with some essential tools. The main technical argument introduces an intermediate Lindbladian $\mathcal{L}_{sec}$

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}_\beta) - \boldsymbol{\rho}_\beta\|_1 \leq \|\boldsymbol{\rho}_{fix}(\mathcal{L}_\beta) - \boldsymbol{\rho}_{fix}(\mathcal{L}_{sec})\|_1 + \|\boldsymbol{\rho}_{fix}(\mathcal{L}_{sec}) - \boldsymbol{\rho}_\beta\|_1$$

and uses the fixed point error bounds (Lemma 2.2.1) for the first term and (Proposition 2.10.3) for the second term.

The first error arises from the *secular approximation* (section 2.6), defined by truncating the transitions in the frequency domain

$$\hat{\boldsymbol{A}}^a(\bar{\omega}) \to \hat{\boldsymbol{S}}^a(\bar{\omega}) \quad \text{such that} \quad \langle\psi_i|\hat{\boldsymbol{S}}^a(\bar{\omega})|\psi_j\rangle = 0 \quad \text{whenever} \quad \left|(E_i - E_j) - \bar{\omega}\right| > \bar{\mu}$$

for a tunable truncation parameter $\bar{\mu}$. See (2.41), (2.45) for the precise definition of the secular approximated jump operators $\hat{\boldsymbol{S}}^a(\bar{\omega})$. The purpose of this truncation is to ensure approximate detailed balance: conjugating $\mathcal{L}_{sec}$ with the Gibbs state $\boldsymbol{\rho}$, as required in comparing with the similarity transformation, remains well-behaved (section 2.6). The truncation parameter $\bar{\mu}$ is not physical but rather a proof artifact. Intuitively, our choice of Gaussian weight ensures its Fourier Transform to remain (approximately) another Gaussian (see section 2.6), which has a rapidly decaying tail. Thus, we expect the error from truncating the Gaussian tail to be small whenever $\bar{\mu} \gtrsim \sigma_t^{-1}$. Thus, with the Gaussian weight, the secular approximation incurs a mild error; this error becomes more severe with the step-function weights given by nature, whose Fourier Transform has a heavy tail (Proposition 2.6.6).

The second error is the most technical part, showing that the secular-approximated operator $\mathcal{L}_{sec}$ satisfies approximate detailed balance (See section 2.7). We highlight the full technical statement as follows.

Lemma 2.2.2 (Approximate detailed balance)**.** *Consider a Lindbladian in the following form*

$$\mathcal{L} = \sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \gamma(\bar{\omega})\hat{\boldsymbol{S}}^a(\bar{\omega})[\cdot]\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger - \frac{\gamma(\bar{\omega})}{2}\{\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger\hat{\boldsymbol{S}}^a(\bar{\omega}), \cdot\},$$

where $\gamma(\bar{\omega})/\gamma(-\bar{\omega}) = \mathrm{e}^{-\beta\bar{\omega}}$ *for each* $\bar{\omega} \in S_{\omega_0}$*. Suppose there exists* $\bar{\mu} \le \beta^{-1}$ *such that the operators satisfy*

$$\langle\psi_i|\hat{\boldsymbol{S}}^a(\bar{\omega})|\psi_j\rangle = 0 \quad \textit{whenever} \quad \left|(E_i - E_j) - \bar{\omega}\right| > \bar{\mu}$$

for the eigenvalue decomposition of $\boldsymbol{H} = \sum_j E_j|\psi_j\rangle\langle\psi_j|$*, and there is a permutation* $\boldsymbol{P}\colon a \to a'$ *such that* $\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger = \hat{\boldsymbol{S}}^{a'}(-\bar{\omega})$ *for each* $a, \bar{\omega}$*. Then, for the Gibbs state* $\boldsymbol{\rho} = \mathrm{e}^{-\beta\boldsymbol{H}}/\mathrm{Tr}[\mathrm{e}^{-\beta\boldsymbol{H}}]$ *we have*

$$\frac{1}{2}\left\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger\right\|_{2-2} \le O\left(\beta\bar{\mu}\left\|\sum_{a\in A}\sum_{\bar{\omega}\in S_{\omega_0}}\gamma(\bar{\omega})\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger\hat{\boldsymbol{S}}^a(\bar{\omega})\right\|\right).$$

The above is a simplified version of Lemma 2.7.3, which we prove in section 2.7. Our normalization further simplifies the RHS to $O(\beta\bar{\mu})$. We now combine the above estimates to prove Theorem 2.1.3.

Proof of Theorem 2.1.3. While Theorem 2.1.3 is stated in the continuum limit $N \to \infty$ (2.8), we give general error bounds at finite N (2.21) and then take the $N \to \infty$ limit (2.8). Introduce the secular-approximated Lindblad operator $\mathcal{L}_{sec}$ to bound the fixed point error

$$\begin{aligned}
&\|\boldsymbol{\rho}_{fix}(\mathcal{L}_\beta) - \boldsymbol{\rho}_\beta\|_1 \\
&\le \|\boldsymbol{\rho}_{fix}(\mathcal{L}_\beta) - \boldsymbol{\rho}_{fix}(\mathcal{L}_{sec})\|_1 + \|\boldsymbol{\rho}_{fix}(\mathcal{L}_{sec}) - \boldsymbol{\rho}_\beta\|_1 \\
&\le 2\|\mathcal{L}_\beta - \mathcal{L}_{sec}\|_{1-1} t_{mix}(\mathcal{L}_\beta) + 10\left\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\right\| t_{mix}(\mathcal{L}_{sec}) \\
&\le O\left(\left(\|\mathcal{L}_\beta - \mathcal{L}_{sec}\|_{1-1} + \left\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\right\|\right) t_{mix}(\mathcal{L}_\beta)\right). \qquad (2.20)
\end{aligned}$$

The second inequality uses Lemma 2.2.1 for the first term and Corollary 2.2.1 for the last term. The third inequality is that $t_{mix}(\mathcal{L}_{sec}) \le t_{mix}(\mathcal{L}_\beta)\left\lceil \frac{\ln(1/2)}{\ln(1/2 + t_{mix}(\mathcal{L}_\beta)\|\mathcal{L}_\beta - \mathcal{L}_{sec}\|_{1-1})} \right\rceil$ (Proposition 2.10.4), which further simplifies to $t_{mix}(\mathcal{L}_{sec}) = O(t_{mix}(\mathcal{L}_\beta))$ since we must have $\|\mathcal{L}_\beta - \mathcal{L}_{sec}\|_{1-1} t_{mix}(\mathcal{L}_\beta) = O(1)$ otherwise the trace distance bound becomes vacuous.

Now, we evaluate approximate detailed balance (Lemma 2.2.2) and the secular approximation error using Lemma 2.6.2, Proposition 2.6.7, and that the Gaussian tail in the time domain is bounded directly by $\sqrt{\sum_{|\bar{t}|\ge T} |f(\bar{t})|^2} = O(\sqrt{T/\sigma_t}^{-1} \mathrm{e}^{-T^2/4\sigma_t^2})$

$$(2.20) \le O\left(\left(T\bar{\omega}_0 + \mathrm{e}^{-T^2/4\sigma_t^2} + \mathrm{e}^{-N^2 t_0^2/16\sigma_t^2} + \mathrm{e}^{-N^2\omega_0^2\sigma_t^2/2} + \mathrm{e}^{-\bar{\mu}^2\sigma_t^2} + \beta\bar{\mu}\right) \cdot t_{mix}(\mathcal{L}_\beta)\right)$$

$$\le O\left((\sigma_t\omega_0\sqrt{\log(1/(\sigma_t\omega_0))} + \frac{\beta}{\sigma_t}\sqrt{\log(\sigma_t/\beta)} + \mathrm{e}^{-N^2\omega_0^2\sigma_t^2/2}) \cdot t_{mix}(\mathcal{L}_\beta)\right).$$

The second inequality chooses the free parameter $T = 2\sigma_t\sqrt{\ln(1/(\sigma_t\omega_0))}$ and $\bar{\mu} = \frac{\beta}{\sigma_t}\sqrt{\ln(\sigma_t/\beta)}$ and uses that $\mathrm{e}^{-N^2 t_0^2/16\sigma_t^2} = \mathrm{e}^{-\pi^2/4\omega_0^2\sigma_t^2} = O(\sigma_t\omega_0\sqrt{\log(1/(\sigma_t\omega_0))})$ to simplify the expression. For the continuum case (2.21), we have the simpler bound

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}_\beta) - \boldsymbol{\rho}_\beta\|_1 = O\left(\frac{\beta}{\sigma_t}\sqrt{\log(\sigma_t/\beta)} \cdot t_{mix}(\mathcal{L}_\beta)\right) \quad \text{if} \quad N\omega_0 \to \infty,\ \omega_0 \to 0,$$

where discretization parameter ω_0 and N disappears in the continuum limit. ∎

2.3 Quantum algorithms for Gibbs sampling

In this section, we present two algorithms for approximately preparing the Gibbs state $\boldsymbol{\rho}_\beta$, both of which are inspired by the dynamical semi-group generated by the Lindbladian $\mathcal{L}_\beta$. This first algorithm, which we call the *incoherent Gibbs sampling algorithm*, directly simulates the time evolution $e^{\mathcal{L}_\beta t}$ by introducing ancillas. The second, which we call the *coherent Gibbs sampling algorithm*, is a Szegedy-type quantum walk algorithm. It enables implementing an orthogonal projector onto the coherent Gibbs state $|\sqrt{\boldsymbol{\rho}_\beta}\rangle\langle\sqrt{\boldsymbol{\rho}_\beta}|$ with a quadratic speedup with respect to the real spectral gap of the generator $\mathcal{L}_\beta$. This projector can then be used in conjunction with simulated annealing (section 2.12) to prepare the purified Gibbs state.

In the circuit constructions, we will extensively use the following rotation gates

$$\boldsymbol{Y}_\theta := e^{-i \arcsin\sqrt{\theta}\boldsymbol{Y}} = \begin{pmatrix} \sqrt{1-\theta} & -\sqrt{\theta} \\ \sqrt{\theta} & \sqrt{1-\theta} \end{pmatrix} \quad \text{with the Pauli-Y matrix} \quad \boldsymbol{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Our quantum Gibbs sampling algorithms

We describe two Lindbladian simulation algorithms: the first exhibits Trotter-like scaling and repeatedly uses a simple (randomized) and weak-measurement gadget (Theorem 2.3.1, Corollary 2.3.1); the second is inspired by [49] and has asymptotically almost optimal scaling with time and error Theorem 2.3.2 but requiring a more involved circuit and slightly more ancilla qubits. Both arguments are general as they assume merely a block-encoding of the Lindbladian (Definition 2.1.2); the particular block-encoding for our proposed Gibbs sampler is constructed explicitly in another section (section 2.3).

Further, we "quantize" the Lindbladians and present *coherent* Gibbs sampling algorithms that prepare the (canonical) purification of an approximate Gibbs state via simulated annealing (section 2.12). The procedure assumes that we have a block-encoding of the discriminant matrix of our Lindbladian, which then enables a Szegedy-type quadratic speedup in the simulation time. However, the total speedup is only sub-quadratic on the gap dependence because of the cost to block-encode the discriminant matrix.

Incoherent Lindbladian simulation algorithms

Guided by [49], we propose two different implementation methods for incoherent (trajectory-based) simulation of the Lindbladians that describe our Gibbs sampler. The first method is based on a product formula and repeatedly uses a weak mea-

surement scheme[25] for implementing a small time step. The resulting scaling is analogous to the performance of "vanilla" Trotter-based Hamiltonian simulation: the complexity for an ϵ-accurate-time-t Lindbladian evolution scales as t^2/ϵ. Our weak measurement scheme gives rise to simple and low-depth circuits for simulating Lindbladians given block-encoding access.

The usefulness of weak measurements should come as no surprise, as they are also extremely helpful in other noncommutative state preparation tasks as well (see, e.g., [72]), and the very recent independent work of [53]. The common theme in these applications is the exploitation of some quantum Zeno-like effect,[26] but on a higher level, these applications also show some conceptual differences. We leave it for future work to explore whether there is a more fundamental connection between our weak measurement scheme and that of [53, 72].

The second method is based on the algorithm of [49], which achieves a close-to-optimal scaling with respect to time and accuracy. Although the asymptotical complexity is much improved, the corresponding circuits are more complicated as they use a linear combination of unitaries (LCU), oblivious amplitude amplification, and advanced "compression" techniques. We leave it to future work to determine how the two schemes perform in practice.

For both algorithms, it suffices to assume that a *purely irreversible* Lindbladian without the Hamiltonian term $\mathcal{L}[\cdot] = \sum_{j\in J} \boldsymbol{L}_j[\cdot]\boldsymbol{L}_j^\dagger - \frac{1}{2}\{\boldsymbol{L}_j^\dagger \boldsymbol{L}_j, \cdot\}$ is provided in the form of a "block-encoding" (i.e., dilation) as Definition 2.1.2.[27] In particular, recall

[25] Our weak measurement scheme is very similar to the short-time evolution by the auxiliary Hamiltonian J utilized in [49]; however, our approach is a bit more direct and made it clear that a block-encoding of the jump operators suffices as input.

[26] In our case, the quantum Zeno-like effect is manifest in the quadratically reduced amplitude of $|0^c\ \perp\rangle$ in (2.22).

[27] Recent work [112, 113] assumes the Lindbladian jumps are individually block-encoded while we assume the *entire* set of jumps is encoded in a *single* unitary. We give strictly better complexity for simulating Lindbladians under this input model, which holds for our Gibbs sampling algorithm and that of [149] (leading to direct improvement for the latter). Remarkably, even if the jumps are individually block-encoded [112, 113], these can be converted to our input model. Still, even accounting for the conversion overhead, we recover (up to polylogarithmic factors) their complexity for Lindbladian simulation. The main innovation here seems to be the generalization of the input model, as the earlier Lindbladian simulation algorithms also seem to work [176] under this more general input assumption.

our proposed Lindbladian Gibbs sampler (as discretization of (2.8))

$$\mathcal{L}_\beta := \sum_{a\in A,\bar{\omega}\in S_{\omega_0}} \gamma(\bar{\omega})\left(\hat{\boldsymbol{A}}^a(\bar{\omega})[\cdot]\hat{\boldsymbol{A}}^a(\bar{\omega})^\dagger - \frac{1}{2}\{\hat{\boldsymbol{A}}^a(\bar{\omega})^\dagger\hat{\boldsymbol{A}}^a(\bar{\omega}),\cdot\}\right) \quad (2.21)$$

$$\text{with Lindblad operators} \quad \{\sqrt{\gamma(\bar{\omega})}\hat{\boldsymbol{A}}^a(\bar{\omega})\}_{a\in A,\bar{\omega}\in\mathbb{R}},$$

and its block-encoding can be found in Eqn. 2.41 in section 2.3. However, working with abstract block encodings makes our simulation results general and also simplifies our presentation and proofs, as the operator Fourier Transform naturally fits this definition (Figure 2.4). Our weak-measurement scheme is not only simple but also improves, e.g., the sparse Lindbladian simulation algorithm of [43, Theorem 9].[28] Also, the lower bound on the "total evolution time" for simple iterative circuits in [49] suggests that the performance of similar schemes may be optimal.

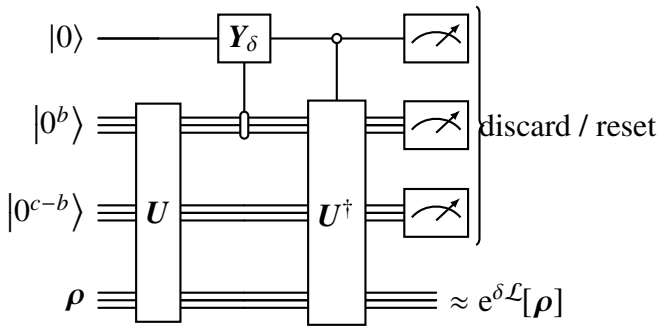


Figure 2.3: Quantum circuit implementation of an approximate δ-time step via a weak measurement scheme. The scheme can be extended to general Lindbladians that include the coherence term $-i[\boldsymbol{H},\boldsymbol{\rho}]$ by applying $O(\delta^2)$-precise Hamiltonian time-evolution for time δ on the system register before the above circuit is applied. For example, one could use Trotterized time-evolution. (In case $\|\boldsymbol{H}\| > 1$, the entire Lindbladian should be first scaled down by a factor of $\|\boldsymbol{H}\|$.)

Theorem 2.3.1 (Weak-measurement for incoherent Lindbladian simulation)**.** *Suppose $\boldsymbol{U}$ is a block-encoding of the purely irreversible Lindbladian $\mathcal{L}$ as in Definition 2.1.2. We can simulate the action of the superoperator $e^{t\mathcal{L}}$ to precision ϵ in diamond norm using*

$$\begin{aligned} & c+1 && \textit{(resettable) ancilla qubits},\\ & O\left(t^2/\epsilon\right) && \textit{(controlled) uses of} \quad \boldsymbol{U},\boldsymbol{U}^\dagger,\\ \textit{and} \quad & O\left((b+1)t^2/\epsilon\right) && \textit{other two-qubit gates}. \end{aligned}$$

[28] Indeed, the complexity is improved by about a factor of k^4, where k is the sparsity.

Proof. We can simulate an approximate δ-time step by $\mathcal{L}$ using the following weak-measurement scheme displayed in Figure 2.3.

1. Apply $\boldsymbol{U}$.
2. Append an ancilla qubit in state $|0\rangle$ and rotate it with angle $\arcsin\sqrt{\delta}$ controlled on the $\left|0^b\right\rangle$ state (indicating the successful application of a jump).
3. Apply $\boldsymbol{U}^\dagger$ controlled on the ancilla qubit being 0.
4. Measure and discard all but the system register.

Assuming the system register is in the pure state $|\psi\rangle$, this circuit $\boldsymbol{C}$ acts as follows:

$$
\begin{aligned}
&|0\rangle\cdot|0^c\rangle|\psi\rangle \\
&\overset{(1)}{\to} |0\rangle\cdot\boldsymbol{U}|0^c\rangle|\psi\rangle \\
&\overset{(2)}{\to} \left(\sqrt{1-\delta}|0\rangle+\sqrt{\delta}|1\rangle\right)\cdot\left(|0^b\rangle\langle 0^b|\otimes\boldsymbol{I}\right)\boldsymbol{U}|0^c\rangle|\psi\rangle \quad + \quad |0\rangle\cdot(\boldsymbol{I}-|0^b\rangle\langle 0^b|\otimes\boldsymbol{I})\boldsymbol{U}|0^c\rangle|\psi\rangle \\
&= |0\rangle\cdot\boldsymbol{U}|0^c\rangle|\psi\rangle \quad + \quad \sqrt{\delta}|1\rangle\cdot\left|0^b\right\rangle\underbrace{(\left\langle 0^b\right|\otimes\boldsymbol{I})\boldsymbol{U}|0^c\rangle|\psi\rangle}_{|\psi_0'\rangle:=} \\
&- \quad (1-\sqrt{1-\delta})|0\rangle\cdot\left(|0^b\rangle\langle 0^b|\otimes\boldsymbol{I}\right)\boldsymbol{U}|0^c\rangle|\psi\rangle \\
&\overset{(3)}{\to} |0\rangle\cdot|0^c\rangle|\psi\rangle \quad + \quad \sqrt{\delta}|1\rangle\cdot\left|0^b\right\rangle|\psi_0'\rangle \quad - \quad (1-\sqrt{1-\delta})|0\rangle\cdot\boldsymbol{U}^\dagger\left(|0^b\rangle\langle 0^b|\otimes\boldsymbol{I}\right)\boldsymbol{U}|0^c\rangle|\psi\rangle \\
&= |0\rangle\cdot|0^c\rangle|\psi\rangle \quad + \quad \sqrt{\delta}|1\rangle\cdot\left|0^b\right\rangle|\psi_0'\rangle \\
&- \quad (1-\sqrt{1-\delta})|0\rangle\cdot|0^c\rangle(\langle 0^c|\otimes\boldsymbol{I})\boldsymbol{U}^\dagger(\left|0^b\right\rangle\otimes\boldsymbol{I})\cdot(\left\langle 0^b\right|\otimes\boldsymbol{I})\boldsymbol{U}|0^c\rangle|\psi\rangle \\
&- (1-\sqrt{1-\delta})|0\rangle\cdot(\boldsymbol{I}-|0^c\rangle\langle 0^c|\otimes\boldsymbol{I})\boldsymbol{U}^\dagger\left(|0^b\rangle\langle 0^b|\otimes\boldsymbol{I}\right)\boldsymbol{U}|0^c\rangle|\psi\rangle \\
&= |0\rangle\cdot|0^c\rangle\left(\boldsymbol{I}-\underbrace{(1-\sqrt{1-\delta})}_{\frac{\delta}{2}+O(\delta^2)}\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\right)|\psi\rangle \\
&+ \sqrt{\delta}|1\rangle\cdot\left|0^b\right\rangle\sum_{j\in J}|j\rangle\boldsymbol{L}_j|\psi\rangle - \underbrace{(1-\sqrt{1-\delta})}_{\frac{\delta}{2}+O(\delta^2)}|0\rangle\cdot|0^c\ \perp\rangle, \qquad (2.22)
\end{aligned}
$$

where $|0^c\ \perp\rangle$ is some quantum state such that $\||0^c\ \perp\rangle\|\le 1$ and $(\langle 0^c|\otimes\boldsymbol{I})\cdot|0^c\ \perp\rangle=0$. Tracing out the first $a+1$ qubits, we get that the resulting state is $O\left(\delta^2\right)$-close to the desired state. Indeed, let $|\psi'\rangle$ denote the final state above in (2.22); we now show that

$$
\|(\mathcal{I}+\delta\mathcal{L})[|\psi\rangle\langle\psi|]-\mathrm{Tr}_{c+1}[|\psi'\rangle\langle\psi'|]\|_1=O\left(\delta^2\right) \qquad (2.23)
$$

by observing that

$$
\begin{aligned}
&\mathrm{Tr}_{c+1}\left[|\psi'\rangle\langle\psi'|\right] \\
&= \mathrm{Tr}_c\Big[(\langle 0|\otimes \boldsymbol{I})\cdot|\psi'\rangle\langle\psi'|\cdot(|0\rangle\otimes\boldsymbol{I})\Big] + \mathrm{Tr}_c\Big[(\langle 1|\otimes \boldsymbol{I})\cdot|\psi'\rangle\langle\psi'|\cdot(|1\rangle\otimes\boldsymbol{I})\Big] \\
&= (\langle 0^{c+1}|\otimes \boldsymbol{I})\cdot|\psi'\rangle\langle\psi'|\cdot(|0^{c+1}\rangle\otimes\boldsymbol{I}) \\
&+ \mathrm{Tr}_c\Big[(\langle 0|\otimes \boldsymbol{I} - |0^c\rangle\langle 0^c|)\cdot|\psi'\rangle\langle\psi'|\cdot(|0\rangle\otimes\boldsymbol{I} - |0^c\rangle\langle 0^c|)\Big] \\
&+ \delta\sum_{j\in J}\boldsymbol{L}_j|\psi\rangle\langle\psi|\boldsymbol{L}_j^\dagger \\
&= \left(\boldsymbol{I} - \frac{\delta}{2}\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j + O\left(\delta^2\right)\right)|\psi\rangle\langle\psi|\left(\boldsymbol{I} - \frac{\delta}{2}\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j + O\left(\delta^2\right)\right) \\
&\quad + O\left(\delta^2\right)\mathrm{Tr}_c|0^c\ \perp\rangle\langle 0^c\ \perp| + \delta\sum_{j\in J}\boldsymbol{L}_j|\psi\rangle\langle\psi|\boldsymbol{L}_j^\dagger \\
&= |\psi\rangle\langle\psi| + \delta\sum_{j\in J}\boldsymbol{L}_j|\psi\rangle\langle\psi|\boldsymbol{L}_j^\dagger - \frac{\delta}{2}\Big\{\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j, |\psi\rangle\langle\psi|\Big\} + O\left(\delta^2\right) \\
&= (\mathcal{I} + \delta\mathcal{L})[|\psi\rangle\langle\psi|] + O\left(\delta^2\right).
\end{aligned}
$$

Convexity implies (2.23) also holds for mixed input states. To extend to the diamond norm, observe that $\mathcal{L}[\cdot]\otimes\boldsymbol{I}[\cdot]\boldsymbol{I}$ has Lindblad operators $\boldsymbol{L}_j\otimes\boldsymbol{I}$ and therefore $\boldsymbol{U}\otimes\boldsymbol{I}$ is a block-encoding for $\mathcal{L}[\cdot]\otimes\boldsymbol{I}[\cdot]\boldsymbol{I}$. This implies that the trace-norm bound of (2.23) holds with respect to $\mathcal{L}[\cdot]\otimes\boldsymbol{I}[\cdot]\boldsymbol{I}$ as well, and so we can conclude that

$$
\left\|(\mathcal{I} + \delta\mathcal{L})[\cdot] - \mathrm{Tr}_{c+1}\boldsymbol{C}\left[|0^{c+1}\rangle\langle 0^{c+1}|\otimes\cdot\right]\boldsymbol{C}^\dagger\right\|_\diamond = O\left(\delta^2\right). \tag{2.24}
$$

The triangle inequality then implies that the implemented map is $O\left(\delta^2\right)$-close in diamond distance to $\mathrm{e}^{\delta\mathcal{L}}$, since $\|(\mathcal{I} + \delta\mathcal{L}) - \mathrm{e}^{\delta\mathcal{L}}\|_\diamond = O\left(\delta^2\right)$ as shown by, e.g., [49, Appendix B].[29]

Choosing $\delta = \Theta(\frac{\epsilon}{t})$ ensures that the error in a single time-step is bounded by $O\left(\frac{\epsilon^2}{t^2}\right)$, and repeating the process $\Theta(\frac{t^2}{\epsilon})$-times induces an error that is bounded by ϵ for the entire time-t evolution. The complexity is then $\Theta(\frac{t^2}{\epsilon})$-times the cost of implementing the circuit in Figure 2.3. ■

In addition to purely irreversible Lindbladian, as noted below (Figure 2.3), the above weak measurement scheme can be amended with the Hamiltonian evolution term.

[29] Here we implicitly used the fact that a block-encoded Lindbladian has norm at most one. This follows from the observation that $\|\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\| \leq 1$, which is a direct consequence of Proposition 2.3.1.

In our Gibbs sampling algorithm, the original random process typically proceeds by a random "jump" operator $\boldsymbol{A}^a$ for a uniformly random $a \in A$. We then obtain the final generators by applying the operator Fourier Transform to these "jump" operators. Naïvely applying our weak measurement scheme to such a Lindbladian would require us to use all the "jump" operators in each iteration. However, we show in the next corollary that it suffices to randomly pick a single "jump" operator in each iteration. In some situations, we could hope for further improvement by parallelization of these jumps if their operator Fourier Transform remains localized.

Corollary 2.3.1 (Improved randomized simulation for convex combinations of Lindbladians)**.** *Suppose that a purely irreversible Lindbladian $\mathcal{L}[\cdot] = \sum_i p_i \mathcal{L}_i[\cdot]$ is a convex combination of the purely irreversible Lindbladians $\mathcal{L}_i[\cdot]$ which are given by their respective block-encodings. In Theorem 2.3.1, we can replace each weak-measurement gadget for $\mathcal{L}[\cdot]$ by an independently sampled weak-measurement gadget for $\mathcal{L}_i[\cdot]$ according to the distribution p_i while keeping the same asymptotic iteration count $O\left(t^2/\epsilon\right)$.*

Proof. It suffices to show that

$$\|\mathrm{e}^{\delta\mathcal{L}}[\cdot] - \sum_i p_i \mathrm{Tr}_{c+1} \boldsymbol{C}_i \Big(|0^{c+1}\rangle\langle 0^{c+1}| \otimes [\cdot]\Big) \boldsymbol{C}_i^\dagger\|_\diamond = O\left(\delta^2\right). \tag{2.25}$$

From (2.24) we know that the weak measurement gadget $\boldsymbol{C}_i$ in Figure 2.3 for $\mathcal{L}_i^\dagger[\cdot]$ satisfies

$$\|(\mathcal{I} + \delta\mathcal{L}_i)[\cdot] - \mathrm{Tr}_{c+1} \boldsymbol{C}_i \Big(|0^{c+1}\rangle\langle 0^{c+1}| \otimes [\cdot]\Big) \boldsymbol{C}_i^\dagger\|_\diamond = O\left(\delta^2\right).$$

By linearity and the triangle inequality, it follows that

$$\|(\mathcal{I} + \delta \sum_i p_i \mathcal{L}_i)[\cdot] - \sum_i p_i \mathrm{Tr}_{c+1} \boldsymbol{C}_i \Big(|0^{c+1}\rangle\langle 0^{c+1}| \otimes [\cdot]\Big) \boldsymbol{C}_i^\dagger\|_\diamond = O\left(\delta^2\right).$$

Since $\|(\mathcal{I} + \delta\mathcal{L}) - \mathrm{e}^{\delta\mathcal{L}}\|_\diamond = O\left(\delta^2\right)$, by the triangle inequality, we get the sought inequality in (2.25). ∎

Now, we turn to our second incoherent simulation result that is roughly based on the algorithm of [49] but contains further improvements and fixes. We obtain improved complexity because we assume that the Lindbladian is provided via a block-encoding, while effectively [49] construct a (potentially suboptimal) block-encoding within their algorithm. Their complexity depends on $\sum_{j\in J} \|\boldsymbol{L}_j^\dagger \boldsymbol{L}_j\|$, while our algorithm can in

principle achieve a dependence like $\|\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j\|$ when an efficient block-encoding is provided — which is the case for our explicit block-encodings outlined in the next section (section 2.3). To our knowledge, this is the first Lindbladian simulation algorithm that achieves both near-linear time dependence and a complexity that scales with $\|\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j\|$. Note that this improvement looks similar to how [12] improved over [92] on the complexity of estimating multiple expectation values, but the techniques are very different. Here, the improvement stems from the following efficient block-encoding construction.

Proposition 2.3.1. *Given a block-encoding of a Lindbladian (Definition 2.1.2), we get a block-encoding of*

$$\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j \quad \textit{via} \quad \boldsymbol{V} := (\boldsymbol{Y}_{\frac{1}{2}} \otimes \boldsymbol{U}^\dagger) \cdot \left(2|0^{b+1}\rangle\langle 0^{b+1}| \otimes \boldsymbol{I} - \boldsymbol{I}\right) \cdot (\boldsymbol{Y}_{\frac{1}{2}} \otimes \boldsymbol{U}),$$

where $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$.

Proof. We calculate

$$\begin{aligned}
&(\langle 0^{c+1}| \otimes \boldsymbol{I}) \cdot \boldsymbol{V} \cdot (|0^{c+1}\rangle \otimes \boldsymbol{I}) \\
&= \left(\langle -| \otimes (\langle 0^c| \otimes \boldsymbol{I})\boldsymbol{U}^\dagger\right) \cdot \left(2|0^{b+1}\rangle\langle 0^{b+1}| \otimes \boldsymbol{I} - \boldsymbol{I}\right) \cdot \left(|+\rangle \otimes \boldsymbol{U}(|0^c\rangle \otimes \boldsymbol{I})\right) \\
&= \left(\langle -| \otimes (\langle 0^c| \otimes \boldsymbol{I})\boldsymbol{U}^\dagger\right) \cdot \left(2|0^{b+1}\rangle\langle 0^{b+1}| \otimes \boldsymbol{I}\right) \cdot \left(|+\rangle \otimes \boldsymbol{U}(|0^c\rangle \otimes \boldsymbol{I})\right) \\
&= (\langle 0^c| \otimes \boldsymbol{I}) \cdot \boldsymbol{U}^\dagger \cdot (|0^b\rangle\langle 0^b| \otimes \boldsymbol{I}) \cdot \boldsymbol{U} \cdot (|0^c\rangle \otimes \boldsymbol{I}) \\
&= \left(\sum_{j\in J} \langle j| \otimes \boldsymbol{L}_j^\dagger\right)\left(\sum_{j'\in J} |j'\rangle \otimes \boldsymbol{L}_{j'}\right) = \sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j.
\end{aligned}$$

■

This block-encoding construction and the following generic Lindbladian simulation algorithm answers an open question[30] recently posed by Rall, Wang, and Wocjan [149], and can significantly improve their complexity. Although we do not use the above block-encoding explicitly, this observation is implicitly used in our weak measurement schemes (Figure 2.3-Figure 2.10) that enable us to prove the following result, whose proof is presented in section 2.11.

Theorem 2.3.2 (Compressed incoherent Lindbladian simulation algorithm)**.** *Suppose* $\boldsymbol{U}$ *is a block-encoding of the Lindblad operators of a purely irreversible Lindbladian*

[30] See [149, Section 7] "That one special Kraus operator involves all the L_j's. Does there exist any special treatment of this special Kraus operator so that we can leverage the special structure of the oracle $\sum_j |j\rangle \otimes L_j$ to get rid of the $O(m)$ dependence?"

$\mathcal{L}$ as in Definition 2.1.2. Let $\epsilon \le 1/2$, then we can simulate the action of the superoperator $e^{t\mathcal{L}}$ to precision ϵ in diamond norm using

$$\begin{aligned} &O((c+\log((t+1)/\epsilon))\log((t+1)/\epsilon)) && \text{(resettable) ancilla qubits,}\\ &O\left((t+1)\frac{\log((t+1)/\epsilon)}{\log\log((t+1)/\epsilon)}\right) && \text{(controlled) uses of } \boldsymbol{U} \text{ and } \boldsymbol{U}^\dagger,\\ \text{and}\quad &O((t+1)(c+1)\mathrm{polylog}((t+1)/\epsilon)) && \text{other two-qubit gates.}\end{aligned}$$

If the Lindbladian has a coherent part $-\mathrm{i}[\boldsymbol{H},\boldsymbol{\rho}]$, and we have access to a block-encoding of $\boldsymbol{H} = (\langle 0^c|\otimes \boldsymbol{I})\boldsymbol{V}(|0^c\rangle\otimes\boldsymbol{I})$, then we can simulate $e^{t\mathcal{L}}$ with $O\left((t+1)\frac{\log((t+1)/\epsilon)}{\log\log((t+1)/\epsilon)}\right)$ additional (controlled) uses of $\boldsymbol{V}$ and $\boldsymbol{V}^\dagger$.

Crucially, the complexity scales almost linearly with time t and poly-logarithmic with the precision ϵ while using very few ancillas, representing a large asymptotic speedup compared to the t^2/ϵ complexity of the weak-measurement scheme (Theorem 2.3.1).

Coherent Lindbladian simulation algorithms

With a quantum computer, we further ask for a *coherent* Gibbs sampler that outputs the *purified* distribution ($\left|\sqrt{\boldsymbol{\rho}_\beta}\right\rangle \propto \sum_i e^{-\beta E_i/2}|\psi_i\rangle\otimes\left|\psi_i^*\right\rangle$) on two copies of the Hilbert space. A desirable coherent Gibbs sampler should satisfy the following conditions:

1. A Hermitian operator $\mathbb{C}$ can be efficiently block-encoded on the duplicated Hilbert space,
2. its top-eigenvector is unique and yields the purified state $\left|\sqrt{\boldsymbol{\rho}}\right\rangle$, and
3. there exists an adiabatic path of operators $\mathbb{C}(s)$ whose top-eigenvalue-gap remains open.

A general coherent Gibbs sampler may not refer to an existing Lindbladian. However, a natural candidate of the operator $\mathbb{C}$ is to take the *vectorized* discriminant $\mathcal{D}(\boldsymbol{\rho},\mathcal{L})^\dagger$ associated with a detailed balance Lindbladian (Definition 2.2.1), as how one quantizes classical Markov chains [166]. Formally, we define *vectorization* of a superoperator by[31]

$$C[\cdot] = \sum_j \alpha_j \boldsymbol{A}_j[\cdot]\boldsymbol{B}_j \to \boldsymbol{C} = \sum_j \alpha_j \boldsymbol{A}_j\otimes\boldsymbol{B}_j^T \quad \text{(vectorization)},$$

[31] One might be tempted to use $\boldsymbol{B}^\dagger$ instead in the vectorization, but that definition leads to inconsistencies. Indeed, if we would use $\boldsymbol{B}^\dagger$ for vectorization then the two different representations of the scalar $1\otimes 1 = 1 = \mathrm{i}\otimes -\mathrm{i}$ would lead to different vectorizations ± 1.

where $\boldsymbol{B}_j^T$ denotes the transpose of the matrix $\boldsymbol{B}_j$ in the computational basis $|i\rangle$. We use curly fonts $\mathbb{C}$ for superoperators and bold fonts $\mathbb{C}$ for the vectorized superoperators (which is, a matrix).[32] For a matrix $\boldsymbol{A}$, let us denote its vectorized (or purified) version by

$$|\boldsymbol{A}\rangle := (I \otimes T^{-1})\boldsymbol{A} \quad \text{(purification)}$$

using the "transpose" map $T|i\rangle = \langle i|$. This automatically ensures the correctness of the fixed point (Point 2).

Proposition 2.3.2. *For any full-rank state $\boldsymbol{\rho}$ and any Lindbladian $\mathcal{L}$, we have that $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger \big|\sqrt{\boldsymbol{\rho}}\big\rangle = 0$. Further, if $\mathcal{L}$ satisfies $\boldsymbol{\rho}$-detailed balance, we also have that $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})\big|\sqrt{\boldsymbol{\rho}}\big\rangle = 0$.*

The above follows from a direct calculation using that any Lindbladian is trace-preserving $\mathcal{L}^\dagger[\boldsymbol{I}] = 0$. However, to turn the above into the advertised efficient algorithm (Theorem 2.1.4), we need two key components reminiscent of the incoherent case. First, we need a proxy for the discriminant with efficient block-encoding. As we mentioned (2.18), we consider

$$\begin{aligned} \mathcal{D}_\beta = & \sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})\gamma(-\bar{\omega})}\hat{\boldsymbol{A}}^a(\bar{\omega}) \otimes \hat{\boldsymbol{A}}^a(\bar{\omega})^* \\ & - \frac{\gamma(\bar{\omega})}{2}\left(\hat{\boldsymbol{A}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{A}}^a(\bar{\omega}) \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \hat{\boldsymbol{A}}^a(\bar{\omega})^{*\dagger} \hat{\boldsymbol{A}}^a(\bar{\omega})^*\right) \\ & \text{as a proxy for} \quad \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_\beta)^\dagger, \end{aligned} \tag{2.26}$$

where the $\hat{\boldsymbol{A}}^a(\bar{\omega})$ are the same operator Fourier Transforms as in the incoherent case (2.7); the block-encoding for $\mathcal{D}$ can be obtained given the block-encoding for $\hat{\boldsymbol{A}}^a(\bar{\omega})$ (section 2.3). The map is self-adjoint $\mathcal{D}_\beta = \mathcal{D}_\beta^\dagger$ due to Hermiticity $\boldsymbol{A}^a = \boldsymbol{A}^{a\dagger}$ and properties of weighted Fourier Transform (section 2.5). More carefully, we do not implement exactly the discriminant $\mathcal{D}_\beta$, but merely an approximation $\mathcal{D}_{impl} \approx \mathcal{D}_\beta$ due to additional implementation errors for the Gaussian weight and truncation errors for the Gaussian tail.

Second, we need to formulate a notion of approximate detailed balance for the above discriminant proxy.

[32] Note that $\boldsymbol{C}^\dagger$ is well defined. The (matrix) adjoint of the vectorized operator is $\sum_j \alpha_j^* \boldsymbol{A}_j^\dagger \otimes \boldsymbol{B}_j^*$. On the other hand, the superoperator adjoint $C^\dagger[\cdot]$ is $\sum_j \alpha_j^* \boldsymbol{A}_j^\dagger[\cdot]\boldsymbol{B}_j^\dagger$ (2.19), whose vectorization is then the same $\sum_j \alpha_j^* \boldsymbol{A}_j^\dagger \otimes \boldsymbol{B}_j^*$.

Definition 2.3.1 (ϵ-Discriminant proxy)**.** *We say a Hermitian matrix $\mathcal{D}$ is an ϵ-discriminant proxy for Lindbladian $\mathcal{L}$ and a full-rank state $\boldsymbol{\rho}$ if*

$$\|\mathcal{D} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger\| \leq \epsilon.$$

Indeed, this implies approximate detailed balance $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger \approx \mathcal{D}(\boldsymbol{\rho}, \mathcal{L})$ for the Lindbladian $\mathcal{L}$ by taking the adjoints. We can think of Definition 2.3.1 as a different form of the approximate detailed balance condition (Definition 2.2.2) that controls the top eigenvector error up to the spectral gap (Point 2).

Proposition 2.3.3 (Fixed point error)**.** *Suppose a gapped Hermitian operator $\mathcal{D}$ is an ϵ-discriminant proxy for a Lindbladian $\mathcal{L}$ and a full-rank state $\boldsymbol{\rho}$. Then, its top eigenvector is approximately the purified state $\left|\sqrt{\boldsymbol{\rho}}\right\rangle$*

$$\left\||\lambda_1(\mathcal{D})\rangle - \left|\sqrt{\boldsymbol{\rho}}\right\rangle\right\| \leq 4\sqrt{2}\frac{\epsilon}{\lambda_{gap}(\mathcal{D})}.$$

Here, the gap dependence naturally arises from eigenvalue (Corollary 2.10.1) and eigenvector perturbation (Proposition 2.10.2) arguments. Unlike Lindbladians, the cost for preparing the coherent Gibbs state scales directly with the gap via quantum simulated annealing; the mixing time of the original Lindbladian is not linked directly to the algorithmic cost.

Thirdly, the block-encoding by itself does not prepare the desired top eigenvector, unlike a Lindbladian; this additionally requires a standard subroutine called *quantum simulated annealing* [179, 186]: adiabatically change the inverse temperature from $\beta' = 0 \longrightarrow \beta' = \beta$. The algorithmic cost is associated with the gaps along the adiabatic path [24]; see section 2.12.

Explicit block-encodings

In this section, we lay out the circuit ingredients to construct the advertised Lindbladians $\mathcal{L}_\beta$ and discriminants $\mathcal{D}_\beta$. First, we show how to construct a block-encoding of the discretized Lindbladian with Lindblad operators (2.16) from a block-encoding of the jump operators $\boldsymbol{A}^a$. Then, we further construct a block-encoding of the corresponding discriminant proxy - with the additional assumption that the set of jump operators is self-adjoint ($\{\boldsymbol{A}^a : a \in A\} = \{\boldsymbol{A}^{a\dagger} : a \in A\}$) and the Fourier weight function f is real. It is not surprising that implementing the discriminant proxy requires more symmetry constraints, as its definition already implicitly draws from these symmetries.

Block-encoding $\mathcal{L}_\beta$

For both incoherent algorithms (Theorem 2.3.1,Theorem 2.3.2), we have assumed that a purely-irreversible Lindbladian $\mathcal{L}[\cdot] = \sum_{j\in J} \boldsymbol{L}_j[\cdot]\boldsymbol{L}_j^\dagger + \{\boldsymbol{L}_j^\dagger \boldsymbol{L}_j, \cdot\}$ is given by a unitary block-encoding $\boldsymbol{U}$ (Definition 2.1.2). Here, we construct a block-encoding unitary $\boldsymbol{U}$ for the advertised Lindblad operators $\sqrt{\gamma(\bar{\omega})}\hat{\boldsymbol{A}}^a(\bar{\omega})$ labeled by $a, \bar{\omega}$:

$$\sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})}|\bar{\omega}\rangle \otimes |a\rangle \otimes \hat{\boldsymbol{A}}^a(\bar{\omega})$$

for the advertised Lindbladian (discretization of (2.8), recap of (2.21)) .

$$\mathcal{L}_\beta := \sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \gamma(\bar{\omega})\left(\hat{\boldsymbol{A}}^a(\bar{\omega})[\cdot]\hat{\boldsymbol{A}}^a(\bar{\omega})^\dagger - \frac{1}{2}\{\hat{\boldsymbol{A}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{A}}^a(\bar{\omega}), \cdot\}\right).$$

We begin by laying out the registers explicitly, including the additional ancillae for block-encoding.

$$\text{registers:} \quad \underbrace{|0\rangle}_{\text{Boltz. weight}} \otimes \underbrace{|\bar{\omega}\rangle}_{\text{Bohr freq.}} \otimes \underbrace{|0^b\rangle}_{\text{block. enc. anc.}} \otimes \underbrace{|a\rangle}_{\text{jump}} \otimes \underbrace{\hat{\boldsymbol{A}}^a(\bar{\omega})}_{\text{system}}$$

From right to left, the registers individually correspond to: the physical system of interest; the jump labels $|a\rangle$ and additional ancillae to accommodate block-encoding access (indicating successful application by the all-zero state $|0^b\rangle$); the frequency register is dedicated to the operator Fourier Transform, storing the weight $|f\rangle$ or the Bohr frequencies $|\bar{\omega}\rangle$; finally, an ancilla qubit for storing the Bohr-frequency dependent Boltzmann weights in the amplitudes $\sqrt{\gamma(\bar{\omega})}|0\rangle + \sqrt{1-\gamma(\bar{\omega})}|1\rangle$.

We specify the discrete Fourier Transform parameters that determine the dimension of the Bohr frequency register as follows. The Fourier frequencies $\bar{\omega}$ and times $\bar{t}$ are integer multiples of ω_0 and t_0 respectively such that

$$\omega_0 t_0 = \frac{2\pi}{N}, \quad \text{and} \quad S^{\lceil N\rfloor} := \left\{-\lceil (N-1)/2\rceil, \ldots, -1, 0, 1, \ldots, \lfloor (N-1)/2\rfloor\right\},$$
$$\text{and} \quad S_{\omega_0}^{\lceil N\rfloor} := \omega_0 \cdot S^{\lceil N\rfloor}, \quad S_{t_0}^{\lceil N\rfloor} := t_0 \cdot S^{\lceil N\rfloor}.$$

We use a "bar" to denote discretized variables; the (Bohr) frequency register takes values $\bar{\omega} \in S_{\omega_0}$. To implement the Fourier Transform when $N = 2^n$, we specify the signed binary representation for the integers $S^{\lceil N\rfloor}$ as follows:

$$10^{n-1}, \ldots, 1^n \quad \text{for each} \quad -N/2, \ldots, -1,$$
$$0^n, \ldots, 01^{n-1} \quad \text{for each} \quad 0, 1, \ldots, N/2-1.$$

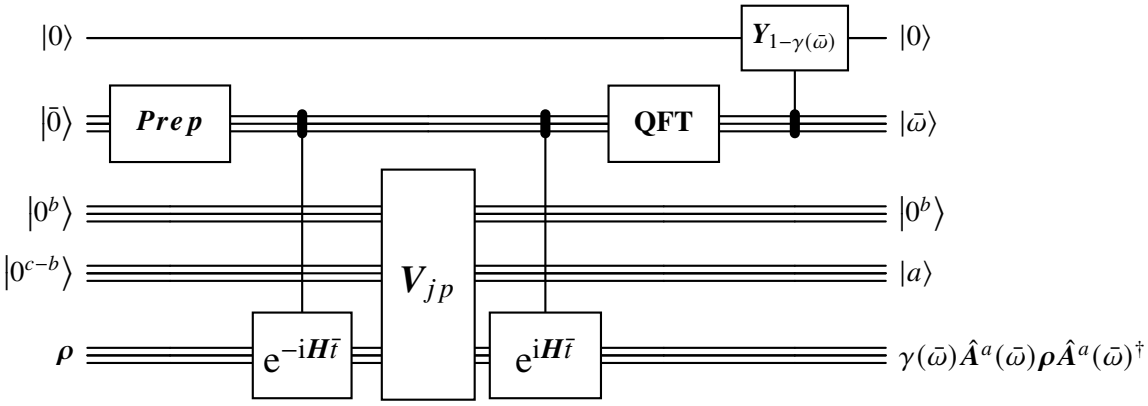


Figure 2.4: Circuit $\boldsymbol{U}$ for block-encoding the Lindbladian. Practically, if we use the simpler weak-measurement-based simulation (Theorem 2.3.1), then by Corollary 2.3.1, we can use a single randomly chosen Lindblad operator $\boldsymbol{A}^a$ at a time. Moreover, if $\boldsymbol{A}^a$ is unitary, we can simply replace $\boldsymbol{V}_{jump}$ with $\boldsymbol{A}^a$, implying $b = c = 0$, i.e., the third and the forth registers can be omitted, thus $n + \lceil \log(N) \rceil + 2$ qubits suffice to simulate the Lindbladian $e^{\mathcal{L}t}$.

Where it does not cause confusion we will drop $\lceil N \rfloor$ from the superscript and will simply write $\bar{\omega} \in S_{\omega_0}$ and $\bar{t} \in S_{t_0}$. We will set the value of N and ω_0 such that the Bohr frequencies $B = \mathrm{spec}(\boldsymbol{H}) - \mathrm{spec}(\boldsymbol{H})$ are contained within the range of energies

$$\|\boldsymbol{H}\| \leq \frac{N}{2}\omega_0.$$

The only "physical" energy scale in the above is the Hamiltonian strength $\|\boldsymbol{H}\|$. Indeed, setting the scale to N requires only $\log(N)$ qubits in the readout register. The circuit consists of the following ingredients, each acting on some appropriate subset of the registers:

- Block-encoding $\boldsymbol{V}_{jump}$ of the jump operators $\boldsymbol{A}^a$ in the form of Definition 2.1.2:

$$(\langle 0^b| \otimes \boldsymbol{I}_a \otimes \boldsymbol{I}_{sys})] \cdot \boldsymbol{V}_{jump} \cdot (|0^c\rangle \otimes \boldsymbol{I}_{sys}) = \sum_{a \in A} |a\rangle \otimes \boldsymbol{A}^a. \tag{2.27}$$

  The operators $\boldsymbol{A}^a$ need not be self-adjoint nor proportional to a unitary. Still, one may conveniently choose $\sqrt{|A|}\boldsymbol{A}^a$ to be unitary for all $a \in A$, (e.g., few-body unitary operators). Then, we can set $b = 0$ and choose

$$\boldsymbol{V}_{jump} = \left(\sum_{a \in A} |a\rangle\langle a| \otimes \boldsymbol{A}^a \right) \cdot (\boldsymbol{B} \otimes \boldsymbol{I}_{sys}) \quad \text{where} \quad \boldsymbol{B}|0^c\rangle = \sum_{a \in A} \frac{|a\rangle}{\sqrt{|A|}} \tag{2.28}$$

Note that implementing the Lindbladian does not require the set of jump operators to contain the adjoints $\{\boldsymbol{A}^a : a \in A\} = \{\boldsymbol{A}^{a\dagger} : a \in A\}$; this assumption is only used for approximate detailed balance and the fixed point correctness (Theorem 2.1.3).

- Controlled Hamiltonian simulation

$$\sum_{\bar{t}\in S_{t_0}} |\bar{t}\rangle\langle\bar{t}| \otimes e^{\pm i\bar{t}\boldsymbol{H}}.$$

- Quantum Fourier Transform

$$\boldsymbol{QFT}_N : |\bar{t}\rangle \to \frac{1}{\sqrt{N}} \sum_{\bar{\omega}\in S_{\omega_0}} \mathrm{e}^{-i\bar{\omega}\bar{t}}|\bar{\omega}\rangle.$$

- State preparation unitary for the Fourier Transform weights, acting on the frequency register

$$\boldsymbol{Prep}_f \quad \text{such that} \quad \boldsymbol{Prep}_f|\bar{0}\rangle = |f\rangle.$$

Naturally, the weight $f(\bar{t})$ as amplitudes of a state is normalized

$$\sum_{\bar{t}\in S_{t_0}} |f(\bar{t})|^2 = 1.$$

It could be, e.g., an easily preparable step function or a Gaussian whose tail decays rapidly. Gaussian states are attractive because they are relatively easy to prepare [127], but as a matter of fact, any other so-called *window function* could be used, such as the Kaiser-window [20, 127] potentially providing further overhead improvements.

- Controlled filter for the Boltzmann factors acting on the frequency register and the Boltzmann weight register

$$\boldsymbol{W} := \sum_{\bar{\omega}\in S_{\omega_0}} \boldsymbol{Y}_{1-\gamma(\bar{\omega})} \otimes |\bar{\omega}\rangle\langle\bar{\omega}| \quad \text{where} \quad 0 \le \gamma(\bar{\omega}) \le 1 \quad \text{and} \quad \gamma(\bar{\omega}) = \gamma(-\bar{\omega})\mathrm{e}^{-\beta\bar{\omega}}.$$

The constraint $0 \le \gamma(\bar{\omega}) \le 1$ ensures the matrix $\boldsymbol{Y}_{1-\gamma(\bar{\omega})}$ is unitary; the symmetry (i.e., the KMS condition) $\gamma(\bar{\omega}) = \gamma(-\bar{\omega})\mathrm{e}^{-\beta\bar{\omega}}$ gives lower weights for "heating" transitions and is closely related to the detailed balance condition. Important examples of weight functions are

$$\text{(Metropolis)} \quad \gamma(\bar{\omega}) = \min(1, \mathrm{e}^{-\beta\bar{\omega}}) \quad \text{and} \quad \text{(Glauber)} \quad \gamma(\bar{\omega}) = \frac{1}{\mathrm{e}^{\beta\bar{\omega}} + 1},$$

which both reduce to the step function in the $\beta \to \infty$ limit. Note that the range of energy labels $\bar{\omega}$ is finite; we choose a large energy readout range $N\omega_0 \geq 4\|\boldsymbol{H}\| + \frac{2}{\beta}$ to ensure all possible transitions are covered by the discretization range (after secular approximation $\mu \leq \frac{1}{\beta}$). We may generally synthesize the controlled filter from elementary gates at cost[33]

$$(\text{polynomial degree of } \gamma) \times \text{Poly}(\log(N), \log(1/\epsilon)),$$

which is $\tilde{O}(1+\beta\|\boldsymbol{H}\|)$ for the Glauber weight. For the Metropolis weight, one can achieve the same scaling by manually switching between 1 and $\mathrm{e}^{-\beta\bar{\omega}}$ at $\bar{\omega} = 0$. In principle, since we are merely controlling a qubit, we may directly implement any efficient computable function (perhaps with terrible overhead).

Further, combining the controlled Hamiltonian simulation and Quantum Fourier Transform yields the advertised operator Fourier Transform (Figure 2.5) acting on the frequency and system register [34]

$$\mathcal{F}[\cdot] : |f(\bar{t})\rangle \otimes \boldsymbol{O} \to \sum_{\bar{\omega}\in S_{\omega_0}} |\bar{\omega}\rangle \otimes \hat{\boldsymbol{O}}_f(\bar{\omega})$$
$$\text{where} \quad \hat{\boldsymbol{O}}_f(\bar{\omega}) := \frac{1}{\sqrt{N}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} f(\bar{t}) \boldsymbol{O}(\bar{t}) \quad \text{and} \quad \boldsymbol{O}(t) := \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{O} \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}.$$

See section 2.5 for basic properties of the operator Fourier Transform. Our implementation is inspired by, but differs from [180]; they sandwich the jump operators with phase estimation and its inverse. The operator picture, inspired by physics, is more natural and tangible. Our construction allows for flexibility in the choice of the weight function $f(\bar{t})$.

As shown in Figure 2.4, we assemble the above ingredients to obtain the unitary $\boldsymbol{U}$ such that

$$\boldsymbol{U}(\boldsymbol{I} \otimes |\bar{0}\rangle \otimes \boldsymbol{I}_{\boldsymbol{V}}) = (\boldsymbol{W} \otimes \boldsymbol{I}_{\boldsymbol{V}}) \cdot \big(\boldsymbol{I} \otimes \mathcal{F}[\boldsymbol{Prep}_f |\bar{0}\rangle \otimes \boldsymbol{V}_{jump}]\big), \tag{2.29}$$

where $\boldsymbol{I}$ is the single qubit identity and $\boldsymbol{I}_{\boldsymbol{V}}$ is the identity on the registers on which $\boldsymbol{V}_{jump}$ acts. As described in Footnote 34, intuitively speaking, we *only* apply the operator Fourier Transform on the "system" register; indeed, in the corresponding

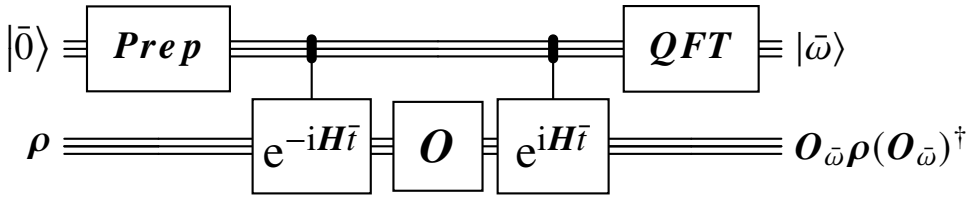


Figure 2.5: Circuit for operator Fourier Transform $\mathcal{F}$ for an operator $\boldsymbol{O}$ acting on the system $\boldsymbol{\rho}$. Of course, in our use, the operator may also act nontrivially on other ancillas.

circuit (Figure 2.5), the jump label register $|a\rangle$ and the block-encoding ancillae $|0^b\rangle$ is only affected by $\boldsymbol{V}_{jump}$.

Lemma 2.3.1 (Explicit Block-encoding)**.** *The untiary* $\boldsymbol{U}$ *in* (2.29) *gives a block-encoding for our Lindbladian* (2.8)

$$\begin{aligned}&\left(\langle 0|\otimes \boldsymbol{I}_{\bar{\omega}}\otimes \langle 0^b|\otimes \boldsymbol{I}_a\otimes \boldsymbol{I}_{sys}\right)\cdot \boldsymbol{U}\cdot \left(|0\rangle\otimes|\bar{0}\rangle\otimes|0^c\rangle\otimes \boldsymbol{I}_{sys}\right)\\ &=\sum_{a\in A,\bar{\omega}\in S_{\omega_0}}\sqrt{\gamma(\bar{\omega})}|\bar{\omega}\rangle\otimes|a\rangle\otimes \hat{\boldsymbol{A}}^a_f(\bar{\omega}).\end{aligned}\tag{2.30}$$

Proof.

(LHS of (2.30))

$$=\left(\langle 0|\otimes \boldsymbol{I}_{\bar{\omega}}\otimes \langle 0^b|\otimes \boldsymbol{I}_a\otimes \boldsymbol{I}_{sys}\right)\cdot(\boldsymbol{W}\otimes \boldsymbol{I}_{\boldsymbol{V}})\cdot\left(|0\rangle\otimes\mathcal{F}[|f\rangle\otimes \boldsymbol{V}_{jump}]\left(|0^c\rangle\otimes \boldsymbol{I}_{sys}\right)\right)$$

(by (2.29))

$$=((\langle 0|\otimes \boldsymbol{I}_{\bar{\omega}})\boldsymbol{W}(|0\rangle\otimes \boldsymbol{I}_{\bar{\omega}})\otimes \boldsymbol{I}_a\otimes \boldsymbol{I}_{sys})\cdot\left(\left(\boldsymbol{I}_{\bar{\omega}}\otimes\langle 0^b|\otimes \boldsymbol{I}_a\otimes \boldsymbol{I}_{sys}\right)\mathcal{F}[|f\rangle\otimes \boldsymbol{V}_{jump}]\left(|0^c\rangle\otimes \boldsymbol{I}_{sys}\right)\right)$$

$$=\sum_{a\in A,\bar{\omega}\in S_{\omega_0}}(\langle 0|\boldsymbol{Y}_{1-\gamma(\bar{\omega})}|0\rangle\otimes|\bar{\omega}\rangle\langle\bar{\omega}|\otimes \boldsymbol{I}_a\otimes \boldsymbol{I}_{sys})\cdot(\mathcal{F}[|f\rangle\otimes|a\rangle\otimes \boldsymbol{A}^a])$$

(by (2.27) and Figure 2.5)

$$=\sum_{a\in A,\bar{\omega}\in S_{\omega_0}}\sqrt{\gamma(\bar{\omega})}|\bar{\omega}\rangle\otimes|a\rangle\otimes \hat{\boldsymbol{A}}^a_f(\bar{\omega}).\qquad\text{(by OFT and controlled filter)}\quad\blacksquare$$

[33] By first applying controlled Hamiltonian simulation for $\boldsymbol{Y}$ rotation $\sum_{\bar{\omega}} e^{i(\bar{\omega}/2\|\boldsymbol{H}\|)\boldsymbol{Y}}\otimes|\bar{\omega}\rangle\langle\bar{\omega}|$ and then apply QSVT to map $\bar{\omega}/2\|\boldsymbol{H}\|\to\arcsin(\sqrt{1-\gamma(\bar{\omega})})$. This is reminiscent of [79].

[34] If the operator $\boldsymbol{O}$ maps between larger Hilbert spaces than $\boldsymbol{H}$, we formally extend its action trivially so that $\boldsymbol{O}(t)=(\boldsymbol{I}\otimes e^{i\boldsymbol{H}t})\boldsymbol{O}(\boldsymbol{I}'\otimes e^{-i\boldsymbol{H}t})$. Of course, this formal extension does not incur any additional cost.

Block-encoding $\mathcal{D}_\beta$

We now describe the explicit and efficient circuit that implements the advertised vectorized discriminant proxy (discretization of (2.18), recap of (2.26))

$$\mathcal{D}_\beta = \sum_{a\in A,\bar\omega\in S_{\omega_0}} \sqrt{\gamma(\bar\omega)\gamma(-\bar\omega)}\hat{\boldsymbol{A}}^a(\bar\omega)\otimes\hat{\boldsymbol{A}}^a(\bar\omega)^* - \frac{\gamma(\bar\omega)}{2}\Big(\hat{\boldsymbol{A}}^a(\bar\omega)^\dagger\hat{\boldsymbol{A}}^a(\bar\omega)\otimes\boldsymbol{I} + \boldsymbol{I}\otimes\hat{\boldsymbol{A}}^a(\bar\omega)^{*\dagger}\hat{\boldsymbol{A}}^a(\bar\omega)^*\Big),$$

assuming that the set of jump operators is self-adjoint in the sense that $\{\boldsymbol{A}^a : a\in A\} = \{\boldsymbol{A}^{a\dagger} : a\in A\}$ and the Fourier weight function f is real. Combining the circuit with simulated annealing then leads to the advertised quadratic speedup.

Our discriminant proxy and its block-encoding is an instantiation of the following general construct that is self-adjoint as a superoperator (and hence Hermitian after vectorization)[35].

Proposition 2.3.4 (Self-adjoint discriminant proxies)**.** *Given a purely irreversible Lindbladian with Lindblad operators $\boldsymbol{L}_j$ for $j\in J$, and a permutation $\boldsymbol{P}: j\to j'$ on the set J, the following superoperator (and hence its vectorization) is self-adjoint:*

$$\mathcal{D}(\boldsymbol{P},\{\boldsymbol{L}_j\}) := \frac{1}{2}\sum_{j\in J}\boldsymbol{L}_j[\cdot]\boldsymbol{L}_{j'}^\dagger + \boldsymbol{L}_j^\dagger[\cdot]\boldsymbol{L}_{j'} - \{\boldsymbol{L}_j^\dagger\boldsymbol{L}_j,\cdot\},$$

$$\mathcal{D}(\boldsymbol{P},\{\boldsymbol{L}_j\}) = \frac{1}{2}\sum_{j\in J}\boldsymbol{L}_j\otimes\boldsymbol{L}_{j'}^{*\dagger} + \boldsymbol{L}_j^\dagger\otimes\boldsymbol{L}_{j'}^* - \boldsymbol{L}_j^\dagger\boldsymbol{L}_j\otimes\boldsymbol{I} - \boldsymbol{I}\otimes\boldsymbol{L}_j^{*\dagger}\boldsymbol{L}_j^*.$$

Proof.

$$\mathcal{D}(\boldsymbol{P},\{\boldsymbol{L}_j\})^\dagger = \frac{1}{2}\sum_{j\in J}\boldsymbol{L}_j^\dagger[\cdot]\boldsymbol{L}_{j'} + \boldsymbol{L}_j[\cdot]\boldsymbol{L}_{j'}^\dagger - \{\boldsymbol{L}_j^\dagger\boldsymbol{L}_j,\cdot\} = \mathcal{D}(\boldsymbol{P},\{\boldsymbol{L}_j\}). \qquad \blacksquare$$

The vectorization is hence also self-adjoint.

Corollary 2.3.2. *If f is real (2.11) and the set of jump operators is self-adjoint (2.10), then the discriminant proxy (2.26) is Hermitian.*

Proof. Due to footnote 32, it suffices to verify that the superoperator is self-adjoint: This follows from Proposition 2.3.4 by setting the permutation $\boldsymbol{P}: (\bar\omega,a)\to(-\bar\omega,a')$

[35] If we additionally have $\sum_{j\in J}\frac{1}{2}\boldsymbol{L}_j\otimes\boldsymbol{L}_{j'}^{*\dagger} + \frac{1}{2}\boldsymbol{L}_{j'}^\dagger\otimes\boldsymbol{L}_j^* \approx \sum_{j\in J}\Big(\boldsymbol{\rho}^{-\frac{1}{4}}\otimes(\boldsymbol{\rho}^*)^{-\frac{1}{4}}\Big)\boldsymbol{L}_j\otimes\boldsymbol{L}_j^*\Big(\boldsymbol{\rho}^{\frac{1}{4}}\otimes(\boldsymbol{\rho}^*)^{\frac{1}{4}}\Big)$ and $\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j \approx \sum_{j\in J}\boldsymbol{\rho}^{\frac{1}{4}}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\boldsymbol{\rho}^{-\frac{1}{4}}$, then the resulting discriminant proxy $\mathcal{D}$ is close to the discriminant $\mathcal{D}(\boldsymbol{\rho},\mathcal{L})^\dagger$. This is exactly what we show in 2.6 for the discriminant proxy (2.26).

such that $\boldsymbol{A}^{a\dagger} = \boldsymbol{A}^{a'}$ and using the operator Fourier Transform property $\hat{\boldsymbol{A}}^{a'}(-\bar{\omega})^{\dagger} = \hat{\boldsymbol{A}^{a'\dagger}}(\bar{\omega})$ for real weight f (2.39), implying that $\hat{\boldsymbol{A}}^{a}(\bar{\omega}) \otimes \hat{\boldsymbol{A}}^{a'}(-\bar{\omega})^{\dagger *} = \hat{\boldsymbol{A}}^{a}(\bar{\omega}) \otimes \hat{\boldsymbol{A}}^{a}(\bar{\omega})^{*} = \hat{\boldsymbol{A}}^{a'}(-\bar{\omega})^{\dagger} \otimes \hat{\boldsymbol{A}}^{a}(\bar{\omega})^{*}$. ■

Now that we have verified the symmetries of the desired discriminant proxy, we move on to our explicit construction. By the standard quantum walk recipe [164], we design an isometry and a reflection such that

$$\boldsymbol{I} + \mathcal{D}(\boldsymbol{P}, \{\boldsymbol{L}_j\}) = \boldsymbol{T}'^{\dagger} \boldsymbol{R} \boldsymbol{T}',$$

which is block-encoded as in Figure 2.6.

Proposition 2.3.5 (A block-encoding for discriminant proxies)**.** *Using the notation of Proposition 2.3.4, let*

$$\boldsymbol{R} := \boldsymbol{I} - (\boldsymbol{I} \otimes \boldsymbol{\Pi}) + \underbrace{\boldsymbol{Z} \otimes |0^{b'}\rangle\langle 0^{b'}| \otimes \boldsymbol{P} \otimes \boldsymbol{I}_{sys} \otimes \boldsymbol{I}_{sys'}}_{=:\boldsymbol{R}_0}$$
$$\textit{where} \quad \boldsymbol{\Pi} := |0^{b'}\rangle\langle 0^{b'}| \otimes \boldsymbol{I}_J \otimes \boldsymbol{I}_{sys} \otimes \boldsymbol{I}_{sys'},$$

and $\boldsymbol{Z}$ is the Pauli-Z operator such that $\boldsymbol{Z}|\pm\rangle = |\mp\rangle$ for $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$. If $\boldsymbol{U}$ is a unitary block-encoding of the Lindbladian such that

$$\left(\langle 0^{b'}| \otimes \boldsymbol{I}_J \otimes \boldsymbol{I}_{sys}\right) \cdot \boldsymbol{U} \cdot \left(|0^{c'}\rangle \otimes \boldsymbol{I}_{sys}\right) = \sum_{j \in J} |j\rangle \otimes \boldsymbol{L}_j,$$

then, we obtain a block encoding for the (shifted) discriminant proxy

$$\left(\langle 0^{c'+1}| \otimes \boldsymbol{I}_{sys} \otimes \boldsymbol{I}_{sys'}\right) \cdot \boldsymbol{U}_{\mathcal{D}(\boldsymbol{P},\{\boldsymbol{L}_j\})} \cdot \left(|0^{c'+1}\rangle \otimes \boldsymbol{I}_{sys} \otimes \boldsymbol{I}_{sys'}\right) = \boldsymbol{I} + \mathcal{D}(\boldsymbol{P}, \{\boldsymbol{L}_j\}).$$

using

$$\boldsymbol{U}_{\mathcal{D}(\boldsymbol{P},\{\boldsymbol{L}_j\})} := \boldsymbol{U}'^{\dagger} \cdot \boldsymbol{R} \cdot \boldsymbol{U}' \quad \textit{where} \quad \boldsymbol{U}' = \left(|+\rangle\langle +| \otimes \boldsymbol{U} \otimes \boldsymbol{I}_{sys'} + |-\rangle\langle -| \otimes \boldsymbol{I}_{sys} \otimes \boldsymbol{U}^{*}\right)$$

and the unitary $\boldsymbol{U}^{*}$ is the conjugate of $\boldsymbol{U}$ but acting on a copy of the system register (sys').

Proof. Consider the isometries $\boldsymbol{T} := \boldsymbol{U}(|0^{c'}\rangle \otimes \boldsymbol{I}_{sys})$ and $\boldsymbol{T}^{*} := \boldsymbol{U}^{*}(|0^{c'}\rangle \otimes \boldsymbol{I}_{sys'})$, and

$$\boldsymbol{T}' := \boldsymbol{U}'\left(|0^{c'+1}\rangle \otimes \boldsymbol{I}_{sys} \otimes \boldsymbol{I}_{sys'}\right) = \frac{1}{\sqrt{2}}\left(|+\rangle \otimes \boldsymbol{T} \otimes \boldsymbol{I}_{sys'} + |-\rangle \otimes \boldsymbol{I}_{sys} \otimes \boldsymbol{T}^{*}\right).$$

To understand the product $\boldsymbol{T}'^{\dagger}\boldsymbol{R}\boldsymbol{T}'$, we first calculate $(\boldsymbol{I}\otimes\boldsymbol{\Pi})\boldsymbol{T}'$ and $\boldsymbol{R}_0\boldsymbol{T}'$:

$$\begin{aligned}(\boldsymbol{I}\otimes\boldsymbol{\Pi})\boldsymbol{T}' &= \frac{1}{\sqrt{2}}\sum_{j\in J}|+\rangle\otimes|0^{b'}\rangle\otimes|j\rangle\otimes\boldsymbol{L}_j\otimes\boldsymbol{I}_{sys'}\\ &\quad+\frac{1}{\sqrt{2}}\sum_{j\in J}|-\rangle\otimes|0^{b'}\rangle\otimes|j\rangle\otimes\boldsymbol{I}_{sys}\otimes\boldsymbol{L}_j^*, \qquad (2.31)\\ \boldsymbol{R}_0\boldsymbol{T}' &= \frac{1}{\sqrt{2}}\sum_{j\in J}|-\rangle\otimes|0^{b'}\rangle\otimes|j'\rangle\otimes\boldsymbol{L}_j\otimes\boldsymbol{I}_{sys'}\\ &\quad+\frac{1}{\sqrt{2}}\sum_{j\in J}|+\rangle\otimes|0^{b'}\rangle\otimes|j'\rangle\otimes\boldsymbol{I}_{sys}\otimes\boldsymbol{L}_j^*, \qquad (2.32)\end{aligned}$$

using the bit-flip $\boldsymbol{Z}$ ($+\leftrightarrow-$), and the permutation $\boldsymbol{P}$ ($j\leftrightarrow j'$). Finally, we get that

$$\begin{aligned}\boldsymbol{T}'^{\dagger}\boldsymbol{R}\boldsymbol{T}' &= \boldsymbol{T}'^{\dagger}\boldsymbol{I}\boldsymbol{T}' - \boldsymbol{T}'^{\dagger}(\boldsymbol{I}\otimes\boldsymbol{\Pi})\boldsymbol{T}' + \boldsymbol{T}'^{\dagger}\boldsymbol{R}_0\boldsymbol{T}'\\ &= \boldsymbol{I} - \boldsymbol{T}'^{\dagger}(\boldsymbol{I}\otimes\boldsymbol{\Pi})\cdot(\boldsymbol{I}\otimes\boldsymbol{\Pi})\boldsymbol{T}' + \boldsymbol{T}'^{\dagger}(\boldsymbol{I}\otimes\boldsymbol{\Pi})\cdot\boldsymbol{R}_0\boldsymbol{T}'\\ &\qquad\qquad\text{(since } (\boldsymbol{I}\otimes\boldsymbol{\Pi})\boldsymbol{R}_0=\boldsymbol{R}_0)\\ &= \boldsymbol{I} - \frac{1}{2}\sum_{j\in J}\boldsymbol{L}_j^{\dagger}\boldsymbol{L}_j\otimes\boldsymbol{I}_{sys'} + \boldsymbol{I}_{sys}\otimes\boldsymbol{L}_j^{*\dagger}\boldsymbol{L}_j^* + \frac{1}{2}\sum_{j\in J}\boldsymbol{L}_j\otimes\boldsymbol{L}_{j'}^{*\dagger} + \boldsymbol{L}_{j'}^{\dagger}\otimes\boldsymbol{L}_j^*\\ &\qquad\qquad\text{(by (2.31)-(2.32))}\\ &= \boldsymbol{I} + \mathcal{D}(\boldsymbol{P},\{\boldsymbol{L}_j\}). \qquad\blacksquare\end{aligned}$$

Specializing the above recipe for (2.26) yields a block-encoding of $\mathcal{D}_\beta$ using the following ingredients:

- A unitary block-encoding for the Lindbladian

  $$\Big(\langle 0^{b'}|\otimes\boldsymbol{I}_{\bar\omega}\otimes\boldsymbol{I}_a\otimes\boldsymbol{I}_{sys}\Big)\cdot\boldsymbol{U}\cdot\Big(|0^{c'}\rangle\otimes\boldsymbol{I}_{sys}\Big) = \sum_{a\in A,\bar\omega\in S_{\omega_0}}\sqrt{\gamma(\bar\omega)}|\bar\omega\rangle\otimes|a\rangle\otimes\hat{\boldsymbol{A}}^a(\bar\omega).$$

  An example would be the block-encoding (2.30) instantiating the parameters $b'=b+1$ and $|0^{c'}\rangle=\big|0^{c+1}\big\rangle\big|\bar 0\big\rangle$ after appropriately rearranging the registers.

- Negation on the Bohr frequency register

  $$\boldsymbol{F} := \sum_{\bar\omega\in S_{\omega_0}}|-\bar\omega\rangle\langle\bar\omega| \quad\text{such that}\quad \boldsymbol{F}^2=\boldsymbol{I}_{\bar\omega}.$$

- Permutation (involution) of the jump operator labels

  $$\boldsymbol{P} := \sum_{a\in A}|a'\rangle\langle a| \quad\text{where}\quad \boldsymbol{A}^{a'}=(\boldsymbol{A}^a)^{\dagger} \quad\text{for each}\quad a\in A. \qquad (2.33)$$

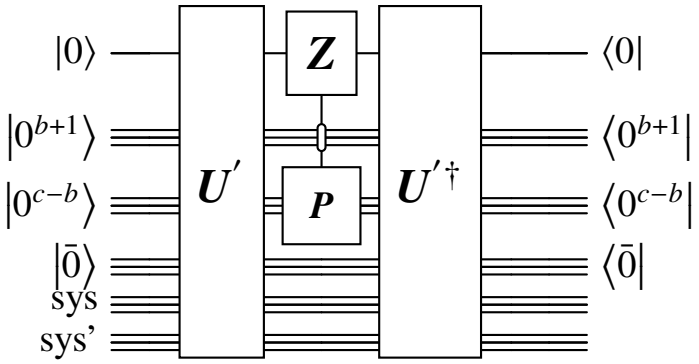


Figure 2.6: Circuit $\boldsymbol{U'}^{\dagger}\boldsymbol{R}\boldsymbol{U'}$ for block-encoding the discriminant in the fashion of Szegedy quantum walk.

> Note that if the jump operators are Hermitian, e.g., Pauli matrices, then we can simply take the permutation to be the identity $\boldsymbol{P} = \boldsymbol{I}_a$.

To reiterate, compared to the incoherent case, the discriminant proxy can be implemented with two extra unitaries $\boldsymbol{F}$ and $\boldsymbol{P}$, an additional copy of the system register (which we denoted by $system'$), and an additional ancilla qubit.

$$\text{registers:} \quad \underbrace{|\pm\rangle}_{\boldsymbol{T}/\boldsymbol{T}^* \text{ selector}} \otimes \underbrace{|0^{b'}\rangle}_{\text{block. enc. anc.}} \otimes \underbrace{|\bar{\omega}\rangle}_{\text{Bohr freq.}} \otimes \underbrace{|a\rangle}_{\text{jump}} \otimes \underbrace{\hat{\boldsymbol{A}}^a(\bar{\omega})}_{\text{system}} \otimes \underbrace{\boldsymbol{A}^a(\bar{\omega})^*}_{\text{system'}}$$

If we combine the constructions of section 2.3-section 2.3, we can see that the number of qubits is

$$\#qubits = 2n + \lceil \log_2(N) \rceil + c + 2,$$

coming from the two copies of the system register, the frequency register, the ancillae for the block-encoding of the jumps, and one additional ancillae introduced in each of section 2.3-section 2.3. When the normalized jump operators $\sqrt{|A|}\boldsymbol{A}^a$ are unitaries, we can have c as small as $\lceil \log_2(|A|) \rceil$; see for example (2.28).

Proof of coherent Gibbs sampler (Theorem 2.1.4)

In this section, we prove guarantees for our coherent Gibbs sampler (Theorem 2.1.4) in a similar vein as the Lindbladian case (Theorem 2.1.3). The required lemmas and propositions are analogous but refer to the spectral gap instead of the mixing time.

Proof of Theorem 2.1.4. We present general bounds on finite N (2.26) and then take the large N limit for the continuum (2.18). We bound the eigenvector distance by the operator norm bounds (Proposition 2.10.1, Proposition 2.10.2 and Proposition 2.3.3): secular approximation error (Lemma 2.6.2 and Proposition 2.6.7), and discriminant proxy (Lemma 2.7.3 and $\mathcal{D}_{sec} = \mathcal{D}_{sec}^{\dagger}$)[36]

$$\begin{aligned}
\| |\lambda_1(\mathcal{D}_\beta)\rangle - |\sqrt{\boldsymbol{\rho}}\rangle \| &\leq \frac{6\|\mathcal{D}_\beta - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\|}{\lambda_{gap}(\mathcal{D}_\beta)} \\
&\leq \frac{6}{\lambda_{gap}(\mathcal{D}_\beta)} \left(\|\mathcal{D}_\beta - \mathcal{D}_{sec}\| + \|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\| \right) \\
&\leq O\left(\frac{(e^{-N^2\omega_0^2\sigma_t^2/2} + e^{-N^2 t_0^2/16\sigma_t^2} + e^{-\bar{\mu}^2\sigma_t^2} + e^{-T^2/4\sigma_t^2} + \omega_0 T) + \beta\bar{\mu}}{\lambda_{gap}(\mathcal{D}_\beta)} \right) \\
&\leq O\left(\frac{\sigma_t\omega_0\sqrt{\log(1/(\sigma_t\omega_0))} + (\beta/\sigma_t)\sqrt{\log(\sigma_t/\beta)} + e^{-N^2\omega_0^2\sigma_t^2/2}}{\lambda_{gap}(\mathcal{D}_\beta)} \right).
\end{aligned}$$

The fourth inequality chooses the free parameter $T = 2\sigma_t\sqrt{\ln(1/(\sigma_t\omega_0))}$ and $\bar{\mu} = \frac{\beta}{\sigma_t}\sqrt{\ln(\frac{\sigma_t}{\beta})}$ and uses that $e^{-N^2 t_0^2/16\sigma_t^2} = e^{-\pi^2/4\omega_0^2\sigma_t^2} = O(\sigma_t\omega_0\sqrt{\log(1/(\sigma_t\omega_0))})$ to reduce the expression.

In the continuum limit (2.18), the discretization parameters ω_0 and N disappear, and the RHS becomes

$$\| |\lambda_1(\mathcal{D}_\beta)\rangle - |\sqrt{\boldsymbol{\rho}}\rangle \| = O\left(\frac{\beta}{\sigma_t} \frac{\sqrt{\log(\sigma_t/\beta)}}{\lambda_{gap}(\mathcal{D}_\beta)} \right)$$

by taking the limit $N\omega_0 \to \infty$, $\omega_0 \to 0$. ∎

Note the user only chooses the time limit T, Gaussian width σ_t, and the Discrete Fourier Transform resolution ω_0 and the number of points N; the truncation parameter $\bar{\mu}$ only appears implicitly in the analysis of secular approximation. Compared with the fixed point error for Lindbladians (section 2.2), the Hermiticity and gap substantially simplifies the analysis.

Metropolis sampling with arbitrary spectral target weights

Looking beyond sampling Gibbs states $\boldsymbol{\rho} \propto \sum_i e^{-\beta E_i}|\psi_i\rangle\langle\psi_i|$, we may modify the circuit to sample from arbitrary weight function $\boldsymbol{\rho} \propto \sum_i p(E_i)|\psi_i\rangle\langle\psi_i|$ that could be

[36] Here, we implicitly assume that $\mathcal{D}_{impl}$ is Hermitian, which holds for example if $\boldsymbol{R}$ in the block-encoding is implemented exactly.

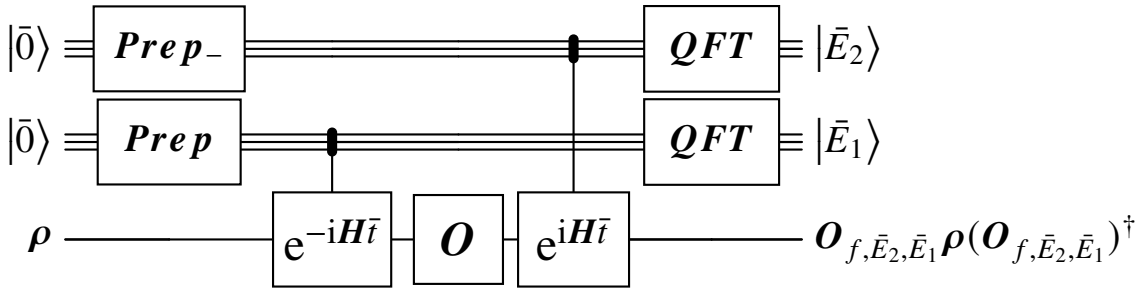


Figure 2.7: The circuit for two-sided operator Fourier Transform. The gate $\boldsymbol{prep}_-$ prepares the flipped function $f_-(t) = f(-t)$. Unlike the one-sided version (Figure 2.5), now we cannot understand the expression by Heisenberg evolution; this is, in spirit, more similar to doing two consecutive phase estimations.

useful in other contexts. We reserve this section to point out the required ingredients and adaptations; we will stick to Gibbs sampling for the rest of the paper, but our coherent and incoherent algorithms should also apply in the general setting.

Instead of applying the operator Fourier Transform for the Bohr frequencies (the energy differences), sampling from arbitrary weight requires accessing *both* energies before and after the jump, resembling [149, 167, 180]. The appropriate Lindbladian takes the following form:

$$\mathcal{L}_{general}[\rho] := \sum_{a\in A,\bar{E}_2,\bar{E}_1\in S_{\omega_0}} \gamma(\bar{E}_2,\bar{E}_1)\left(\hat{A}^a_f(\bar{E}_2,\bar{E}_1)\rho\hat{A}^a_f(\bar{E}_2,\bar{E}_1)^\dagger - \frac{1}{2}\{\hat{A}^a_f(\bar{E}_2,\bar{E}_1)^\dagger\hat{A}^a_f(\bar{E}_2,\bar{E}_1),\rho\}\right). \tag{2.34}$$

To implement a block-encoding of the above Lindbladian, we need different Fourier Transform components and a controlled filter for the two-argument Metropolis weight.

- Phase estimation isometry (in the Schrödinger picture)

$$\Phi_f := \frac{1}{\sqrt{N}}\sum_{\bar{E}\in S_{\omega_0}}\sum_{\bar{t}\in S_{\bar{t}_0}} f(\bar{t})|\bar{E}\rangle \otimes e^{-i(\bar{E}-\boldsymbol{H})\bar{t}}.$$

  This circumvents the impossibility results as it does not take the shift-invariant form (section 2.13).

- Two-sided operator Fourier Transform (section 2.3)

$$\mathcal{F}[|f_2\rangle \otimes |f_1\rangle \otimes \boldsymbol{O}] := (\Phi_{f_2} \otimes \boldsymbol{I}_{\bar{E}_1}) \circ (\boldsymbol{I}_{\bar{E}_1} \otimes \boldsymbol{O}) \circ \Phi_{f_1}$$
$$= \sum_{\bar{E}_2, \bar{E}_1 \in S_{\omega_0}} \left|\bar{E}_2\right\rangle \otimes \left|\bar{E}_1\right\rangle \otimes \hat{\boldsymbol{O}}_{f_1,f_2}(\bar{E}_2, \bar{E}_1),$$

$$\text{where} \quad \hat{\boldsymbol{O}}_{f_1,f_2}(\bar{E}_2, \bar{E}_1) := \frac{1}{N} \sum_{\bar{t}_2, \bar{t}_1 \in S_{\bar{t}_0}} f_2(\bar{t}_2) f_1(\bar{t}_1) \mathrm{e}^{-\mathrm{i}(\bar{E}_2 - \boldsymbol{H})\bar{t}_2} \boldsymbol{O} \mathrm{e}^{-\mathrm{i}(\bar{E}_1 - \boldsymbol{H})\bar{t}_1}.$$

  In our particular case, we set $f_1(\bar{t}) = f(\bar{t})$ and $f_2(\bar{t}) = f(-\bar{t}) =: f_-(\bar{t})$ for a normalized function f to get

$$\boldsymbol{O}_{f,f_-}(\bar{E}_2, \bar{E}_1) = \frac{1}{N} \sum_{\bar{t}_2, \bar{t}_1 \in S_{\bar{t}_0}} f(-\bar{t}_2) f(\bar{t}_1) \mathrm{e}^{-\mathrm{i}(\bar{E}_2 - \boldsymbol{H})\bar{t}_2} \boldsymbol{O} \mathrm{e}^{-\mathrm{i}(\bar{E}_1 - \boldsymbol{H})\bar{t}_1} \quad (2.35)$$

  which will be short-handed by $\boldsymbol{O}_{f,\bar{E}_2,\bar{E}_1}$. This construction is reminiscent of [180], but it does not require a rounding promise.

- Controlled filter for the Metropolis weight

$$\boldsymbol{W} := \sum_{\bar{E}_2, \bar{E}_1 \in S_{\omega_0}} \boldsymbol{Y}_{1-\gamma(\bar{E}_2, \bar{E}_1)} \otimes |\bar{E}_2\rangle\langle\bar{E}_2| \otimes |\bar{E}_1\rangle\langle\bar{E}_1|$$

$$\text{such that} \quad 0 \leq \gamma(\bar{E}_2, \bar{E}_1) \leq 1 \quad \text{and} \quad \frac{\gamma(\bar{E}_2, \bar{E}_1)}{\gamma(\bar{E}_1, \bar{E}_2)} = \frac{p(\bar{E}_2)}{p(\bar{E}_1)}.$$

  The ratio constraint ensures approximate detailed balance.

If the set of jump operators is self-adjoint and f is real, we may construct a valid Hermitian discriminant proxy as outlined in Proposition 2.3.4 — the corresponding involution is $\mathbf{SWAP} \otimes \boldsymbol{P}$, where **SWAP** acts on the two phase estimation registers, and $\boldsymbol{P}$ is a permutation as in (2.33). This construction yields a discriminant proxy analogous to (2.26) due to the following "skew symmetry" of the two-sided operator Fourier Transform. We also include the two-index version of operator Parseval's identity.

Proposition 2.3.6 (Parseval's identity)**.** *For a set of matrices $\boldsymbol{A}^a \in \mathbb{C}^{n \times n}$, consider their two-sided discrete Fourier Transform weighted by a complex-valued function*

$f\colon S_{t_0} \to \mathbb{C}$ as in (2.35). Then, $\hat{\boldsymbol{A}}^a_f(\bar{E}_2,\bar{E}_1)^\dagger = (\hat{\boldsymbol{A}^{a\dagger}})_{f^*}(\bar{E}_2,\bar{E}_1)$, and

$$\begin{aligned}\sum_{a\in A}\sum_{\bar{E}_1,\bar{E}_2\in S_{\omega_0}} \hat{\boldsymbol{A}}^a_f(\bar{E}_2,\bar{E}_1)^\dagger \hat{\boldsymbol{A}}^a_f(\bar{E}_2,\bar{E}_1) &= \|f\|_2^2 \sum_{\bar{t}_1\in S_{t_0}} |f(\bar{t}_1)|^2 \mathrm{e}^{-\mathrm{i}\boldsymbol{H}\bar{t}_1}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\mathrm{e}^{\mathrm{i}\boldsymbol{H}\bar{t}_1} \\ &\preceq \|f\|_2^4 \Big\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\Big\|\cdot\boldsymbol{I} \quad \textit{and} \\ \sum_{a\in A}\sum_{\bar{E}_1,\bar{E}_2\in S_{\omega_0}} \hat{\boldsymbol{A}}^a_f(\bar{E}_2,\bar{E}_1)\hat{\boldsymbol{A}}^a_f(\bar{E}_2,\bar{E}_1)^\dagger &= \|f\|_2^2 \sum_{a\in A}\sum_{\bar{t}_2\in S_{t_0}} |f(\bar{t}_2)|^2 \mathrm{e}^{\mathrm{i}\boldsymbol{H}\bar{t}_2}\boldsymbol{A}^a\boldsymbol{A}^{a\dagger}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}\bar{t}_2} \\ &\preceq \|f\|_2^4 \Big\|\sum_{a\in A}\boldsymbol{A}^a\boldsymbol{A}^{a\dagger}\Big\|\cdot\boldsymbol{I}.\end{aligned}$$

Proof. For each $a\in A$, we have

$$\begin{aligned}\hat{\boldsymbol{A}}^a_f(\bar{E}_2,\bar{E}_1)^\dagger &= \left(\frac{1}{N}\sum_{\bar{t}_2,\bar{t}_1\in S_{\bar{t}_0}} f(-\bar{t}_2)f(\bar{t}_1)\mathrm{e}^{-\mathrm{i}(\bar{E}_2-\boldsymbol{H})\bar{t}_2}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}(\bar{E}_1-\boldsymbol{H})\bar{t}_1}\right)^\dagger \\ &= \left(\frac{1}{N}\sum_{\bar{t}_2,\bar{t}_1\in S_{\bar{t}_0}} f(\bar{t}_2)f(\bar{t}_1)\mathrm{e}^{\mathrm{i}(\bar{E}_2-\boldsymbol{H})\bar{t}_2}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}(\bar{E}_1-\boldsymbol{H})\bar{t}_1}\right)^\dagger \\ &= \frac{1}{N}\sum_{\bar{t}_2,\bar{t}_1\in S_{\bar{t}_0}} f(\bar{t}_2)^*f(\bar{t}_1)^*\mathrm{e}^{\mathrm{i}(\bar{E}_1-\boldsymbol{H})\bar{t}_1}\boldsymbol{A}^{a\dagger}\mathrm{e}^{-\mathrm{i}(\bar{E}_2-\boldsymbol{H})\bar{t}_2} = (\hat{\boldsymbol{A}^{a\dagger}})_{f^*}(\bar{E}_2,\bar{E}_1).\end{aligned}$$

■

Our analytic arguments (section 2.5, section 2.7) can be adapted to the above Lindbladian (2.34) and discriminant variant, but we will stick to the operator Fourier Transform for simplicity throughout the paper.

2.4 Discussion

Our work aimed to lay the algorithmic and analytic foundation for Monte Carlo-style Quantum Gibbs samplers. We have presented families of efficiently implementable algorithms for Gibbs sampling in terms of Lindbladians with guarantees for fixed-point accuracy. We have confronted technical issues from energy uncertainty (which have haunted quantum Gibbs sampling algorithms for a decade) by highlighting a key algorithmic component, the operator Fourier Transform, and introducing a general analytic framework: the secular approximation and approximate detailed balance. These arguments are compatible with the Szegedy-type speedup and preparation of the purified Gibbs state.

Our construction is conceptually simple as it draws inspiration from the physical mechanism of open-system thermalization, especially the Davies' generator. Conversely,

our argument completes the first proof of Gibbs state stationarity for physically derived Lindbladians, especially the coarse-grained master equations [121, 134]. Potentially, this could lead to proposals of quantum Gibbs samplers on analog quantum simulators. Still, our analysis is restricted to the open system setting where the bath is Markovian by assumption; we hope the precise statement in the open system settings inspires further insight toward closed-system thermodynamics.

Would quantum Gibbs samplers be the ultimate solution to the ground state preparation problem by setting $\beta \gg 1$? Our work only answers the first half of this problem by writing down some candidate Lindbladians with a provably accurate Gibbs fixed point and efficient simulation algorithm. Still, the missing piece of the puzzle is the mixing time (or the spectral gap) of the proposed Lindbladians. A scientifically informative first step is to directly benchmark the performance of quantum Gibbs samplers numerically for viable system sizes. This would give concrete estimates of the realistic costs of quantum simulation. From a mathematical physics perspective, there have been efforts to prove rapid mixing [16, 31] (convergence at a logarithmic depth $\log(n)$) of Quantum Gibbs samplers for lattice Hamiltonian in the spirit of classical Ising models [124]. There, most results have been restricted to commuting Hamiltonians due to the lack of a satisfactory formulation of noncommuting Gibbs samplers, which this work provides[37]. As a direct implication of this work, we provide a candidate algorithm for preparing a gapped ground state at a potentially very low depth: simply setting the Gaussian width to be $\sigma_t \sim \tilde{O}(\frac{1}{\Delta_{gap}})$ and $\beta = \tilde{O}(\frac{1}{\Delta_{gap}})$ ensures the ground state to be approximately the common kernel of $\tilde{O}(\frac{1}{\Delta_{gap}})$-local Lindbladians[38].

From a complexity perspective, quantum Gibbs samplers provide a new *dynamic* angle to study the complexity of thermal states and even ground states. Indeed, existing complexity results for gapped ground states, especially the area law, have beaten the static properties to death (local gap, decay of correlation, etc.). Conceptually, these approaches neglect the instinctive experimental origin of ground states: cool the system in a fridge. It would be curious to bridge this thermodynamics process to the area-law literature (e.g., [10, 84, 106]). Practically, we hope noncommuting Gibbs samplers will inspire new tensor network algorithms or even new ansatz, which could lead to a better grasp of 2D gapped physics.

[37] The followup work [42] gives an even nicer noncommutative Gibbs sampler with *exact detailed balance*.

[38] This observation was later exploited in [61] for ground state preparation.

Acknowledgments

We thank Simon Apers, Mario Berta, Garnet Chan, Alex Dalzell, Zhiyan Ding, Hsin-Yuan (Robert) Huang, Lin Lin, Yunchao Liu, Sam McArdle, Jonathan Moussa, Evgeny Mozgunov, Tobias Osborne, Wocjan Pawel, Patrick Rall, Mehdi Soleimanifar, Kristan Temme, Umesh Vazirani, and Tong Yu for helpful discussions. We also thank anonymous referees for their helpful feedback. AG thanks Chunhao Wang, Dávid Matolcsi, Cambyse Rouzé, and Daniel Stilck França for useful discussions. CFC is supported by the Eddlemen Fellowship and the AWS Center for Quantum Computing internship. AG acknowledges funding from the AWS Center for Quantum Computing.

2.5 Appendix:Operator Fourier Transform: properties and error bounds

In this section, we study properties of the *operator Fourier Transform*. Given a Hamiltonian $\boldsymbol{H}$, an operator $\boldsymbol{A}$, and a complex-valued function $f : \mathbb{R} \to \mathbb{C}$, let

$$\hat{\boldsymbol{A}}_f(\bar{\omega}) := \frac{1}{\sqrt{N}} \sum_{\bar{t} \in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} f(\bar{t}) \boldsymbol{A}(\bar{t}) \quad \text{where} \quad \boldsymbol{A}(t) := \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A} \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}. \tag{2.36}$$

Note the normalization $\frac{1}{\sqrt{N}}$. The transformed operators $\hat{\boldsymbol{A}}_f(\bar{\omega})$ satisfy the desirable *exact* symmetry of a standard Fourier Transform as well as an operator version of Parseval's identity.

Proposition 2.5.1 (Symmetry and operator Parseval's identity)**.** *For a set of matrices $\{\boldsymbol{A}^a\}_{a \in A}$ and a Hamiltonian $\boldsymbol{H}$, consider their discrete operator Fourier Transform weighted by a complex-valued function $f : S_{t_0} \to \mathbb{C}$ as in (2.36). Then, the symmetry holds $\hat{\boldsymbol{A}}^a_f(\bar{\omega})^\dagger = (\hat{\boldsymbol{A}^{a\dagger}})_{f^*}(-\bar{\omega})$, moreover*

$$\sum_{a \in A} \sum_{\bar{\omega} \in S_{\omega_0}} \hat{\boldsymbol{A}}^a_f(\bar{\omega})^\dagger \hat{\boldsymbol{A}}^a_f(\bar{\omega}) = \sum_{a \in A} \sum_{\bar{t} \in S_{t_0}} |f(\bar{t})|^2 \mathrm{e}^{\mathrm{i}\boldsymbol{H}\bar{t}} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a \mathrm{e}^{-\mathrm{i}\boldsymbol{H}\bar{t}} \preceq \|\sum_{a \in A} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a\| \|f\|_2^2 \cdot \boldsymbol{I} \quad \textit{and} \tag{2.37}$$

$$\sum_{a \in A} \sum_{\bar{\omega} \in S_{\omega_0}} \hat{\boldsymbol{A}}^a_f(\bar{\omega}) \hat{\boldsymbol{A}}^a_f(\bar{\omega})^\dagger = \sum_{a \in A} \sum_{\bar{t} \in S_{t_0}} |f(\bar{t})|^2 \mathrm{e}^{\mathrm{i}\boldsymbol{H}\bar{t}} \boldsymbol{A}^a \boldsymbol{A}^{a\dagger} \mathrm{e}^{-\mathrm{i}\boldsymbol{H}\bar{t}} \preceq \|\sum_{a \in A} \boldsymbol{A}^a \boldsymbol{A}^{a\dagger}\| \|f\|_2^2 \cdot \boldsymbol{I}. \tag{2.38}$$

For our Lindbladian Gibbs samplers, the weights will be normalized $\sum_{\bar{t} \in S_{t_0}} |f(\bar{t})|^2 = 1$, which means that they can be implemented by amplitudes of a state. In the special case $\sum_{a \in A} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a = \boldsymbol{I}$ (i.e., these operators can be interpreted as a quantum channel), then the inequality (2.37) hold with equality, and as a consequence, the operators resolve the identity $\sum_{a \in A} \sum_{\bar{\omega} \in S_{\omega_0}} \hat{\boldsymbol{A}}^a_f(\bar{\omega})^\dagger \hat{\boldsymbol{A}}^a_f(\bar{\omega}) = \boldsymbol{I}$.

Proof. First observe that by definition

$$\hat{\boldsymbol{A}}_f^a(\bar{\omega})^\dagger = \frac{1}{\sqrt{N}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{\mathrm{i}\bar{\omega}\bar{t}} f^*(\bar{t})(\boldsymbol{A}^a(\bar{t}))^\dagger \tag{2.39}$$
$$= \frac{1}{\sqrt{N}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{\mathrm{i}\bar{\omega}\bar{t}} f^*(\bar{t})\boldsymbol{A}^{a\dagger}(\bar{t}) = (\hat{\boldsymbol{A}^{a\dagger}})_{f^*,-\bar{\omega}}.$$

Next, we prove (2.37) by direct computation as follows

$$\sum_{a\in A}\sum_{\bar{\omega}\in S_{\omega_0}} \hat{\boldsymbol{A}}_f^a(\bar{\omega})^\dagger \hat{\boldsymbol{A}}_f^a(\bar{\omega}) = \sum_{a\in A}\frac{1}{N}\sum_{\bar{\omega}\in S_{\omega_0}}\sum_{\bar{t}'\in S_{t_0}} \mathrm{e}^{\mathrm{i}\bar{\omega}\bar{t}'} f^*(\bar{t})\boldsymbol{A}^a(\bar{t}')^\dagger \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} f(\bar{t})\boldsymbol{A}^a(\bar{t})$$

$$= \sum_{a\in A}\sum_{\bar{t}\in S_{t_0}} |f(\bar{t})|^2 \boldsymbol{A}^a(\bar{t})^\dagger \boldsymbol{A}^a(\bar{t})$$
$$= \sum_{a\in A}\sum_{\bar{t}\in S_{t_0}} |f(\bar{t})|^2 \mathrm{e}^{\mathrm{i}\boldsymbol{H}\bar{t}} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a \mathrm{e}^{-\mathrm{i}\boldsymbol{H}\bar{t}}$$
$$\preceq \|\sum_{a\in A} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \sum_{\bar{t}\in S_{t_0}} |f(\bar{t})|^2 \mathrm{e}^{\mathrm{i}\boldsymbol{H}\bar{t}}\boldsymbol{I}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}\bar{t}}$$
$$= \boldsymbol{I}\cdot\|\sum_{a\in A} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \sum_{\bar{t}\in S_{t_0}} |f(\bar{t})|^2.$$

The second equality uses the Fourier representation of the discrete delta function

$$\sum_{\bar{\omega}\in S_{\omega_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}(\bar{t}-\bar{t}')} = N\delta_{\bar{t},\bar{t}'}. \tag{2.40}$$

The proof of (2.38) is completely analogous. ∎

We also include the analogous analysis in the continuum limit where the discretization parameter N disappears. We will assume throughout that the weight function f is square integrable, i.e., $f \in \ell_2(\mathbb{R})$. In the continuous case, the operator Fourier Transform is a matrix-valued function, and to emphasize this, we change the notation to $\hat{\boldsymbol{A}}_f(\bar{\omega}) \to \hat{\boldsymbol{A}}_f(\omega)$. We could directly copy the above proof; however, arguing about the Dirac delta function in the continuous case is tricky. We resolve this by relying on Parseval-Plancherel's identity.

For studying the operator Fourier Transform, it is useful to decompose the operator according to the Bohr frequencies

$$\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} = \sum_{\nu\in B(\boldsymbol{H})} \mathrm{e}^{\mathrm{i}\nu t}\boldsymbol{A}_\nu, \tag{2.41}$$
$$\text{where}\quad \boldsymbol{A}_\nu := \sum_{E_2-E_1=\nu} \boldsymbol{P}_{E_2}\boldsymbol{A}\boldsymbol{P}_{E_1} \quad \text{satisfies that} \quad (\boldsymbol{A}_\nu)^\dagger = (\boldsymbol{A}^\dagger)_{-\nu},$$

where $\boldsymbol{P}_E$ denotes the orthogonal projector onto the subspace spanned by energy E eigenstates of $\boldsymbol{H}$. If $f \in \ell_1(\mathbb{R})$, using this decomposition, we can conveniently express the operator Fourier Transform as follows

$$\begin{aligned}\hat{\boldsymbol{A}}_f(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} f(t) \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A} \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} \mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} f(t) \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \sum_{\nu \in B(\boldsymbol{H})} \boldsymbol{A}_\nu \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} \mathrm{d}t \\ &= \sum_{\nu \in B(\boldsymbol{H})} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}(\omega-\nu)t} f(t) \boldsymbol{A}_\nu \mathrm{d}t = \sum_{\nu \in B(\boldsymbol{H})} \hat{f}(\omega - \nu) \boldsymbol{A}_\nu, \quad (2.42)\end{aligned}$$

where $\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t$ is the Fourier Transform of the weight function $f(t)$. More generally, if $f \in \ell_2(\mathbb{R})$, then we use (2.42) as the definition of the operator Fourier Transform because the Fourier Transform uniquely extends to a unitary map $\mathcal{F} \colon \ell_2(\mathbb{R}) \to \ell_2(\mathbb{R})$.

Proposition 2.5.2 (Symmetry and operator Parseval's identity)**.** *For a set of matrices $\{\boldsymbol{A}^a\}_{a \in A}$ and a Hamiltonian $\boldsymbol{H}$, consider their continuous operator Fourier Transform weighted by a complex-valued function $f \in \ell_2(\mathbb{R})$*

$$\hat{\boldsymbol{A}}^a_f(\omega) := \sum_{\nu \in B(\boldsymbol{H})} \hat{f}(\omega - \nu) \boldsymbol{A}^a_\nu,$$

then $(\hat{\boldsymbol{A}}^a_f(\omega))^\dagger = (\hat{\boldsymbol{A}^{a\dagger}})_{f^*}(-\omega)$, moreover

$$\sum_{a \in A} \int_{-\infty}^{\infty} \hat{\boldsymbol{A}}^a_f(\omega)^\dagger \hat{\boldsymbol{A}}^a_f(\omega) \mathrm{d}\omega = \sum_{a \in A} \int_{-\infty}^{\infty} |f(t)|^2 \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} \mathrm{d}t \preceq \|f\|_2^2 \Big\| \sum_{a \in A} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a \Big\| \cdot \boldsymbol{I}, \quad \textit{and} \tag{2.43}$$

$$\sum_{a \in A} \int_{-\infty}^{\infty} \hat{\boldsymbol{A}}^a_f(\omega) \hat{\boldsymbol{A}}^a{}_f(\omega)^\dagger \mathrm{d}\omega = \sum_{a \in A} \int_{-\infty}^{\infty} |f(t)|^2 \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^a \boldsymbol{A}^{a\dagger} \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} \mathrm{d}t \preceq \|f\|_2^2 \Big\| \sum_{a \in A} \boldsymbol{A}^a \boldsymbol{A}^{a\dagger} \Big\| \cdot \boldsymbol{I}. \tag{2.44}$$

Similarly as before, if $\sum_{a \in A} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a = \boldsymbol{I}$, then the inequality (2.43) hold with equality.

Proof. First observe that by definition

$$\begin{aligned}\hat{\boldsymbol{A}}^a_f(\omega)^\dagger &= \sum_{\nu \in B(\boldsymbol{H})} (\hat{f}(\omega - \nu))^* (\boldsymbol{A}^a_\nu)^\dagger \\ &= \sum_{\nu \in B(\boldsymbol{H})} \widehat{f^*}(\nu - \omega) (\boldsymbol{A}^{a\dagger})_{-E} \\ &= \sum_{-\nu \in B(\boldsymbol{H})} \widehat{f^*}(-\omega - \nu) (\boldsymbol{A}^{a\dagger})_\nu = \hat{\boldsymbol{A}^{a\dagger}}_{f^*}(-\omega).\end{aligned}$$

Next, we prove (2.43) by direct computation as follows:

$$\sum_{a\in A}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}_f^a(\omega)^\dagger\hat{\boldsymbol{A}}_f^a(\omega)\mathrm{d}\omega = \sum_{a\in A}\int_{-\infty}^{\infty}\sum_{\nu\in B(\boldsymbol{H})}(\boldsymbol{A}_\nu^a)^\dagger(\hat{f}(\omega-\nu))^*\sum_{\nu'\in B(\boldsymbol{H})}\hat{f}(\omega-\nu')\boldsymbol{A}_{\nu'}^a\mathrm{d}\omega$$

(by definition)

$$= \sum_{a\in A}\sum_{\nu,\nu'\in B(\boldsymbol{H})}(\boldsymbol{A}_\nu^a)^\dagger\boldsymbol{A}_{\nu'}^a\int_{-\infty}^{\infty}(\hat{f}(\omega))^*\hat{f}(\omega-(\nu'-\nu))\mathrm{d}\omega$$

(shift $\omega\longrightarrow\omega-\nu$ and use $|B|\leq\infty$)

$$= \sum_{a\in A}\sum_{\nu''\in B(\boldsymbol{H})}(\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a)_{\nu''}\int_{-\infty}^{\infty}(\hat{f}(\omega))^*\hat{f}(\omega-\nu'')\mathrm{d}\omega$$

(by the definition of $\boldsymbol{A}_\nu^a$)

$$= \sum_{a\in A}\sum_{\nu''\in B(\boldsymbol{H})}(\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a)_{\nu''}\int_{-\infty}^{\infty}(f(t))^*f(t)\mathrm{e}^{\mathrm{i}\nu''t}\mathrm{d}t$$

(since $\mathcal{F}$ is unitary)

$$= \sum_{a\in A}\sum_{\nu''\in B(\boldsymbol{H})}\int_{-\infty}^{\infty}|f(t)|^2\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}(\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a)_{\nu''}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}\mathrm{d}t$$

(as in (2.42))

$$= \sum_{a\in A}\int_{-\infty}^{\infty}|f(t)|^2\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}\mathrm{d}t$$

(since $|B(\boldsymbol{H})|\leq\infty$)

$$\preceq \|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\int_{-\infty}^{\infty}|f(t)|^2\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{I}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}\mathrm{d}t$$

(since $\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\preceq\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\boldsymbol{I}$)

$$= \boldsymbol{I}\cdot\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\int_{-\infty}^{\infty}|f(t)|^2\mathrm{d}t.$$

The proof of (2.44) is completely analogous.[39] ∎

2.6 Appendix:Nonasymptotic secular approximation

In this section, we define the secular approximation of the Fourier Transformed operators $\hat{\boldsymbol{A}}_f(\bar{\omega})$ and analyze the resulting error. The secular approximation applies truncation to the Fourier-transformed operators in the frequency domain by suppressing Bohr frequencies $\nu\in B(\boldsymbol{H})$ that deviate substantially from the frequency label ω via some filter function $s\in\ell_\infty(\mathbb{R})$. For example truncation at

[39] Intuitively speaking the fourth line can be viewed as consequence of the Fourier representation of the Dirac delta distribution $\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\omega t}\mathrm{d}\omega = 2\pi\delta(t)$, analogous to (2.40). Not introducing delta functions makes the proof completely general.

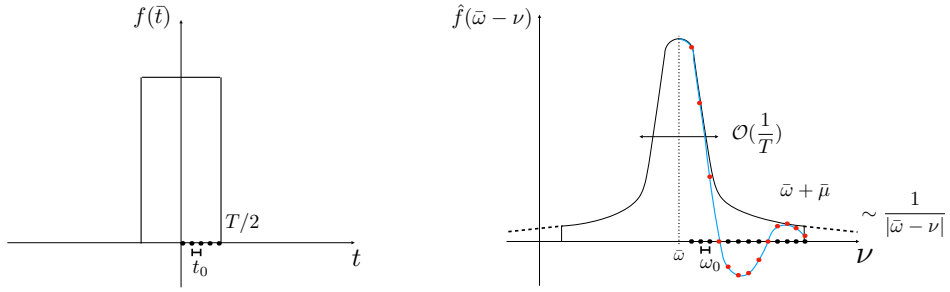


Figure 2.8: Left: the weight as the step function. Right: An illustration for the Fourier Transformed amplitudes $\boldsymbol{A}_{\bar{\omega}} = \sum_{\nu\in S_{\omega_0}} \boldsymbol{A}_\nu f(\bar{\omega}-\nu) = \sum_{\nu\in S_{\omega_0}} \boldsymbol{A}_\nu \frac{e^{i(\nu-\bar{\omega})T}-e^{-i(\nu-\bar{\omega})T}}{e^{i(\nu-\bar{\omega})t_0}-1}$ for Bohr frequency ν given the energy label $\bar{\omega}$. The expression coincides with the phase estimation profile. It peaks near energy $\nu=\bar{\omega}$ with a width $\sim T^{-1}$ and decays polynomially. The profile in absolute value oscillates (blue), but we also display the norm bound to guide the eye (black). The secular approximation truncates the profile at an energy $\bar{\mu}$ far in the tail $\bar{\mu} \gg T^{-1}$.

energy difference μ can be achieved by setting $s(\omega) := \mathbb{1}(|\omega| < \mu)$ and defining the following secular-approximated operators as follows:

$$\hat{\boldsymbol{S}}_{f,s}(\omega) := \sum_{\nu\in B(\boldsymbol{H})} \hat{f}(\omega-\nu)s(\omega-\nu)\boldsymbol{A}_\nu. \tag{2.45}$$

We will sometimes consider alternative filter functions $s\colon \mathbb{R}\to\mathbb{C}$, so we will treat general s throughout our discussion.

The key observation in our analysis is that due to the definition of the operator Fourier Transform (2.42) we have

$$\hat{\boldsymbol{S}}_{f,s}(\omega) = \hat{\boldsymbol{A}}_{f_s}(\omega),$$

where $f_s = \mathcal{F}^{-1}(\hat{f}\cdot s)$ is the inverse Fourier Transform of the function $\hat{f}(\omega)s(\omega)$. Furthermore, this also implies that

$$\hat{\boldsymbol{A}}_f(\omega) - \hat{\boldsymbol{S}}_{f,s}(\omega) = \hat{\boldsymbol{A}}_{f-f_s}(\omega),$$

where $f - f_s = f_{1-s}$ is the inverse Fourier Transform of the function $\hat{f}(\omega)(1-s(\omega))$. The significance of this equation is that it enables us to bound the error induced by the secular approximation via bounding the tail $\|\hat{f}(\omega)(1-s(\omega))\|_2$.

Proposition 2.6.1. *If $f(t)$ is real and $s(t)$ is real and even, then $f_s(t)$ is real.*

Proof. If f is real, then $\hat{f}$ has an even real and odd imaginary part, which remains true for $\hat{f} \cdot s$, and therefore $\mathcal{F}^{-1}(\hat{f} \cdot s)$ is real as well.[40] ∎

Analogously, we define the secular approximation for the discrete Fourier Transform as

$$\hat{\boldsymbol{S}}_{f,s}(\bar{\omega}) := \sum_{\nu \in B(\boldsymbol{H})} \bar{\mathcal{F}}\Big(f(\bar{t}) \cdot \mathrm{e}^{\mathrm{i}\nu\bar{t}}\Big)(\bar{\omega}) \cdot s(\bar{\omega} - \nu)\boldsymbol{A}_\nu,$$

where $\bar{\mathcal{F}}$ denotes the discrete Fourier Transform. In case ν is an integer multiple of the base frequency ω_0 of $\bar{\mathcal{F}}$, then $\bar{\mathcal{F}}\Big(f(\bar{t}) \cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\Big)(\bar{\omega})$ above simplifies to $\hat{f}(\bar{\omega} - \nu) = \bar{\mathcal{F}}(f)(\bar{\omega} - \nu)$. Indeed, the discrete Fourier Transform $\hat{f}$ is only defined at points $\bar{\omega} \in S_{\omega_0}$, and thus translation by some value ν which is not an integer multiple of ω_0 can cause troubles.

The simple analysis working nicely in the continuous case can be directly translated to the discrete case if all the Bohr frequencies are multiples of the base frequency ω_0 of $\bar{\mathcal{F}}$, i.e., $B \subset \omega_0\mathbb{Z}$. This is the reason why we introduce a slightly rounded Hamiltonian in the analysis of the secular approximation.

Proposition 2.6.2 (Truncation via modifying weight function)**.** *Let $f\colon S_{t_0} \to \mathbb{C}$ and $s\colon S_{\omega_0} \to \mathbb{C}$ and suppose that $B(\boldsymbol{H}) \subset \omega_0\mathbb{Z}$, then the secular-approximated operator can be represented as plain operator Fourier Transform corresponding to a perturbed weight function*

$$\hat{\boldsymbol{S}}_{f,s}(\bar{\omega}) = \hat{\boldsymbol{A}}_{f_s}(\bar{\omega}),$$

where $f_s(\bar{\omega})$ is the inverse discrete Fourier Transform of the function $\hat{f}(\bar{\omega})s(\bar{\omega})$ and $\hat{f}(\bar{\omega}) := \frac{1}{\sqrt{N}} \sum_{\bar{t} \in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} f(\bar{t})$.

[40] If we work with the discrete Fourier Transform, and N is even, then $-\omega_0 \cdot N/2$ (or $-t_0 \cdot N/2$ if we work in the time domain) should be treated as its own inverse due to reasons of parity and modular arithmetic. In particular an even function can take arbitrary value on $-\omega_0 \cdot N/2$, but an odd function must be 0, similarly to how such functions must behave on 0.

Proof.

$$\begin{aligned}\hat{\boldsymbol{S}}_{f,s}(\bar{\omega}) &= \sum_{\bar{\nu}\in B(\boldsymbol{H})} \hat{f}(\bar{\omega}-\bar{\nu})\cdot s(\bar{\omega}-\bar{\nu})\boldsymbol{A}_{\bar{\nu}} \\ &= \sum_{\bar{\nu}\in B(\boldsymbol{H})} \hat{f}_s(\bar{\omega}-\bar{\nu})\boldsymbol{A}_{\bar{\nu}} \\ &= \sum_{\bar{\nu}\in B(\boldsymbol{H})} \frac{1}{\sqrt{N}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}(\bar{\omega}-\nu)\bar{t}} f_s(\bar{t})\boldsymbol{A}_{\bar{\nu}} \\ &= \frac{1}{\sqrt{N}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} f_s(\bar{t})\boldsymbol{A}(\bar{t}).\end{aligned}$$

The last equality recombines the Bohr frequencies $\sum_{\bar{\nu}\in B(\boldsymbol{H})} \mathrm{e}^{\mathrm{i}\bar{\nu}\bar{t}}\boldsymbol{A}_{\bar{\nu}} = \boldsymbol{A}(\bar{t})$ analogously to (2.42). ∎

To reiterate, the condition $B \subseteq \omega_0\mathbb{Z}$ need not hold for the original Hamiltonian $\boldsymbol{H}$. Proceeding with the discretized Hamiltonian $\bar{\boldsymbol{H}}$ introduces a small additive error.[41] We present error bounds for both the Lindbladian and our discriminant proxy.

Lemma 2.6.1 (Perturbation bounds)**.** *Let $f, f' \in \mathbb{C}^{S_{t_0}}$, and $\gamma \in \mathbb{C}^{S_{\omega_0}}$ such that $\|f\|_2, \|f'\|_2, \|\gamma\|_\infty \leq 1$. If $\|\sum_{a\in A} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \leq 1$, then for every $T > 0$ and $f_T(t) := f(t)\mathbb{1}(|t| \leq T)$ we have that*

$$\begin{aligned}&\Big\| \sum_{\bar{\omega}\in S_{\omega_0}, a\in A} \sqrt{\gamma(\bar{\omega})}|\bar{\omega}, a\rangle\langle\bar{0}| \otimes \hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega}) - \sqrt{\gamma(\bar{\omega})}|\bar{\omega}, a\rangle\langle\bar{0}| \otimes \hat{\boldsymbol{A}}^a_{(f',\boldsymbol{H}')}(\bar{\omega})\Big\| \\ &\leq \|f - f_T\|_2 + \|f_T - f'\|_2 + 2T\|\boldsymbol{H} - \boldsymbol{H}'\|.\end{aligned}$$

Proof. This directly follows from the (not necessarily unitary) block-encoding construction of Figure 2.5-Figure 2.4 and triangle inequalities. Indeed, for any function $g \in \mathbb{C}^{S_{t_0}}$ let $\boldsymbol{Prep}_g := |g\rangle\langle\bar{0}|$ and let $\boldsymbol{B}_{(g,\boldsymbol{H})}$ denote the block-encoding given by Figure 2.4 when setting $\boldsymbol{Prep} \leftarrow \boldsymbol{Prep}_g$, $\boldsymbol{V}_{jp} \leftarrow \sum_a |a\rangle \otimes \boldsymbol{A}^a$, and using the Hamiltonian $\boldsymbol{H}$. Then, we have that

$$\begin{aligned}\|\boldsymbol{B}_{(g,\boldsymbol{H})} - \boldsymbol{B}_{(f_T,\boldsymbol{H}')}\| &\leq \|\boldsymbol{B}_{(g,\boldsymbol{H})} - \boldsymbol{B}_{(f_T,\boldsymbol{H})}\| + \|\boldsymbol{B}_{(f_T,\boldsymbol{H})} - \boldsymbol{B}_{(f_T,\boldsymbol{H}')}\| \\ &\leq \|g - f_T\|_2\|\boldsymbol{V}_{jp}\| + 2T\|\boldsymbol{H} - \boldsymbol{H}'\|\|\boldsymbol{V}_{jp}\| \\ &\leq \|g - f_T\|_2 + 2T\|\boldsymbol{H} - \boldsymbol{H}'\|.\end{aligned}$$

[41] This differs from the *unphysical* rounding assumption [180] where the Hamiltonian needs to have "large" gaps in the spectrum. Here, ω_0 is not related to the resolution of energy estimates but rather the discretization of the register.

Using the above inequality twice, we obtain the desired result

$$\begin{aligned}
&\| \sum_{\bar{\omega}\in S_{\omega_0}, a\in A} \sqrt{\gamma(\bar{\omega})}|\bar{\omega}, a\rangle\langle\bar{0}| \otimes \left(\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega}) - \hat{\boldsymbol{A}}^a_{(f',\boldsymbol{H}')}(\bar{\omega})\right)\| \\
&= \|\boldsymbol{B}_{(f,\boldsymbol{H})} - \boldsymbol{B}_{(f',\boldsymbol{H}')}\| \\
&\leq \|\boldsymbol{B}_{(f,\boldsymbol{H})} - \boldsymbol{B}_{(f_T,\boldsymbol{H}')}\| + \|\boldsymbol{B}_{(f_T,\boldsymbol{H}')} - \boldsymbol{B}_{(f',\boldsymbol{H}')}\| \\
&\leq \|f - f_T\|_2 + \|f_T - f'\|_2 + 2T\|\boldsymbol{H} - \boldsymbol{H}'\|. \qquad \blacksquare
\end{aligned}$$

Corollary 2.6.1 (Perturbation bounds on Lindbladians and discriminant proxies)**.** *Consider*

$$\bar{\mathcal{L}}_{(f,\boldsymbol{H})} := \sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \gamma(\bar{\omega})\left(\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega})[\cdot]\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega})^\dagger - \frac{1}{2}\{\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega})^\dagger\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega}), \cdot\}\right), \quad \textit{and}$$

$$\begin{aligned}
\bar{\boldsymbol{\mathcal{D}}}_{(f,\boldsymbol{H})} := &\sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})\gamma(-\bar{\omega})}\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega}) \otimes \hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega})^* \\
&- \frac{\gamma(\bar{\omega})}{2}\left(\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega})^\dagger\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega}) \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega})^{*\dagger}\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\bar{\omega})^*\right).
\end{aligned}$$

Assuming the conditions and the notation of Lemma 2.6.1 hold, we have

$$\|\bar{\mathcal{L}}_{(f,\boldsymbol{H})} - \bar{\mathcal{L}}_{(f',\boldsymbol{H}')}\|_{1-1} \leq 4(\|f - f_T\|_2 + \|f_T - f'\|_2 + 2T\|\boldsymbol{H} - \boldsymbol{H}'\|).$$

Further assuming the symmetry and normalization conditions (2.10),(2.11) are satisfied by f, f' and $\{\boldsymbol{A}^a : a \in A\}$, we have

$$\|\bar{\boldsymbol{\mathcal{D}}}_{(f,\boldsymbol{H})} - \bar{\boldsymbol{\mathcal{D}}}_{(f',\boldsymbol{H}')}\| \leq 4(\|f - f_T\|_2 + \|f_T - f'\|_2 + 2T\|\boldsymbol{H} - \boldsymbol{H}'\|).$$

Proof. The superoperator $\mathcal{L}_{(f,\boldsymbol{H})}$ acts as follows:

$$\begin{aligned}
&\bar{\mathcal{L}}_{(f,\boldsymbol{H})}[\boldsymbol{\rho}] \\
&= \mathrm{Tr}_{a,\bar{\omega}}\left[\left(\sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})}|\bar{\omega}, a\rangle\langle\bar{0}|\hat{\boldsymbol{A}}^a(\bar{\omega})\right)|\bar{0}\rangle\langle\bar{0}| \otimes \boldsymbol{\rho}\left(\sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})}|\bar{0}\rangle\langle\bar{\omega}, a|\hat{\boldsymbol{A}}^a(\bar{\omega})^\dagger\right)\right] \\
&- \frac{1}{2}\mathrm{Tr}_{\bar{0}}\left\{\left(\sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})}|\bar{0}\rangle\langle\bar{\omega}, a|\hat{\boldsymbol{A}}^a(\bar{\omega})^\dagger\right)\left(\sum_{a\in A, \bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})}|\bar{\omega}, a\rangle\langle\bar{0}|\hat{\boldsymbol{A}}^a(\bar{\omega})\right), |\bar{0}\rangle\langle\bar{0}| \otimes \boldsymbol{\rho}\right\}.
\end{aligned}$$

The conclusion about $\|\bar{\mathcal{L}}_{(f,\boldsymbol{H})} - \bar{\mathcal{L}}_{(f',\boldsymbol{H}')}\|_{1-1}$ follows from Lemma 2.6.1 using the triangle and Hölder inequalities and that taking partial trace contracts trace-distance.

The proof of Proposition 2.3.5 shows that if $\boldsymbol{B}$ is a (nonunitary) block-encoding of $\bar{\mathcal{L}}_{(f,\boldsymbol{H})}$, then $\boldsymbol{B}'^{\dagger}\cdot(\boldsymbol{R}-\boldsymbol{I})\cdot\boldsymbol{B}'$ is a block-encoding of $\bar{\mathcal{D}}_{(f,\boldsymbol{H})}$. Since $\|f\|_2, \|f'\|_2, \|\gamma\|_\infty \leq 1$ without loss of generality we can assume $\|\boldsymbol{B}'\| \leq 1$, which together with $\|\boldsymbol{R}-\boldsymbol{I}\| \leq 2$ implies the bound on $\|\bar{\mathcal{D}}_{(f,\boldsymbol{H})} - \bar{\mathcal{D}}_{(f',\boldsymbol{H}')}\|$ via a triangle inequality. ∎

Note that under the conditions of Theorem 2.8.1, the same bounds also hold in the continuous case, as can be shown by a limit argument using the results of Theorem 2.8.1.

Corollary 2.6.2 (Perturbation bounds on continuous Lindbladians and discriminant proxies)**.** *Consider*

$$\mathcal{L}_{(f,\boldsymbol{H})} := \sum_{a\in A}\int_{-\infty}^{\infty}\gamma(\omega)\left(\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)[\cdot]\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)^\dagger - \frac{1}{2}\{\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)^\dagger\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega),\cdot\}\right)\mathrm{d}\omega, \quad \textit{and}$$

$$\begin{aligned}\mathcal{D}_{(f,\boldsymbol{H})} := &\sum_{a\in A}\int_{-\infty}^{\infty}\sqrt{\gamma(\omega)\gamma(-\omega)}\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)\otimes\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)^* \\ &- \frac{\gamma(\omega)}{2}\left(\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)^\dagger\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)\otimes\boldsymbol{I} + \boldsymbol{I}\otimes\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)^{*\dagger}\hat{\boldsymbol{A}}^a_{(f,\boldsymbol{H})}(\omega)^*\right)\mathrm{d}\omega.\end{aligned}$$

If $\gamma \in \ell_\infty(\mathbb{R})$, $f \in \ell_2(\mathbb{R})$, and γ, f are continuous almost everywhere (i.e., the set of points of discontinuity has measure zero) while f is bounded on every finite interval, then assuming the conditions and the notation of Lemma 2.6.1 hold, we have

$$\left\|\mathcal{L}_{(f,\boldsymbol{H})} - \mathcal{L}_{(f',\boldsymbol{H}')}\right\|_{1-1} \leq 4(\|f-f_T\|_2 + \|f_T - f'\|_2 + 2T\|\boldsymbol{H}-\boldsymbol{H}'\|).$$

Further assuming the symmetry and normalization conditions (2.10),(2.11) are satisfied by f, f' and $\{\boldsymbol{A}^a : a\in A\}$, we have

$$\|\mathcal{D}_{(f,\boldsymbol{H})} - \mathcal{D}_{(f',\boldsymbol{H}')}\| \leq 4(\|f-f_T\|_2 + \|f_T - f'\|_2 + 2T\|\boldsymbol{H}-\boldsymbol{H}'\|).$$

Proof. The objects $\mathcal{L}_{(f,\boldsymbol{H})}, \mathcal{L}_{(f',\boldsymbol{H}')}, \mathcal{D}_{(f,\boldsymbol{H})}, \mathcal{D}_{(f,\boldsymbol{H})}$ can be obtained as limits of their respective discretizations $\bar{\mathcal{L}}_{(f,\boldsymbol{H})}, \bar{\mathcal{L}}_{(f',\boldsymbol{H}')}, \bar{\mathcal{D}}_{(f,\boldsymbol{H})}, \bar{\mathcal{D}}_{(f,\boldsymbol{H})}$ as per Theorem 2.8.1, for which the discretized versions of these bounds hold due to Corollary 2.6.2. As shown in the proof of Theorem 2.8.1, if a function $g \in \ell_2(\mathbb{R})$ is continuous almost everywhere while also bounded on every finite interval, then $\|g\|_2^2 = \lim_{K\to\infty}\lim_{N\to\infty}\sum_{\bar{t}\in S_{t_0}^{\lceil N\rceil}}|\bar{g}_K(\bar{t})|^2$, where $t_0 = \sqrt{2\pi/N}$ and $\bar{g}_K(t) = \sqrt{t_0}g(t)\mathbb{1}(|t|\leq K)$, therefore the RHS of the discretized bounds also converge to their continuous counterpart implying the validity of the continuous versions of these bounds. ∎

Lemma 2.6.2 (Secular approximation)**.** *Let* $\|\gamma\|_\infty \leq 1$*, and consider the Lindbladian* $\mathcal{L}_\beta$ (2.21) *and discriminant* $\mathcal{D}_\beta$ (2.26) *with* $\hat{\boldsymbol{A}}^a(\bar{\omega})$ *being the operator Fourier Transforms of* $\boldsymbol{A}^a$ *with* $\boldsymbol{H}$ *and their secular approximations*

$$\begin{aligned}\mathcal{L}_{sec}[\cdot] &= \sum_{a\in A,\bar{\omega}\in S_{\omega_0}} \gamma(\bar{\omega})\left(\hat{\boldsymbol{S}}^a(\bar{\omega})[\cdot]\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger - \frac{1}{2}\{\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger\hat{\boldsymbol{S}}^a(\bar{\omega}),\cdot\}\right),\\ \mathcal{D}_{sec} &= \sum_{a\in A,\bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})\gamma(-\bar{\omega})}\hat{\boldsymbol{S}}^a(\bar{\omega})\otimes\hat{\boldsymbol{S}}^a(\bar{\omega})^*\\ &\quad - \frac{\gamma(\bar{\omega})}{2}\left(\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger\hat{\boldsymbol{S}}^a(\bar{\omega})\otimes\boldsymbol{I} + \boldsymbol{I}\otimes\hat{\boldsymbol{S}}^a(\bar{\omega})^{\dagger *}\hat{\boldsymbol{S}}^a(\bar{\omega})^*\right),\end{aligned}$$

with the operators

$$\hat{\boldsymbol{S}}^a(\bar{\omega}) := \sum_{\bar{\nu}\in B(\bar{\boldsymbol{H}})} \hat{f}_s(\bar{\omega}-\bar{\nu})\boldsymbol{A}^a_{\bar{\nu}} \quad \textit{where} \quad \hat{f}_s(\bar{\omega}) := \hat{f}(\bar{\omega})\cdot\mathbb{1}(|\bar{\omega}| < m\omega_0)$$

defined by the discretized Hamiltonian $\bar{\boldsymbol{H}}$ *and cut-off frequency* $m\omega_0$*. If* $\|f\|_2 \leq 1$ *and* $\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \leq 1$*, then for every* $T > 0$ *and* $f_T(t) := f(t)\mathbb{1}(|t| \leq T)$*, we have*

$$\left\|\mathcal{L}_\beta - \mathcal{L}_{sec}\right\|_{1-1} \leq 4\|\hat{f} - \hat{f}_s\|_2 + 8\|f - f_T\|_2 + 4T\omega_0.$$

Moreover, assuming the symmetry and normalization conditions (2.10),(2.11)*, we have* $\mathcal{D}_{sec} = \mathcal{D}_{sec}^\dagger$ *and*

$$\|\mathcal{D}_\beta - \mathcal{D}_{sec}\| \leq 4\|\hat{f} - \hat{f}_s\|_2 + 8\|f - f_T\|_2 + 4T\omega_0.$$

The truncation introduces an error scaling with the tail in the frequency domain, while the last two error terms arise from discretizing the Hamiltonian spectrum $\boldsymbol{H} \to \bar{\boldsymbol{H}}$ for discrete Fourier Transforms; this is more of a technical artifact and merely introduces a minor error shrinking with finer Fourier frequency resolution ω_0.

Proof. Since the secular approximation amounts to changing the real function and discretizing the Hamiltonian, i.e., $\mathcal{L}_{sec} = \mathcal{L}_{(f_s,\bar{\boldsymbol{H}})}$ we can apply Corollary 2.6.1. The final bound follows using the observation that $\|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \leq \omega_0/2$, and $\|f - f_s\| = \|\hat{f} - \hat{f}_s\|$ since the discrete Fourier Transformation is unitary.

Finally, since $\mathcal{D}_{sec} = \mathcal{D}_{(f_s,\bar{\boldsymbol{H}})}$ we have that $\mathcal{D}_{sec}$ is self-adjoint due to Proposition 2.6.1 and Corollary 2.3.2. ∎

Taking the continuum limit further simplifies the expression.

Lemma 2.6.3 (Secular approximation in the continuum)**.** *Let* $\|\gamma\|_\infty \leq 1$*, and consider the Lindbladian* $\mathcal{L}_\beta$ (2.21) *and discriminant* $\mathcal{D}_\beta$ (2.26) *with* $\hat{\boldsymbol{A}}^a(\omega)$ *being the operator Fourier Transforms of* $\boldsymbol{A}^a$ *with* $\boldsymbol{H}$ *and their secular approximations*

$$\begin{aligned}
\mathcal{L}_{sec}[\cdot] &= \sum_{a\in A}\int_{-\infty}^{\infty}\gamma(\omega)\left(\hat{\boldsymbol{S}}^a(\omega)[\cdot]\hat{\boldsymbol{S}}^a(\omega)^\dagger - \frac{1}{2}\{\hat{\boldsymbol{S}}^a(\omega)^\dagger\hat{\boldsymbol{S}}^a(\omega),\cdot\}\right)\mathrm{d}\omega,\\
\mathcal{D}_{sec} &= \sum_{a\in A}\int_{-\infty}^{\infty}\sqrt{\gamma(\omega)\gamma(-\omega)}\hat{\boldsymbol{S}}^a(\omega)\otimes\hat{\boldsymbol{S}}^a(\omega)^*\\
&\quad - \frac{\gamma(\omega)}{2}\left(\hat{\boldsymbol{S}}^a(\omega)^\dagger\hat{\boldsymbol{S}}^a(\omega)\otimes\boldsymbol{I} + \boldsymbol{I}\otimes\hat{\boldsymbol{S}}^a(\omega)^{\dagger *}\hat{\boldsymbol{S}}^a(\omega)^*\right)\mathrm{d}\omega,
\end{aligned}$$

with the operators

$$\hat{\boldsymbol{S}}^a(\omega) := \sum_{\nu\in B(\boldsymbol{H})}\hat{f}_s(\omega-\nu)\boldsymbol{A}^a_\nu \quad \textit{where} \quad \hat{f}_s(\omega) := \hat{f}(\omega)\cdot\mathbb{1}(|\omega|<\mu).$$

If $\|f\|_2 \leq 1$ *and* $\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \leq 1$*, then,*

$$\left\|\mathcal{L}_\beta - \mathcal{L}_{sec}\right\|_{1-1} \leq 4\|\hat{f}-\hat{f}_s\|_2$$

Moreover, assuming the symmetry and normalization conditions (2.10),(2.11)*, we have* $\mathcal{D}_{sec} = \mathcal{D}_{sec}^\dagger$ *and*

$$\|\mathcal{D}_\beta - \mathcal{D}_{sec}\| \leq 4\|\hat{f}-\hat{f}_s\|_2.$$

Let us quickly note that the above argument controls the implementation error for truncating the Hamiltonian simulation. Indeed, in practice, we will only implement Hamiltonian simulation up to time T, and this is perfectly accounted for by Lemma 2.6.1,Corollary 2.6.1 by setting $f' = f_T$ and $\boldsymbol{H}' = \boldsymbol{H}$.

Uniform weights

Consider the simplest Fourier Transform with uniform weights

$$\hat{\boldsymbol{A}}_f(\bar{\omega}) := \frac{1}{\sqrt{2NT/t_0}}\sum_{-T\leq\bar{t}<T}\mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}}\boldsymbol{A}(\bar{t}).$$

Proposition 2.6.3 (Preparing uniform weights)**.** *Suppose that* $T/t_0 = 2^k$*. Then, the state*

$$\sum_{\bar{t}} f(\bar{t})|\bar{t}\rangle \quad \textit{for} \quad f(\bar{t}) = \frac{1}{\sqrt{2T/t_0}}\cdot\begin{cases}1 & \textit{if} \quad -T\leq\bar{t}<T\\ 0 & \textit{else.}\end{cases}$$

can be prepared using $k+1$ *Hadamard gates and* $n-k-1$ *CNOT gates.*

Proof. Prepare with the GHZ state on the first $n-k$ qubits using 1 Hadamard gate and $n-k-1$ CNOT gates

$$\frac{1}{\sqrt{2}}\left(|1^{n-k}\rangle + |0^{n-k}\rangle\right)|0^k\rangle$$

and then apply Hadamard gates on the last k qubits. ∎

Since the weights are real $f(t) = f^*(t)$ and normalized $\sum_{\bar{t}\in S_{t_0}} |f(\bar{t})|^2 = 1$, the transformed operator satisfies the properties listed in Proposition 2.5.1.

Gaussian ansatz

Instead of the plain Fourier Transform, consider the Gaussian-weighted Fourier Transform

$$\hat{\boldsymbol{A}}_f(\bar{\omega}) := \frac{1}{\sqrt{N}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} f(\bar{t}) \boldsymbol{A}(\bar{t}) \quad \text{for} \quad f(\bar{t}) := \frac{1}{\sqrt{\sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\frac{\bar{t}^2}{2\sigma_t^2}}}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\frac{\bar{t}^2}{4\sigma_t^2}} |\bar{t}\rangle.$$

Again, since the weight is real $f(t) = f^*(t)$ and normalized $\int_{-\infty}^{\infty} |f(t)|^2 \mathrm{d}t = 1$, the transformed operator satisfies the symmetry properties listed in Proposition 2.5.1. To implement the above operator, we just need to prepare the initial state approximately.

Proposition 2.6.4 (Preparing a truncated Gaussian state [127])**.** *Suppose* $(Nt_0)^2/16\sigma_t^2 \geq \log(1/\epsilon)$. *Then, the state*

$$\frac{1}{\sqrt{\sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\frac{\bar{t}^2}{2\sigma_t^2}}}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\frac{\bar{t}^2}{4\sigma_t^2}} |\bar{t}\rangle$$

can be prepared using $O(n \log(1/\epsilon)^{5/4})$ *gates up to error* ϵ.

The main advantage of using a Gaussian weight is that its Fourier Transform remains a Gaussian, which has a rapidly decaying tail. Indeed, for the *continuous* Gaussian, we can evaluate the Gaussian integral by completing the square

$$\frac{1}{\sqrt{\sigma_t\sqrt{2\pi}}} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{e}^{-\frac{t^2}{4\sigma_t^2}} \mathrm{d}t = \sqrt{\sigma_t\sqrt{2\pi}} \mathrm{e}^{-\omega^2\sigma_t^2}.$$

The uncertainty in energy is inversely proportional to the uncertainty in time σ_t^{-1}, as a manifestation of the energy-time uncertainty principle.

Discretizing continuous functions via periodic summation

It is not obvious how to carefully derive bounds on the discretization errors that appear in the Riemann sums of the discrete Fourier Transform, as the Fourier phases are highly oscillatory. Nevertheless, the discrete Fourier Transform remains Gaussian, up to a well-controlled error.

Proposition 2.6.5 (DFT of Gaussian)**.** *There is a choice of parameter* $(Nt_0)^2/\sigma_t^2 = \Omega(\log(1/\epsilon))$ *and* $(N\omega_0)^2\sigma_t^2 = \Omega(\log(1/\epsilon))$ *such that the discrete Fourier Transform for*

$$f(\bar{t}) = \frac{1}{\sqrt{\sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\frac{\bar{t}^2}{2\sigma_t^2}}}} \mathrm{e}^{-\frac{\bar{t}^2}{4\sigma_t^2}} \quad \textit{is approximately} \quad \frac{1}{\sqrt{\sum_{\bar{\omega}\in S_{\omega_0}} \mathrm{e}^{-\bar{\omega}^2\sigma_t^2}}} \mathrm{e}^{-\bar{\omega}^2\sigma_t^2}$$

up to error $O(\frac{1}{Nt_0\sigma_t}\mathrm{e}^{-N^2v_0^2\sigma_t^2/2} + \frac{1}{Nt_0/\sigma_t}\mathrm{e}^{-N^2t_0^2/16\sigma_t^2})$ *in 2-norm.*

To relate the continuous Fourier Transform to the discrete one, we apply the discretization to a continuous-variable function after periodic summation. This is related to the *Poisson Summation Formula* [190, Chapter II §13], but pushes the idea one step further to the realm of discrete Fourier Transform. Similar ideas are used, e.g., in lattice cryptography (c.f., [151]), but we include a self-contained treatment for completeness.

To state the following general result, we introduce the notation $\ell_1(\mathbb{R})$ for (Lebesque) integrable $\mathbb{R} \to \mathbb{C}$ functions.

Fact 2.6.1. *Consider the Fourier Transform* $\hat{f}(\omega) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} f(t)\mathrm{d}t$ *of a function* $f \in \ell_1(\mathbb{R})$*. Suppose that a "wrapped around" version of* f *can be defined such that* $p(t) = \sum_{n=-\infty}^{\infty} f(t + nNt_0)$ *for almost every* $t \in \mathbb{R}$ *(i.e., the set of points where the equality does not hold has Lebesgue measure* 0*),* $p(t)$ *is continuous at every* $\bar{t} \in S_{t_0}$ *and Riemann integrable on the interval* $[-\frac{N}{2}t_0, \frac{N}{2}t_0]$*. If the sequence* $k \cdot \hat{f}(k\omega_0) \colon k \in \mathbb{Z}$ *is bounded in absolute value, then the limit* $\hat{p}(\bar{\omega}) := \lim_{B\to\infty} \sum_{\ell=-B}^{B} \hat{f}(\bar{\omega} + \ell N\omega_0)$ *exists for every* $\bar{\omega} \in S_{\omega_0}$ *and* $\hat{p}(\bar{\omega})$ *is the discrete Fourier Transform of* $p(\bar{t})$*, i.e.,*

$$\hat{p}(\bar{\omega}) = \frac{t_0}{\sqrt{2\pi}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} p(\bar{t}).$$

Proof. For every $\omega \in \omega_0 \cdot \mathbb{Z}$ we have that

$$\begin{aligned}
\hat{f}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} f(t)\mathrm{d}t && \text{(since } f \in \ell_1(\mathbb{R})\text{)} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \int_{-\frac{N}{2}t_0}^{\frac{N}{2}t_0} \mathrm{e}^{-\mathrm{i}\omega(t+nNt_0)} f(t+nNt_0)\mathrm{d}t \\
& && \text{(by Fubini's theorem since } f \in \ell_1(\mathbb{R})\text{)} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \int_{-\frac{N}{2}t_0}^{\frac{N}{2}t_0} \mathrm{e}^{-\mathrm{i}\omega t} f(t+nNt_0)\mathrm{d}t && \text{(since } \omega \in \omega_0 \cdot \mathbb{Z} \text{ and } \omega_0 t_0 = \tfrac{2\pi}{N}\text{)} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\frac{N}{2}t_0}^{\frac{N}{2}t_0} \sum_{n=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} f(t+nNt_0)\mathrm{d}t \\
& && \text{(by Fubini's theorem since } f \in \ell_1(\mathbb{R})\text{)} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\frac{N}{2}t_0}^{\frac{N}{2}t_0} \mathrm{e}^{-\mathrm{i}\omega t} p(t)\mathrm{d}t. && \text{(since } p(t) = \textstyle\sum_{n=-\infty}^{\infty} f(t+nNt_0) \text{ a.s.)}
\end{aligned}$$

The above equality means that $c_k := \frac{\sqrt{2\pi}}{Nt_0} \cdot \hat{f}(k \cdot \omega_0)$ is the Fourier series of $p(t)$. Since p is Riemann integrable on the interval $[-\frac{N}{2}t_0, \frac{N}{2}t_0]$, continuous at every $\bar{t} \in S_{t_0}$ point and $k \cdot c_k$ is bounded by assumption, we have for all $\bar{t} \in S_{t_0}$ that

$$\begin{aligned}
p(\bar{t}) &= \lim_{B\to\infty} \sum_{k=-B}^{B} c_k \mathrm{e}^{\mathrm{i}\bar{t}k\omega_0} && \text{(due to [105, Theorem 15.3])} \\
&= \lim_{B\to\infty} \sum_{k=-BN}^{BN} c_k \mathrm{e}^{\mathrm{i}\bar{t}k\omega_0} \\
&= \lim_{B\to\infty} \sum_{k=-BN-\lceil (N-1)/2 \rceil}^{BN+\lfloor (N-1)/2 \rfloor} c_k \mathrm{e}^{\mathrm{i}\bar{t}k\omega_0} && \text{(since } |c_k| = O\Big(\tfrac{1}{k}\Big)\text{)} \\
&= \lim_{B\to\infty} \sum_{\ell=-B}^{B} \sum_{j=-\lceil (N-1)/2 \rceil}^{\lfloor (N-1)/2 \rfloor} c_{\ell N+j} \mathrm{e}^{\mathrm{i}\bar{t}(\ell N+j)\omega_0} && \text{(set } k = \ell N + j\text{)} \\
&= \lim_{B\to\infty} \sum_{\ell=-B}^{B} \sum_{\bar{\omega}\in S_{\omega_0}} \frac{\sqrt{2\pi}}{Nt_0} \hat{f}(\ell N\omega_0 + \bar{\omega}) \mathrm{e}^{\mathrm{i}\bar{t}\bar{\omega}}. \\
& && \text{(set } \bar{\omega} = j\omega_0\text{; use } \bar{t} \in t_0 \cdot \mathbb{Z} \text{ and } \omega_0 t_0 = \tfrac{2\pi}{N}\text{)}
\end{aligned}$$

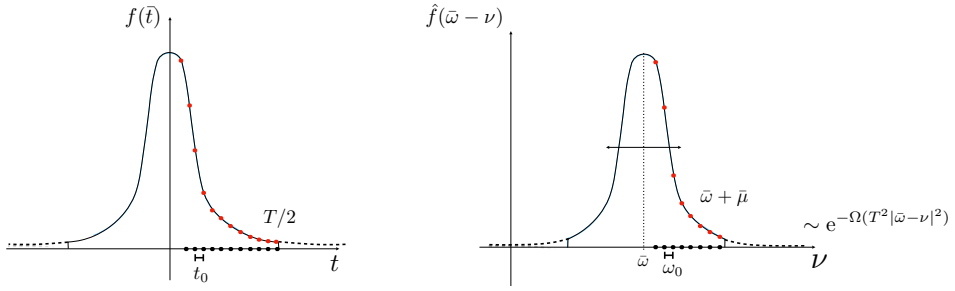


Figure 2.9: Left: the weight function as (approximately) the Gaussian distribution but truncated at T. Right: An illustration for the Fourier Transformed amplitudes, which is also (approximately) Gaussian. It peaks near energy $\nu = \bar{\omega}$ with a width $\sim T^{-1}$ and decays exponentially. The secular approximation truncates the profile at an energy $\bar{\mu}$ in the tail $\bar{\mu} \sim T^{-1}$.

Finally, for all $\bar{\omega}' \in S_{\omega_0}$ we have

$$\begin{aligned}
\frac{t_0}{\sqrt{2\pi}} \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{t}\bar{\omega}'} p(\bar{t}) &= \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{t}\bar{\omega}'} \lim_{B\to\infty} \sum_{\ell=-B}^{B} \sum_{\bar{\omega}\in S_{\omega_0}} \frac{1}{N} \hat{f}(\bar{\omega}+\ell N\omega_0) \mathrm{e}^{\mathrm{i}\bar{t}\bar{\omega}} \\
&= \lim_{B\to\infty} \sum_{\ell=-B}^{B} \sum_{\bar{\omega}\in S_{\omega_0}} \frac{1}{N} \hat{f}(\bar{\omega}+\ell N\omega_0) \sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{-\mathrm{i}\bar{t}\bar{\omega}'} \mathrm{e}^{\mathrm{i}\bar{t}\bar{\omega}} \\
&= \lim_{B\to\infty} \sum_{\ell=-B}^{B} \hat{f}(\bar{\omega}'+\ell N\omega_0) \\
&\qquad \left(\text{since } \textstyle\sum_{\bar{t}\in S_{t_0}} \mathrm{e}^{\mathrm{i}\bar{t}(\bar{\omega}-\bar{\omega}')} = N\cdot\delta_{\bar{\omega},\bar{\omega}'}\right) \\
&= \hat{p}(\bar{\omega}'). \qquad \blacksquare
\end{aligned}$$

Using the above, we prove Proposition 2.6.5.

Proof of Proposition 2.6.5. Apply Fact 2.6.1 for the Gaussian $f(t) = \sqrt{g_\sigma(t)} = \frac{1}{\sqrt{\sigma_t\sqrt{2\pi}}} \mathrm{e}^{-\frac{t^2}{4\sigma_t^2}}$. The problem reduces to implementing the periodic sum $p(\bar{t})$ (Figure 2.9) approximately. Up to a small error from the Gaussian tail, we may only keep the centered Gaussian ($n = 0$) as long as the Gaussian is largely confined in the window for both the time domain $(Nt_0)^2 \gg \sigma_t^2$ and frequency domain $(N\omega_0)^2 \gg \sigma_t^2$. $\blacksquare$

Tail bounds

We evaluate the tail bounds that appear in the analysis of the secular approximation. First, we consider the case of the uniform weights whose Fourier Transform has a

heavy tail impacting the accuracy.

Proposition 2.6.6 (Tail bound for uniform weights)**.** *Let* $f(\bar{t}) := \frac{\mathbb{1}(-T\le\bar{t}<T)}{\sqrt{2T/t_0}}\mathrm{e}^{\mathrm{i}\nu\bar{t}}$*. Then, its discrete Fourier Transform is*

$$\hat{f}(\bar{\omega}) = \frac{1}{\sqrt{2TN/t_0}}\frac{\mathrm{e}^{\mathrm{i}(\nu-\bar{\omega})T} - \mathrm{e}^{-\mathrm{i}(\nu-\bar{\omega})T}}{\mathrm{e}^{\mathrm{i}(\nu-\bar{\omega})t_0} - 1}$$

with a tail bound

$$\sum_{|\bar{\omega}|>m\omega_0}\left|\hat{f}(\bar{\omega})\right|^2 \le \frac{\pi}{2m\omega_0 T}.$$

Proof. The Fourier Transform gives a geometric series with ratio $\mathrm{e}^{\mathrm{i}(\nu-\bar{\omega})\bar{t}}$

$$\frac{1}{\sqrt{N}}\sum_{\bar{t}\in S_{t_0}}\mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}}f(\bar{t}) = \frac{1}{\sqrt{2TN/t_0}}\sum_{-T\le\bar{t}<T}\mathrm{e}^{\mathrm{i}(\nu-\bar{\omega})\bar{t}} = \frac{1}{\sqrt{2TN/t_0}}\frac{\mathrm{e}^{\mathrm{i}(\nu-\bar{\omega})T} - \mathrm{e}^{-\mathrm{i}(\nu-\bar{\omega})T}}{\mathrm{e}^{\mathrm{i}(\nu-\bar{\omega})t_0} - 1}.$$

This function scales inversely with $\bar{\omega}$

$$\begin{aligned}
\sum_{|\bar{\omega}|>m\omega_0}\left|\hat{f}(\bar{\omega})\right|^2 &\le \frac{1}{2NT/t_0}\sum_{|\bar{\omega}|>m\omega_0}\frac{4}{\left|\mathrm{e}^{\mathrm{i}\bar{\omega}t_0}-1\right|^2}\\
&\le \frac{t_0}{2NT}\sum_{|\bar{\omega}|>m\omega_0}\frac{\pi^2}{(\bar{\omega}t_0)^2} \qquad (\text{since } \left|\mathrm{e}^{\mathrm{i}x}-1\right| \ge \tfrac{2}{\pi}|x| \text{ for } x\in[-\pi,\pi])\\
&\le \frac{\pi^2}{NTt_0\omega_0^2}\sum_{n=m+1}^{\infty}\frac{1}{n^2}\\
&\le \frac{\pi}{2T\omega_0}\sum_{n=m+1}^{\infty}\frac{1}{n(n-1)} \qquad (\text{since } \omega_0 t_0 = \tfrac{2\pi}{N})\\
&= \frac{\pi}{2m\omega_0 T}. \qquad \blacksquare
\end{aligned}$$

In retrospect, it is very important that we consider the 2-norm of the tail here; the 1-norm would be divergent.

Now, we consider Gaussians, which have a rapidly decaying tail, greatly improving the accuracy.

Proposition 2.6.7 (Tail bound for Gaussian weights)**.** *For the function* $f(\bar{t}) = (\sum_{\bar{t}\in S_{t_0}}\mathrm{e}^{-\frac{\bar{t}^2}{2\sigma_t^2}})^{-1/2}\mathrm{e}^{-\frac{\bar{t}^2}{4\sigma_t^2}}$*, the Fourier-transformed tail satisfies*

$$\sqrt{\sum_{\substack{\bar{\omega}\in S_{\omega_0}^{\lceil N\rfloor}\\ |\bar{\omega}|\ge\bar{\mu}}}\left|\hat{f}(\bar{\omega})\right|^2} \le O\left(\frac{1}{\sqrt{N\omega_0\sigma_t}}\mathrm{e}^{-N^2\omega_0^2\sigma_t^2/2} + \frac{1}{\sqrt{Nt_0/\sigma_t}}\mathrm{e}^{-N^2t_0^2/16\sigma_t^2} + \frac{1}{\sqrt{\bar{\mu}\sigma_t}}\mathrm{e}^{-\bar{\mu}^2\sigma_t^2}\right).$$

Proof. By Proposition 2.6.5, the discrete Fourier Transform $\hat{f}(\bar{\omega})$ is approximately $\frac{1}{\sum_{\bar{\omega}\in S_{\omega_0}} e^{-\bar{\omega}^2\sigma_t^2}} e^{-2\bar{\omega}^2\sigma_t^2}$ up to error $O(\frac{1}{\sqrt{N\omega_0\sigma_t}} e^{-N^2 v_0^2\sigma_t^2/2} + \frac{1}{\sqrt{Nt_0/\sigma_t}} e^{-N^2 t_0^2/16\sigma_t^2})$. We then control the tail bound

$$\frac{1}{\sum_{\bar{\omega}\in S_{\omega_0}} e^{-\bar{\omega}^2\sigma_t^2}} \sum_{|\bar{\omega}|\geq\bar{\mu}} e^{-2\bar{\omega}^2\sigma_t^2} = O(\frac{1}{\bar{\mu}\sigma_t} e^{-2\bar{\mu}^2\sigma_t^2}).$$ ∎

2.7 Appendix:Proving approximate detailed balance

In this section, we prove approximate detailed balance (or discriminant proxy) for the constructed discriminant. It amounts to controlling the error arising from Boltzmann factors due to the finite resolution of the operator Fourier Transform.

A simpler but weaker bound

We begin with a simpler but weaker bound. This will be enough for the Gaussian-damped discriminant due to its rapidly decaying tail. We can bootstrap this weaker bound using a more careful truncation scheme for the special case of uniform weight (which has a heavy tail) as shown in section 2.7. To prove our error bound, we introduce two useful technical lemmas.

Lemma 2.7.1 (Norm bound on block-band matrices)**.** *Let $V_i \subseteq \mathcal{H}$ and $W_i \subseteq \mathcal{H}'$ be systems of mutually orthogonal subspaces of $\mathcal{H}$ and $\mathcal{H}'$ respectively. If $\boldsymbol{M} = \bigoplus_i \boldsymbol{B}_i$ where $\boldsymbol{B}_i : V_i \to W_i$, then $\|\boldsymbol{M}\| = \max_i \|\boldsymbol{B}_i\|$.*

Proof. We can get a singular value decomposition of $\boldsymbol{M}$ by taking singular value decompositions of each $\boldsymbol{B}_i$ and then merging them. Since $\|\boldsymbol{M}\|$ is the largest singular value, we get the claimed equality. ∎

Lemma 2.7.2 (Norm bounds on sums of tensor products of matrices)**.** *Let $|I| \leq \infty$ and $\boldsymbol{A}_i \in \mathbb{C}^{n\times m}$, $\boldsymbol{B}_i \in \mathbb{C}^{n'\times m'}$ for each $i \in I$, then*

$$\|\sum_{i\in I} \boldsymbol{A}_i[\cdot]\boldsymbol{B}_i^\dagger\|_{2-2} = \|\sum_{i\in I} \boldsymbol{A}_i \otimes \boldsymbol{B}^*\| \leq \sqrt{\|\sum_{i\in I} \boldsymbol{A}_i\boldsymbol{A}_i^\dagger\|\|\sum_{i\in I} \boldsymbol{B}_i^\dagger\boldsymbol{B}_i\|}.$$

Proof. Define the maps

$$\boldsymbol{V} := \sum_{i\in I} \boldsymbol{A}_i \otimes \boldsymbol{I} \otimes \langle i| \quad \text{and} \quad \boldsymbol{U} := \sum_{i\in I} \boldsymbol{I} \otimes \boldsymbol{B}^* \otimes |i\rangle, \qquad \text{then}$$

$$\|\sum_{i\in I} \boldsymbol{A}_i \otimes \boldsymbol{B}^*\| = \|\boldsymbol{V}\boldsymbol{U}\| \leq \|\boldsymbol{V}\|\|\boldsymbol{U}\| \leq \|\sum_{i\in I} \boldsymbol{A}_i\boldsymbol{A}_i^\dagger\|^{1/2}\|\sum_{i\in I} \boldsymbol{B}^{*\dagger}\boldsymbol{B}^*\|^{1/2}.$$

Take complex conjugate to conclude the proof. ∎

Lemma 2.7.3 (Secular approximation gives discriminant proxy)**.** *Consider the following discriminant proxy and a closely related Lindbladian*

$$\begin{aligned}\mathcal{D}_{sec} &:= \sum_{a\in A}\sum_{\bar\omega\in S_{\omega_0}} \sqrt{\gamma(\bar\omega)\gamma(-\bar\omega)}\hat{\boldsymbol{S}}^a(\bar\omega)\otimes\hat{\boldsymbol{S}}^a(\bar\omega)^* \\ &\quad - \frac{\gamma(\bar\omega)}{2}\Big(\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\hat{\boldsymbol{S}}^a(\bar\omega)\otimes\boldsymbol{I} + \boldsymbol{I}\otimes\hat{\boldsymbol{S}}^a(\bar\omega)^{\dagger *}\hat{\boldsymbol{S}}^a(\bar\omega)^*\Big), \\ \mathcal{L}_{sec} &= \sum_{a\in A}\sum_{\bar\omega\in S_{\omega_0}} \gamma(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)[\cdot]\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger - \frac{\gamma(\bar\omega)}{2}\{\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\hat{\boldsymbol{S}}^a(\bar\omega),\cdot\},\end{aligned}$$

such that the nonnegative weights satisfy[42]

$$\gamma(\bar\omega)/\gamma(-\bar\omega) = \mathrm{e}^{-\beta\bar\omega} \quad \textit{or} \quad \gamma(\bar\omega)=\gamma(-\bar\omega)=0 \quad \textit{for each} \quad \bar\omega\in S_{\omega_0}. \tag{2.46}$$

Suppose the operators satisfy that

$$\langle\psi_i|\hat{\boldsymbol{S}}^a(\bar\omega)\big|\psi_j\big\rangle = 0 \quad \textit{whenever} \quad \big|(E_i-E_j)-\bar\omega\big| > \bar\mu \tag{2.47}$$

for the eigenvalue decomposition of $\boldsymbol{H}=\sum_j E_j|\psi_j\rangle\langle\psi_j|$*. Then, for any* $\beta,\bar\mu>0$ *such that* $\beta\bar\mu\le 1$ *and the Gibbs state* $\boldsymbol{\rho}=\mathrm{e}^{-\beta\boldsymbol{H}}/\mathrm{Tr}[\mathrm{e}^{-\beta\boldsymbol{H}}]$*,*

$$\begin{aligned}&\|\mathcal{D}_{sec}-\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})\| \\ &\le \beta\bar\mu\sqrt{\|\sum_{a\in A,\bar\omega\in S_{\omega_0}}\gamma(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\hat{\boldsymbol{S}}^a(\bar\omega)\|} \\ &\left(7\sqrt{\|\sum_{a\in A,\bar\omega\in S_{\omega_0}}\gamma(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\hat{\boldsymbol{S}}^a(\bar\omega)\|}+125\sqrt{\|\sum_{a\in A,\bar\omega\in S_{\omega_0}}\gamma(-\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\|}\right),\end{aligned}$$

where $\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})$ *is the vectorization of* $\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})$*. Finally, if there is a permutation* $\boldsymbol{P}\colon a\to a'$ *such that*[42] $\sqrt{\gamma(-\bar\omega)}\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger = \sqrt{\gamma(-\bar\omega)}\hat{\boldsymbol{S}}^{a'}(-\bar\omega)$ *for each* a *and* $\bar\omega$*, then we have that* $\mathcal{D}_{sec}=\mathcal{D}_{sec}^\dagger$ *and that*

$$\|\mathcal{D}_{sec}-\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})^\dagger\| = \|\mathcal{D}_{sec}-\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})\| \le 132\beta\bar\mu\|\sum_{a\in A}\sum_{\bar\omega\in S_{\omega_0}}\gamma(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\hat{\boldsymbol{S}}^a(\bar\omega)\|.$$

[42] Potentially allowing zero values of $\gamma(\bar\omega)=\gamma(-\bar\omega)=0$ might be needed for dealing with the case when N is even and therefore the smallest label in S_{ω_0} would be its own inverse (due to parity and the modulo arithmetic of S_{ω_0}). Also note that if $\hat{\boldsymbol{S}}^a(\bar\omega)=\boldsymbol{S}^a_{-\bar\omega}=0$ for all $a\in A$, then we can assume without loss of generality that the corresponding weight is $\gamma(\bar\omega)=\gamma(-\bar\omega)=0$.

This also quickly leads to approximate detailed balance for Lindbladians (Lemma 2.2.2) by a triangle inequality

$$\left\|\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})-\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})^{\dagger}\right\|_{2-2}\leq\|\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})-\mathcal{D}_{sec}\|_{2-2}+\|\mathcal{D}_{sec}-\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})^{\dagger}\|_{2-2}$$

$$=\|\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})-\mathcal{D}_{sec}\|+\|\mathcal{D}_{sec}-\mathcal{D}(\boldsymbol{\rho},\mathcal{L}_{sec})^{\dagger}\|.$$

Note that the term $\|\sum_{a,\bar{\omega}}\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{\dagger}\hat{\boldsymbol{S}}^{a}(\bar{\omega})\|$ can be thought of as the "strength" of the interaction, and in our case can be simply bounded by 1 due to Proposition 2.5.1 as follows:

$$\|\sum_{a,\bar{\omega}}\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{\dagger}\hat{\boldsymbol{S}}^{a}(\bar{\omega})\|=\|\sum_{a,\bar{\omega}}\left(\boldsymbol{A}^{a}_{f_{sec}}(\bar{\omega})\right)^{\dagger}\boldsymbol{A}^{a}_{f_{sec}}(\bar{\omega})\|=\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^{a}\|\cdot\||f_{sec}\rangle\|^{2}\leq 1.$$

Proof of Lemma 2.7.3. Our proof adapts from the strategy of [38] for dealing with approximate detailed balance. First, let us define projectors that partition the spectrum per truncation frequency $\bar{\mu}$ as follows:

$$\boldsymbol{P}_{i}:=\sum_{j:\,\frac{E_j}{\bar{\mu}}\in[(i-\frac{1}{2}),(i+\frac{1}{2}))}|\psi_j\rangle\langle\psi_j|\quad\text{for each}\quad i\in\mathbb{Z}\quad\text{such that}\quad\sum_{i}\boldsymbol{P}_{i}=\boldsymbol{I}\quad\text{and}\quad\boldsymbol{P}_{i}\boldsymbol{H}=\boldsymbol{H}\boldsymbol{P}_{i}.$$

In other words, these projectors provide a resolution of the identity, and moreover, they commute with $\boldsymbol{H}$. We proceed by decomposing the difference of the operators $\mathcal{D}_{sec}-\mathcal{D}'$ as follows:

$$\mathcal{D}_{sec}-\mathcal{D}'=\overbrace{\sum_{a,\bar{\omega}}\sqrt{\gamma(\bar{\omega})\gamma(-\bar{\omega})}\hat{\boldsymbol{S}}^{a}(\bar{\omega})\otimes\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{*}-\gamma(\bar{\omega})\boldsymbol{\rho}^{-1/4}\hat{\boldsymbol{S}}^{a}(\bar{\omega})\boldsymbol{\rho}^{1/4}\otimes\boldsymbol{\rho}^{*-1/4}\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{*}\boldsymbol{\rho}^{*1/4}}^{\delta\mathcal{A}:=}$$

$$+\underbrace{\sum_{a,\bar{\omega}}\frac{\gamma(\bar{\omega})}{2}\left(\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{\dagger}\hat{\boldsymbol{S}}^{a}(\bar{\omega})\otimes\boldsymbol{I}-\boldsymbol{\rho}^{-1/4}\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{\dagger}\hat{\boldsymbol{S}}^{a}(\bar{\omega})\boldsymbol{\rho}^{1/4}\otimes\boldsymbol{I}\right)}_{\frac{1}{2}\delta\mathcal{R}\otimes\boldsymbol{I}:=}$$

$$+\underbrace{\sum_{a,\bar{\omega}}\frac{\gamma(\bar{\omega})}{2}\left(\boldsymbol{I}\otimes\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{\dagger *}\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{*}-\boldsymbol{I}\otimes\boldsymbol{\rho}^{*-1/4}\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{*\dagger}\hat{\boldsymbol{S}}^{a}(\bar{\omega})^{*}\boldsymbol{\rho}^{*1/4}\right)}_{\frac{1}{2}\boldsymbol{I}\otimes\delta\mathcal{R}^{*}=}.\tag{2.48}$$

By the triangle inequality we have that $\|\mathcal{D}_{sec} - \mathcal{D}'\| \leq \|\delta\mathcal{A}\| + \frac{1}{2}\|\delta\mathcal{R} \otimes \boldsymbol{I}\| + \frac{1}{2}\|\boldsymbol{I} \otimes \delta\mathcal{R}^*\| = \|\delta\mathcal{A}\| + \|\delta\mathcal{R}\|$. Now, we bound the above two terms individually, starting from the term $\|\delta\mathcal{R}\|$. Our proof crucially relies on the fact that the frequency label of $\hat{\boldsymbol{S}}^a(\bar{\omega})$ closely approximates the true Bohr frequency, up to the truncation frequency $\bar{\mu}$ as expressed by (2.47). This implies that $\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega})$ roughly preserves energy that

$$\langle\psi_i|\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega})|\psi_j\rangle = 0 \quad \text{whenever} \quad |E_i - E_j| > 2\bar{\mu}.$$

Indeed,

$$\begin{aligned}\langle\psi_i|\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega})|\psi_j\rangle &= \sum_k \langle\psi_i|\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger|\psi_k\rangle\langle\psi_k|\hat{\boldsymbol{S}}^a(\bar{\omega})|\psi_j\rangle \\ &= \sum_{k:\ \substack{|(E_k-E_i)-\bar{\omega}|\leq\bar{\mu}\ \&\\ |(E_k-E_j)-\bar{\omega}|\leq\bar{\mu}}} \langle\psi_i|\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger|\psi_k\rangle\langle\psi_k|\hat{\boldsymbol{S}}^a(\bar{\omega})|\psi_j\rangle, \\ &\qquad\qquad \text{(due to (2.47))}\end{aligned}$$

since $|E_i - E_j| = |(E_k - E_j) - \bar{\omega} - ((E_k - E_i) - \bar{\omega})| \leq |(E_k - E_j) - \bar{\omega}| + |(E_k - E_i) - \bar{\omega}|$, meaning that the above summands can only be nonzero when $|E_i - E_j| \leq 2\bar{\mu}$. This observation enables us to introduce the following decomposition

$$\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) = \sum_{i,j} \boldsymbol{P}_i \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \boldsymbol{P}_j = \sum_{\ell=-2}^{2} \sum_i \boldsymbol{P}_i \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \boldsymbol{P}_{i+\ell}. \tag{2.49}$$

Let us define $\delta\boldsymbol{H}_j := \frac{\beta}{4}(\boldsymbol{P}_j \boldsymbol{H} \boldsymbol{P}_j - j\bar{\mu}\boldsymbol{P}_j)$ and $\delta\boldsymbol{H} := \sum_j \delta\boldsymbol{H}_j$. Since $\boldsymbol{H}$ and $\boldsymbol{P}_j$ commute and $\boldsymbol{P}_j = \boldsymbol{P}_j^2$ we have

$$\boldsymbol{\rho}^{\mp\frac{1}{4}} \boldsymbol{P}_j = \boldsymbol{P}_j \boldsymbol{\rho}^{\mp\frac{1}{4}} = (\mathrm{Tr}(\mathrm{e}^{-\beta\boldsymbol{H}}))^{\pm\frac{1}{4}} \cdot \mathrm{e}^{\pm\frac{\beta\bar{\mu}}{4} j} \boldsymbol{P}_j \mathrm{e}^{\pm\delta\boldsymbol{H}}. \tag{2.50}$$

We use (2.49)-(2.50) to exploit the "approximate energy preservation" of the operator $\hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega})$ as follows:

$$\begin{aligned}\boldsymbol{\rho}^{-1/4} \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \boldsymbol{\rho}^{1/4} &= \sum_{\ell=-2}^{2} \sum_i \boldsymbol{\rho}^{-1/4} \boldsymbol{P}_{i+\ell} \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \boldsymbol{P}_i \boldsymbol{\rho}^{1/4} \\ &= \sum_{\ell=-2}^{2} \sum_i \mathrm{e}^{\frac{\beta\bar{\mu}}{4}\ell} \boldsymbol{P}_{i+\ell} \mathrm{e}^{\delta\boldsymbol{H}} \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \mathrm{e}^{-\delta\boldsymbol{H}} \boldsymbol{P}_i. \end{aligned} \tag{2.51}$$

Let us define $\boldsymbol{S}' := \sum_{a,\bar{\omega}} \gamma(\bar{\omega}) \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega})$, then we get the following bound on $\|\delta\mathcal{R}\|$:

$$\begin{aligned}
\|\delta\mathcal{R}\| &= \left\| \sum_{a,\bar{\omega}} \gamma(\bar{\omega}) \left(\boldsymbol{\rho}^{-1/4} \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \boldsymbol{\rho}^{1/4} - \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \right) \right\| && \text{(by (2.48))} \\
&= \left\| \sum_{a,\bar{\omega}} \gamma(\bar{\omega}) \left(\sum_{\ell=-2}^{2} \sum_i \mathrm{e}^{\frac{\beta\bar{\mu}}{4}\ell} \boldsymbol{P}_{i+\ell} \mathrm{e}^{\delta\boldsymbol{H}} \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \mathrm{e}^{-\delta\boldsymbol{H}} \boldsymbol{P}_i - \boldsymbol{P}_{i+\ell} \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \boldsymbol{P}_i \right) \right\| \\
& && \text{(by (2.49)-(2.51))} \\
&\leq \sum_{\ell=-2}^{2} \left\| \sum_i \boldsymbol{P}_{i+\ell} \sum_{a,\bar{\omega}} \gamma(\bar{\omega}) \left(\mathrm{e}^{\frac{\beta\bar{\mu}}{4}\ell} \mathrm{e}^{\delta\boldsymbol{H}} \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \mathrm{e}^{-\delta\boldsymbol{H}} - \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \right) \boldsymbol{P}_i \right\| \\
& && \text{(by triangle inequality)} \\
&= \sum_{\ell=-2}^{2} \max_i \left\| \boldsymbol{P}_{i+\ell} \sum_{a,\bar{\omega}} \gamma(\bar{\omega}) \left(\mathrm{e}^{\frac{\beta\bar{\mu}}{4}\ell} \mathrm{e}^{\delta\boldsymbol{H}} \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \mathrm{e}^{-\delta\boldsymbol{H}} - \hat{\boldsymbol{S}}^a(\bar{\omega})^\dagger \hat{\boldsymbol{S}}^a(\bar{\omega}) \right) \boldsymbol{P}_i \right\|. \\
& && \text{(by Lemma 2.7.1)}
\end{aligned}$$

We may now drop the project $\boldsymbol{P}_i$ and $\boldsymbol{P}_{i+\ell}$ and simplify via elementary bounds.

$$\begin{aligned}
(cont'd) &\leq \sum_{\ell=-2}^{2} \left\| \mathrm{e}^{(\delta\boldsymbol{H} - \frac{\beta\bar{\mu}}{4}\ell\boldsymbol{I})} \boldsymbol{S}' \mathrm{e}^{-\delta\boldsymbol{H}} - \boldsymbol{S}' \right\| \\
&\leq \sum_{\ell=-2}^{2} \left\| \mathrm{e}^{(\delta\boldsymbol{H} - \frac{\beta\bar{\mu}}{4}\ell\boldsymbol{I})} \boldsymbol{S}' \mathrm{e}^{-\delta\boldsymbol{H}} - \boldsymbol{S}' \mathrm{e}^{-\delta\boldsymbol{H}} \right\| + \left\| \boldsymbol{S}' \mathrm{e}^{-\delta\boldsymbol{H}} - \boldsymbol{S}' \right\| \\
&\leq \sum_{\ell=-2}^{2} \| \mathrm{e}^{(\delta\boldsymbol{H} - \frac{\beta\bar{\mu}}{4}\ell\boldsymbol{I})} - \boldsymbol{I} \| \| \boldsymbol{S}' \| \| \mathrm{e}^{-\delta\boldsymbol{H}} \| + \| \boldsymbol{S}' \| \left\| \mathrm{e}^{-\delta\boldsymbol{H}} - \boldsymbol{I} \right\| \\
&\leq \| \boldsymbol{S}' \| \sum_{\ell=-2}^{2} \| 2\delta\boldsymbol{H} - \frac{\beta\bar{\mu}}{2}\ell\boldsymbol{I} \| (1 + \| 2\delta\boldsymbol{H} \|) + \| 2\delta\boldsymbol{H} \| \\
& && \text{(since } |\mathrm{e}^x - 1| \leq 2|x| \text{ for } |x| \leq \tfrac{5}{4}\text{)} \\
&\leq \| \boldsymbol{S}' \| \sum_{\ell=-2}^{2} \beta\bar{\mu} \left(\frac{1}{4} + \frac{|\ell|}{2} \right) \left(1 + \frac{\beta\bar{\mu}}{4} \right) + \frac{\beta\bar{\mu}}{4} \\
& && \text{(since } \|\delta\boldsymbol{H}\| \leq \tfrac{\beta\bar{\mu}}{8} \text{ and } \beta\bar{\mu} \leq 1\text{)} \\
&\leq 7\beta\bar{\mu} \| \boldsymbol{S}' \|.
\end{aligned}$$

Next, we bound $\delta\mathcal{A}$ in a similar fashion. The expression will be more cumbersome

because of the double Hilbert spaces. Decompose $\delta\mathcal{A} = \sum_{a\in A}\delta\mathcal{A}^a$, where

$$\delta\mathcal{A}^a := \sum_{\bar\omega\in S_{\omega_0}} \underbrace{\sqrt{\gamma(\bar\omega)\gamma(-\bar\omega)}}_{\mathring\gamma(\bar\omega):=}\hat{\boldsymbol{S}}^a(\bar\omega)\otimes\hat{\boldsymbol{S}}^a(\bar\omega)^* - \sum_{\bar\omega\in S_{\omega_0}}\gamma(\bar\omega)\boldsymbol{\rho}^{-1/4}\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{\rho}^{1/4}\otimes\boldsymbol{\rho}^{*-1/4}\hat{\boldsymbol{S}}^a(\bar\omega)^*\boldsymbol{\rho}^{*1/4}$$

$$\overset{\text{(by (2.46))}}{=} \sum_{\bar\omega\in S_{\omega_0}}\mathring\gamma(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)\otimes\hat{\boldsymbol{S}}^a(\bar\omega)^* - \mathring\gamma(\bar\omega)\mathrm{e}^{-\frac{\beta}{2}\bar\omega}\boldsymbol{\rho}^{-1/4}\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{\rho}^{1/4}\otimes\boldsymbol{\rho}^{*-1/4}\hat{\boldsymbol{S}}^a(\bar\omega)^*\boldsymbol{\rho}^{*1/4}. \tag{2.52}$$

Let $\lfloor\bar\omega\rceil$ denote the rounding of $\frac{\bar\omega}{\bar\mu}$ to the closest integer, and suppose that $\boldsymbol{P}_i\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{P}_j \neq 0$. Then, there must exist some $|\psi_i\rangle, |\psi_j\rangle$ eigenvectors in the images of $\boldsymbol{P}_i, \boldsymbol{P}_j$ respectively such that $\langle\psi_i|\hat{\boldsymbol{S}}^a(\bar\omega)|\psi_j\rangle \neq 0$. Due to (2.47) we have that $|E_i - E_j - \bar\omega| \leq \bar\mu$. Then $|i-j-\lfloor\bar\omega\rceil| \leq |i-\frac{E_i}{\bar\mu}| + |\frac{E_j}{\bar\mu} - j| + |\frac{\bar\omega}{\bar\mu} - \lfloor\bar\omega\rceil| + |\frac{E_i}{\bar\mu} - \frac{E_j}{\bar\mu} - \frac{\bar\omega}{\bar\mu}| \leq \frac{1}{2}+\frac{1}{2}+\frac{1}{2}+1 < 3$, thus we can define a bisection analogously to (2.49) as follows:

$$\hat{\boldsymbol{S}}^a(\bar\omega) = \sum_{i,j}\boldsymbol{P}_i\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{P}_j = \sum_{i,j\,:\,|i-j-\lfloor\bar\omega\rceil|<3}\boldsymbol{P}_i\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{P}_j = \sum_{\ell=-2}^{2}\sum_i \boldsymbol{P}_i\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{P}_{i-\lfloor\bar\omega\rceil+\ell},$$

which leads to the following "tensor-slicing" assuming that $\boldsymbol{P}_n$ and $\boldsymbol{M}_m$ commute:

$$\begin{aligned}&\boldsymbol{M}_1\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{M}_2\otimes\boldsymbol{M}_3^*\hat{\boldsymbol{S}}^a(\bar\omega)^*\boldsymbol{M}_4^*\\ &= \sum_{\ell,\ell'=-2}^{2}\sum_{i,j}(\boldsymbol{P}_j\otimes\boldsymbol{P}_{j+i}^*)\Big(\boldsymbol{M}_1\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{M}_2\otimes\boldsymbol{M}_3^*\hat{\boldsymbol{S}}^a(\bar\omega)^*\boldsymbol{M}_4^*\Big)(\boldsymbol{P}_{j-\lfloor\bar\omega\rceil+\ell}\otimes\boldsymbol{P}_{j+i-\lfloor\bar\omega\rceil+\ell'}^*).\end{aligned}$$

Using this we get the following decomposition analogously to (2.51) by expressing $\boldsymbol{\rho}^{-1/4}\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{\rho}^{1/4}$ via (2.52) and (2.50)

$$\begin{aligned}\delta\mathcal{A}^a &= \sum_{\ell,\ell'=-2}^{2}\sum_i\delta\mathcal{A}^a_{i,\ell,\ell'} = \sum_{\ell,\ell'=-2}^{2}\sum_i\\ &\sum_{j,\bar\omega}\mathring\gamma(\bar\omega)(\boldsymbol{P}_j\otimes\boldsymbol{P}_{j+i}^*)\Big(\hat{\boldsymbol{S}}^a(\bar\omega)\otimes\hat{\boldsymbol{S}}^a(\bar\omega)^* - \mathrm{e}^{\frac{\beta\bar\mu}{2}(\lfloor\bar\omega\rceil-\frac{\bar\omega}{\bar\mu}-\frac{\ell+\ell'}{2})}\mathrm{e}^{\delta\boldsymbol{H}}\hat{\boldsymbol{S}}^a(\bar\omega)\mathrm{e}^{-\delta\boldsymbol{H}}\otimes\mathrm{e}^{\delta\boldsymbol{H}^*}\hat{\boldsymbol{S}}^a(\bar\omega)^*\mathrm{e}^{-\delta\boldsymbol{H}^*}\Big)\\ &\cdot(\boldsymbol{P}_{j-\lfloor\bar\omega\rceil+\ell}\otimes\boldsymbol{P}_{j+i-\lfloor\bar\omega\rceil+\ell'}^*).\end{aligned}$$

Since $|\frac{\bar\omega}{\bar\mu} - \lfloor\bar\omega\rceil| \leq \frac{1}{2}$ and $|\ell|, |\ell'| \leq 2$, we have that $|\lfloor\bar\omega\rceil - \frac{\bar\omega}{\bar\mu} - \frac{\ell+\ell'}{2}| \leq \frac{5}{2}$ and therefore the above factor is close to 1:

$$|\mathrm{e}^{\frac{\beta\bar\mu}{2}(\lfloor\bar\omega\rceil-\frac{\bar\omega}{\bar\mu}-\frac{\ell+\ell'}{2})} - 1| \leq \beta\bar\mu\left|\lfloor\bar\omega\rceil - \frac{\bar\omega}{\bar\mu} - \frac{\ell+\ell'}{2}\right| \leq \frac{5}{2}\beta\bar\mu. \tag{2.53}$$

At this point, it seems intuitively clear that the error coming from the Boltzmann factor is small. However, we need to argue how the sum over j does not blow up the error. We proceed by using the triangle inequality over ℓ, ℓ' and then Lemma 2.7.1 over i to get that

$$\|\delta\mathcal{A}\| = \|\sum_{a\in A}\delta\mathcal{A}^a\| \le 5^2 \max_{i,\ell,\ell'} \Big\| \underbrace{\sum_a \delta\mathcal{A}^a_{i,\ell,\ell'}}_{\delta\mathcal{A}_{i,\ell,\ell'}:=} \Big\|.$$

The key for bounding the norm of $\delta\mathcal{A}_{i,\ell,\ell'}$ is applying Lemma 2.7.2 to estimate the following for some weights $\alpha(\bar\omega)$:

$$\begin{aligned}
&\|\sum_{j,a,\bar\omega}\alpha(\bar\omega)\mathring\gamma(\bar\omega)(\boldsymbol{P}_j\otimes\boldsymbol{P}^*_{j+i})\Big(\hat{\boldsymbol{S}}^a(\bar\omega)\otimes\hat{\boldsymbol{S}}^a(\bar\omega)^*\Big)(\boldsymbol{P}_{j-\lfloor\bar\omega\rceil+\ell}\otimes\boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'})\|^2\\
&=\|\sum_{j,a,\bar\omega}\alpha(\bar\omega)\boldsymbol{P}_j\sqrt{\gamma(-\bar\omega)}\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{P}_{j-\lfloor\bar\omega\rceil+\ell}\otimes\boldsymbol{P}^*_{j+i}\sqrt{\gamma(\bar\omega)}\hat{\boldsymbol{S}}^a(\bar\omega)^*\boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'}\|^2\\
&\le\|\sum_{j,a,\bar\omega}|\alpha(\bar\omega)|^2\gamma(-\bar\omega)\boldsymbol{P}_j\quad\underbrace{\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{P}_{j-\lfloor\bar\omega\rceil+\ell}\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger}_{\le\sum_k\hat{\boldsymbol{S}}^a(\bar\omega)\boldsymbol{P}_k\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger=\hat{\boldsymbol{S}}^a(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger}\quad\boldsymbol{P}_j\|\\
&\cdot\|\sum_{j,a,\bar\omega}\gamma(\bar\omega)\boldsymbol{P}_{j+i-\lfloor\bar\omega\rceil+\ell'}\underbrace{\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\boldsymbol{P}_{j+i}\hat{\boldsymbol{S}}^a(\bar\omega)}_{\le\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\hat{\boldsymbol{S}}^a(\bar\omega)}\boldsymbol{P}_{j+i-\lfloor\bar\omega\rceil+\ell'}\|\qquad\text{(by Lemma 2.7.2)}\\
&\le\|\alpha\|^2_\infty\Big\|\sum_j\boldsymbol{P}_j\underbrace{\sum_{a,\bar\omega}\gamma(-\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger}_{\boldsymbol{S}''}\boldsymbol{P}_j\Big\|\Big\|\sum_j\boldsymbol{P}_j\underbrace{\sum_{a,\bar\omega}\gamma(\bar\omega)\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger\hat{\boldsymbol{S}}^a(\bar\omega)}_{=\boldsymbol{S}'}\boldsymbol{P}_j\Big\|\\
&\qquad\qquad\text{since } 0\le\boldsymbol{A}\le\boldsymbol{B}\Rightarrow\|\boldsymbol{A}\|\le\|\boldsymbol{B}\|\\
&=\|\alpha\|^2_\infty\underbrace{\max_{j''}\Big\|\boldsymbol{P}_{j''}\boldsymbol{S}''\boldsymbol{P}_{j''}\Big\|\cdot\max_{j'}\Big\|\boldsymbol{P}_{j'}\boldsymbol{S}'\boldsymbol{P}_{j'}\Big\|}_{\text{due to Lemma 2.7.1}}\le\|\alpha\|^2_\infty\|\boldsymbol{S}'\|\|\boldsymbol{S}''\|.\qquad(2.54)
\end{aligned}$$

Let us introduce a telescoping sum $\delta\mathcal{A}_{i,\ell,\ell'}=\sum_{s=1}^3\delta\mathcal{A}^{(s)}_{i,\ell,\ell'}$, where $\mathring\gamma_{\ell,\ell'}(\bar\omega):=\mathrm{e}^{\frac{\beta\bar\mu}{2}(\lfloor\bar\omega\rceil-\frac{\bar\omega}{\bar\mu}-\frac{\ell+\ell'}{2})}\mathring\gamma(\bar\omega)$ and

$$\begin{aligned}
&\delta\mathcal{A}^{(1)}_{i,\ell,\ell'}\\
&:=\sum_{j,a,\bar\omega}\mathring\gamma(\bar\omega)(\boldsymbol{P}_j\otimes\boldsymbol{P}^*_{j+i})\Big(\hat{\boldsymbol{S}}^a(\bar\omega)\otimes\hat{\boldsymbol{S}}^a(\bar\omega)^*-\mathrm{e}^{\frac{\beta\bar\mu}{2}(\lfloor\bar\omega\rceil-\frac{\bar\omega}{\bar\mu}-\frac{\ell+\ell'}{2})}\hat{\boldsymbol{S}}^a(\bar\omega)\otimes\hat{\boldsymbol{S}}^a(\bar\omega)^*\Big)(\boldsymbol{P}^*_{j-\lfloor\bar\omega\rceil+\ell}\otimes\boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'}),
\end{aligned}$$

$$\delta\mathcal{A}^{(2)}_{i,\ell,\ell'}$$
$$:= \sum_{j,a,\bar\omega} \mathring\gamma_{\ell,\ell'}(\bar\omega)(\boldsymbol{P}_j \otimes \boldsymbol{P}^*_{j+i})\Big(\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \hat{\boldsymbol{S}}^a(\bar\omega)^* - \mathrm{e}^{-\delta\boldsymbol{H}}\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}\hat{\boldsymbol{S}}^a(\bar\omega)^*\Big)$$
$$(\boldsymbol{P}^*_{j-\lfloor\bar\omega\rceil+\ell} \otimes \boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'}),$$

$$\delta\mathcal{A}^{(3)}_{i,\ell,\ell'} := \sum_{j,a,\bar\omega}$$
$$\mathring\gamma_{\ell,\ell'}(\bar\omega)(\boldsymbol{P}_j \otimes \boldsymbol{P}^*_{j+i})\Big(\mathrm{e}^{-\delta\boldsymbol{H}}\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}\hat{\boldsymbol{S}}^a(\bar\omega)^* - \mathrm{e}^{-\delta\boldsymbol{H}}\hat{\boldsymbol{S}}^a(\bar\omega)\mathrm{e}^{\delta\boldsymbol{H}} \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}\hat{\boldsymbol{S}}^a(\bar\omega)^*\mathrm{e}^{\delta\boldsymbol{H}^*}\Big)$$

$$(\boldsymbol{P}^*_{j-\lfloor\bar\omega\rceil+\ell} \otimes \boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'}).$$

Due to (2.53) we can bound $\|\delta\mathcal{A}^{(1)}_{i,\ell,\ell'}\| \leq \frac{5}{2}\beta\bar\mu\sqrt{\|\boldsymbol{S}'\|\|\boldsymbol{S}''\|}$ via (2.54). For bounding $\|\delta\mathcal{A}^{(2)}_{i,\ell,\ell'}\|$ observe that

$$\|\delta\mathcal{A}^{(2)}_{i,\ell,\ell'}\|$$
$$= \|\sum_{j,a,\bar\omega} \mathring\gamma_{\ell,\ell'}(\bar\omega)(\boldsymbol{P}_j \otimes \boldsymbol{P}^*_{j+i})\Big(\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \hat{\boldsymbol{S}}^a(\bar\omega)^* - \mathrm{e}^{-\delta\boldsymbol{H}}\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}\hat{\boldsymbol{S}}^a(\bar\omega)^*\Big)(\boldsymbol{P}^*_{j-\lfloor\bar\omega\rceil+\ell} \otimes \boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'})\|$$

$$= \|(\boldsymbol{I}\otimes\boldsymbol{I} - \mathrm{e}^{-\delta\boldsymbol{H}} \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}) \sum_{j,\bar\omega} \mathring\gamma_{\ell,\ell'}(\bar\omega)(\boldsymbol{P}_j \otimes \boldsymbol{P}^*_{j+i})\Big(\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \hat{\boldsymbol{S}}^a(\bar\omega)^*\Big)(\boldsymbol{P}^*_{j-\lfloor\bar\omega\rceil+\ell} \otimes \boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'})\|$$

$$\leq \underbrace{\|\boldsymbol{I}\otimes\boldsymbol{I} - \mathrm{e}^{-\delta\boldsymbol{H}} \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}\|}_{\leq\frac{2}{7}\beta\bar\mu \text{ since } \|\delta\boldsymbol{H}\|\leq\frac{\beta\bar\mu}{8}} \underbrace{\|\sum_{j,\bar\omega} \mathring\gamma_{\ell,\ell'}(\bar\omega)(\boldsymbol{P}_j \otimes \boldsymbol{P}^*_{j+i})\Big(\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \hat{\boldsymbol{S}}^a(\bar\omega)^*\Big)(\boldsymbol{P}^*_{j-\lfloor\bar\omega\rceil+\ell} \otimes \boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'})\|}_{\leq(1+\frac{5}{2}\beta\bar\mu)\sqrt{\|\boldsymbol{S}'\|\|\boldsymbol{S}''\|} \text{ due to (2.53) and (2.54)}}$$

$$\leq \beta\bar\mu\sqrt{\|\boldsymbol{S}'\|\|\boldsymbol{S}''\|}.$$

Analogously we can bound $\|\delta\mathcal{A}^{(3)}_{i,\ell,\ell'}\|$ as follows:

$$\begin{aligned}
&\|\delta\mathcal{A}^{(3)}_{i,\ell,\ell'}\| \\
&= \|(\mathrm{e}^{-\delta\boldsymbol{H}} \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}) \sum_{j,\bar\omega} \mathring{\gamma}_{\ell,\ell'}(\bar\omega)(\boldsymbol{P}_j \otimes \boldsymbol{P}^*_{j+i})\Big(\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \hat{\boldsymbol{S}}^a(\bar\omega)^*\Big)(\boldsymbol{P}^*_{j-\lfloor\bar\omega\rceil+\ell} \otimes \boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'})
\end{aligned}$$

$$\begin{aligned}
&(\boldsymbol{I}\otimes\boldsymbol{I} - \mathrm{e}^{\delta\boldsymbol{H}} \otimes \mathrm{e}^{\delta\boldsymbol{H}^*})\| \\
&\le \underbrace{\|\mathrm{e}^{-\delta\boldsymbol{H}} \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}\|}_{\le 1+\frac{2}{7}\beta\bar\mu}\| \underbrace{\sum_{j,\bar\omega} \mathring{\gamma}_{\ell,\ell'}(\bar\omega)(\boldsymbol{P}_j \otimes \boldsymbol{P}^*_{j+i})\Big(\hat{\boldsymbol{S}}^a(\bar\omega) \otimes \hat{\boldsymbol{S}}^a(\bar\omega)^*\Big)(\boldsymbol{P}^*_{j-\lfloor\bar\omega\rceil+\ell} \otimes \boldsymbol{P}^*_{j+i-\lfloor\bar\omega\rceil+\ell'})\|}_{\le(1+\frac{5}{2}\beta\bar\mu)\sqrt{\|\boldsymbol{S}'\|\|\boldsymbol{S}''\|} \text{ due to (2.53) and (2.54)}}
\end{aligned}$$

$$\begin{aligned}
&\underbrace{\|\boldsymbol{I}\otimes\boldsymbol{I} - \mathrm{e}^{-\delta\boldsymbol{H}} \otimes \mathrm{e}^{-\delta\boldsymbol{H}^*}\|}_{\le\frac{2}{7}\beta\bar\mu \text{ since } \|\delta\boldsymbol{H}\|\le\frac{\beta\bar\mu}{8}} \\
&\le \frac{3}{2}\beta\bar\mu\sqrt{\|\boldsymbol{S}'\|\|\boldsymbol{S}''\|}.
\end{aligned}$$

Putting everything together, we get that

$$\|\delta\mathcal{A}\| \le 25 \max_{i,\ell,\ell'} \|\delta\mathcal{A}_{i,\ell,\ell'}\| \le 25 \max_{i,\ell,\ell'}\Big(\|\delta\mathcal{A}^{(1)}_{i,\ell,\ell'}\| + \|\delta\mathcal{A}^{(2)}_{i,\ell,\ell'}\| + \|\delta\mathcal{A}^{(3)}_{i,\ell,\ell'}\|\Big) \le 125\beta\bar\mu\sqrt{\|\boldsymbol{S}'\|\|\boldsymbol{S}''\|},$$

which concludes the proof for the first bound.

Finally, if $\sqrt{\gamma(-\bar\omega)}\hat{\boldsymbol{S}}^a(\bar\omega)^\dagger = \sqrt{\gamma(-\bar\omega)}\hat{\boldsymbol{S}}^{a'}(-\bar\omega)$, then $\boldsymbol{S}' = \boldsymbol{S}''$ and $\mathcal{D}_{sec} = \mathcal{D}^\dagger_{sec}$, so we easily get the other bound

$$\begin{aligned}
\|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\| &= \|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}^\dagger_{sec})\|_{2\text{–}2} = \|\mathcal{D}^\dagger_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})\|_{2\text{–}2} \\
&= \|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})\|_{2\text{–}2} = \|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})\| \le 132\beta\bar\mu\|\boldsymbol{S}'\|.
\end{aligned}$$

■

Bootstrapping the secular approximation

Lemma 2.7.4 (Bootstrapping the secular approximation)**.** *Consider the decomposition $f = \sum_{j\in J} f_j$ of the weight function, where $f_j\colon S_{t_0} \to \mathbb{C}$, and let $\mu_j := \min\{\mu \ge 0\colon \|\hat f_j(\omega)\cdot\mathbb{1}(|\omega|>\mu)\| = 0\}$. If the Hamiltonian $\bar{\boldsymbol{H}}$ has discretized spectrum so that $B \subset \omega_0\mathbb{Z}$, $\beta\mu_j \le 1$, $\max_{\bar\omega\in B, j\in J}\bar\omega + \mu_j =: \nu \in S_{\omega_0}$, $\gamma\colon S_{\omega_0} \to \mathbb{R}_+$ is such that $\gamma(\bar\omega)/\gamma(-\bar\omega) = \mathrm{e}^{-\beta\bar\omega}$ for all $\bar\omega \in [-\nu,\nu]\cap S_{\omega_0}$, and the set of jumps is*

self-adjoint $\{\boldsymbol{A}^a : a \in A\} = \{\boldsymbol{A}^{a\dagger} : a \in A\}$, *then*[43]

$$\|\mathcal{D}_f - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_f)^\dagger\| = \|\mathcal{D}_f - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_f)\| \leq \sum_{i\in J} \|f_i\|\mu_i \sum_{j\in J} \|f_j\| \|\gamma\|_\infty 1056\beta \| \sum_{a\in A} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|,$$

where $\boldsymbol{\rho} = \mathrm{e}^{-\beta \boldsymbol{H}}/\mathrm{Tr}[\mathrm{e}^{-\beta \boldsymbol{H}}]$,

$$\mathcal{D}_f := \boldsymbol{I}[\cdot]\boldsymbol{I} + \sum_{a,\bar{\omega}\in S_{\omega_0}} \sqrt{\gamma(\bar{\omega})\gamma(-\bar{\omega})}\hat{\boldsymbol{A}}^a_f(\bar{\omega})[\cdot]\hat{\boldsymbol{A}}^a_f(\bar{\omega})^\dagger - \frac{\gamma(\bar{\omega})}{2}\{\hat{\boldsymbol{A}}^a_f(\bar{\omega})^\dagger \hat{\boldsymbol{A}}^a_f(\bar{\omega}), \cdot\}, \text{ and}$$

$$\mathcal{L}_f := \sum_{a\in A,\bar{\omega}\in S_{\omega_0}} \gamma(\bar{\omega})\hat{\boldsymbol{A}}^a_f(\bar{\omega})[\cdot]\hat{\boldsymbol{A}}^a_f(\bar{\omega})^\dagger - \frac{\gamma(\bar{\omega})}{2}\{\hat{\boldsymbol{A}}^a_f(\bar{\omega})^\dagger \hat{\boldsymbol{A}}^a_f(\bar{\omega}), \cdot\}.$$

Proof. The proof builds on the following "polarization" identity: for all $\boldsymbol{M}_i$ matrices and $c_1, c_2 \in \mathbb{R}$

$$\boldsymbol{M}_1 \star \boldsymbol{M}_4 + \boldsymbol{M}_2 \star \boldsymbol{M}_3 = \frac{c_1 c_2}{2} \sum_{s=\pm 1} s\left(\frac{\boldsymbol{M}_1}{c_1} + s\frac{\boldsymbol{M}_2}{c_2}\right) \star \left(\frac{\boldsymbol{M}_3}{c_1} + s\frac{\boldsymbol{M}_4}{c_2}\right),$$

where $\star$ stands for any operation that is distributive with $+$, e.g., matrix product $\star = \cdot$ or tensor product $\star = \otimes$.

Due to the linearity of the operator Fourier Transform, we have that $\hat{\boldsymbol{A}}^a_f(\bar{\omega}) = \sum_{j\in J} \boldsymbol{A}^a_{f_j,\bar{\omega}}$ and consequently

$$\mathcal{D}_f = \sum_{i,j\in J}\left(1 - \frac{\delta_{ij}}{2}\right)\frac{\|f_i\|\|f_j\|}{2} \sum_{s=\pm 1} s\mathcal{D}_{\frac{f_i}{\|f_i\|} + s\frac{f_j}{\|f_j\|}},$$
$$\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_f) = \sum_{i,j\in J}\left(1 - \frac{\delta_{ij}}{2}\right)\frac{\|f_i\|\|f_j\|}{2} \sum_{s=\pm 1} s\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{\frac{f_i}{\|f_i\|} + s\frac{f_j}{\|f_j\|}}).$$

Due to the properties of the operator Fourier Transform (Proposition 2.5.1), we have that

$$\begin{aligned}\|\sum_{a\in A}\sum_{\bar{\omega}\in S_{\omega_0}} \gamma(\bar{\omega})\boldsymbol{A}^a_g(\bar{\omega})^\dagger \hat{\boldsymbol{A}}^a_g(\bar{\omega})\| &\leq \|\gamma\|_\infty \|\sum_{a\in A}\sum_{\bar{\omega}\in S_{\omega_0}} \boldsymbol{A}^a_g(\bar{\omega})^\dagger \hat{\boldsymbol{A}}^a_g(\bar{\omega})\| \\ &= \|\sum_{a\in A}\sum_{\bar{t}\in S_{t_0}} |g(\bar{t})|^2 \mathrm{e}^{\mathrm{i}\boldsymbol{H}\bar{t}}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a \mathrm{e}^{-\mathrm{i}\boldsymbol{H}\bar{t}}\| \leq \|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \|g\|^2,\end{aligned}$$

[43] Note that here we use notation $\mathcal{D}_f$, $\mathcal{L}_f$ instead of $\mathcal{D}_\beta$, $\mathcal{L}_\beta$ to spell out the dependence on f instead of β.

since $\|\frac{f_i}{\|f_i\|} + s\frac{f_j}{\|f_j\|}\| \le 2$, $\beta\mu_j \le 1$ for all $j \in J$, and[42] $\nu \in S_{\omega_0}$ (i.e., no wrapping around), by Lemma 2.7.3 we get

$$\begin{aligned}
&\|\mathcal{D}_{\frac{f_i}{\|f_i\|}+s\frac{f_j}{\|f_j\|}} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{\frac{f_i}{\|f_i\|}+s\frac{f_j}{\|f_j\|}})\| \\
&\le \max\{\mu_i,\mu_j\}\|\gamma\|_\infty 4\beta\sqrt{\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|}\left(7\sqrt{\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|} + 125\sqrt{\|\sum_{a\in A}\boldsymbol{A}^a\boldsymbol{A}^{a\dagger}\|}\right) \\
&\le \max\{\mu_i,\mu_j\}\underbrace{\|\gamma\|_\infty 528\beta\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|}_{K:=}.
\end{aligned}$$

The second line uses that $\{\boldsymbol{A}^a : a \in A\} = \{\boldsymbol{A}^{a\dagger} : a \in A\}$ thus $\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| = \|\sum_{a\in A}\boldsymbol{A}^a\boldsymbol{A}^{a\dagger}\|$.

Finally, by the triangle inequality, we get that

$$\begin{aligned}
\|\mathcal{D}_f - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_f)\| &\le \sum_{i,j\in J}\left(1-\frac{\delta_{ij}}{2}\right)\frac{\|f_i\|\|f_j\|}{2}\sum_{s=\pm1}\|\mathcal{D}_{\frac{f_i}{\|f_i\|}+s\frac{f_j}{\|f_j\|}} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{\frac{f_i}{\|f_i\|}+s\frac{f_j}{\|f_j\|}})\| \\
&\le \sum_{i,j\in J}\left(1-\frac{\delta_{ij}}{2}\right)\|f_i\|\|f_j\|\max\{\mu_i,\mu_j\}K \\
&\le \sum_{i,j\in J}\|f_i\|\|f_j\|(\mu_i+\mu_j)K \\
&= 2\sum_{i\in J}\|f_i\|\mu_i\sum_{j\in J}\|f_j\|K. \qquad \blacksquare
\end{aligned}$$

Corollary 2.7.1 (Improved bounds for uniform weights)**.** *In the setting of Lemma 2.7.4, consider the uniform weight function* $f(\bar{t}) = \sqrt{\frac{t_0}{2T}}\mathbb{1}(-T \le \bar{t} < T)$ *and the secular approximation with* $s(\bar{\omega}) = \mathbb{1}(|\bar{\omega}| \le \mu)$ *for some* $\mu \ge \pi/T$*, then*

$$\|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\| = \|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})\| = O\left(\beta\|\gamma\|_\infty\sqrt{\frac{\mu}{T}}\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\right).$$

Proof. Let $f_s := \mathcal{F}^{*-1}(\hat{f}\cdot s)$; by Proposition 2.6.2, we know that $\mathcal{D}_{sec}$ can be obtained by utilizing the weight function f_s.

We decompose s into exponentially increasing intervals. We set $s_0 := \mathbb{1}(|\bar{\omega}| \le \pi/T)$ and

$$s_j := \mathbb{1}(4^{j-1}\pi/T < |\bar{\omega}| \le \min\{4^j\pi/T, \mu\}) \quad \forall j \in \mathbb{Z}_+.$$

Let $f_j := \mathcal{F}^{*-1}(\hat{f} \cdot s_j)$; since $s = \sum_{j=0}^{\lceil \log_4(\mu T/\pi)\rceil} s_j$ we have $f_s = \sum_{j=0}^{\lceil \log_4(\mu T/\pi)\rceil} f_j$. Observe that due to Proposition 2.6.6

$$\|f_j\| = \|\hat{f} \cdot s_j\| \le \|\hat{f} \cdot \mathbb{1}(4^{j-1}\pi/T < |\bar{\omega}|)\| \le \sqrt{4^{1-j}} = 2^{1-j}, \quad \text{and thus} \quad \sum_j \|f_j\| \le 4.$$

The result follows from Lemma 2.7.4 since $\hat{f}_j = \hat{f} \cdot s_j$ and by the definition of s_j we have $\mu_j \le 4^j \pi/T$ and thus

$$\sum_{j=0}^{\lceil \log_4(\mu T/\pi)\rceil} \|f_j\|\mu_j \le \sum_{j=0}^{\lceil \log_4(\mu T/\pi)\rceil} \frac{2\pi}{T} 2^j \le \frac{4\pi}{T} 2^{\lceil \log_4(\mu T/\pi)\rceil} \le \frac{8\pi}{T} 2^{\log_4(\mu T/\pi)}$$
$$= \frac{8\pi}{T}\sqrt{\mu T/\pi} = 8\sqrt{\frac{\pi\mu}{T}}.$$

■

Fourier Transform with uniform weights

For simpler implementation, we can also work with the Fourier Transform with uniform weight (which is not smooth), leading to slightly worse bounds than the Gaussian damped case of Theorem 2.1.3.

Theorem 2.7.1 (Uniform weight for Fourier Transform)**.** *Consider the discriminant proxy $\mathcal{D}_\beta$ (2.26) with the plain Fourier Transform $\hat{\boldsymbol{A}}^a(\bar{\omega}) :\propto \sum_{-T\le \bar{t}<T} \boldsymbol{A}^a(\bar{t}) \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}}$. Let $\nu := \frac{1}{\beta} + \max_{\omega\in B} \omega$ such that $\nu \le \max_{\bar{\omega}\in S_{\omega_0}}$, $\gamma\colon S_{\omega_0} \to \mathbb{R}_+$ is such that $\gamma(\bar{\omega})/\gamma(-\bar{\omega}) = \mathrm{e}^{-\beta\bar{\omega}}$ for all $\bar{\omega} \in [-\nu, \nu] \cap S_{\omega_0}$ and the set of jumps are self-adjoint and normalized (2.10), then the (normalized) top eigenvector approximates the purified Gibbs state $\left|\sqrt{\boldsymbol{\rho}_\beta}\right\rangle$ such that*

$$\left\| |\lambda_1(\mathcal{D}_\beta)\rangle - \left|\sqrt{\boldsymbol{\rho}_\beta}\right\rangle \right\| \le O\left(\frac{1}{\lambda_{gap}(\mathcal{D}_\beta)}(\omega_0 T + \sqrt{\frac{\beta}{T}})\right).$$

The block-encoding for the discriminant proxy can be implemented exactly using Hamiltonian simulation time $O(T)$ using the construction outlined in section 2.3-section 2.3.

Even though with a worse asymptotic bound, the plain Fourier Transform is simpler to implement and closer to thermalization in nature (section 2.9). The proof is even simpler than the Gaussian case, partly because $\mathcal{D} = \mathcal{D}_{impl}$ as the uniform weights can be prepared exactly.

Proof of Theorem 2.7.1. We can assume without loss of generality that $T/\beta \geq \pi$, since otherwise, the bound is vacuous. We bound the eigenvector distance by the operator norm using Proposition 2.3.3 and recall the secular approximation (Lemma 2.6.2, Proposition 2.6.6) and the improved bounds on approximate detailed balance (Corollary 2.7.1):

$$\begin{aligned}
\left\| |\lambda_1(\mathcal{D}_\beta)\rangle - |\sqrt{\boldsymbol{\rho}}\rangle \right\| &\leq \frac{6\left\|\mathcal{D}_\beta - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\right\|}{\lambda_{gap}(\mathcal{D}_\beta)} \\
&\leq \frac{6}{\lambda_{gap}(\mathcal{D}_\beta)}\left(\|\mathcal{D}_\beta - \mathcal{D}_{sec}\| + \|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\|\right) \\
&\leq O\left(\frac{1}{\lambda_{gap}(\mathcal{D}_\beta)}(\omega_0 T + \frac{1}{\sqrt{\mu T}} + \beta\sqrt{\frac{\mu}{T}})\right) \quad \text{(now set } \mu := 1/\beta) \\
&\leq O\left(\frac{1}{\lambda_{gap}(\mathcal{D}_\beta)}(\omega_0 T + \sqrt{\frac{\beta}{T}})\right). \qquad \blacksquare
\end{aligned}$$

Unfortunately, the above result suggests that the Hamiltonian simulation time needs to scale with the inverse gap squared λ_{gap}^{-2}; we do not know if better bounds are possible.

2.8 Appendix:Discretization error for Lindbladians and discriminant proxies

In this appendix, we bound the discretization error for continuous Lindbladians. We use the notation established in section 2.5. In addition, for a function $f : \mathbb{R} \to \mathbb{C}$, by $\bar{\mathcal{F}}(f(\bar{t}))$, we mean the discrete Fourier Transform of the vector obtained by evaluating f at the points $\bar{t} \in S_{t_0}^{\lceil N \rfloor}$. Also, we define the "discretized" version $\bar{f}(t) := \sqrt{t_0} f(t)$ with a natural rescaling.

We begin with a seemingly loose bound that will, however, be sufficient.

Lemma 2.8.1. *If $f, g, h \in \ell_2(\mathbb{R})$ and $\gamma \in \ell_\infty(\mathbb{R})$, then for any norm $|||\cdot|||$, we have that*

$$\begin{aligned}
&\left|\left|\left| \int_{-\infty}^{\infty} \gamma(\omega) \hat{\boldsymbol{A}}_f(\omega)^\dagger \star \hat{\boldsymbol{A}}_f(\omega) \mathrm{d}\omega - \sum_{\bar{\omega} \in S_{\omega_0}^{\lceil N \rfloor}} g(\bar{\omega}) \hat{\boldsymbol{A}}_h(\bar{\omega})^\dagger \star \hat{\boldsymbol{A}}_h(\bar{\omega}) \right|\right|\right| \\
&\quad \leq \sum_{\nu, \nu' \in B} \left|\left|\left| (\boldsymbol{A}_\nu)^\dagger \star \boldsymbol{A}_{\nu'} \right|\right|\right| \\
&\left| \int_{-\infty}^{\infty} \gamma(\omega) \hat{f}^*(\omega - \nu) \hat{f}(\omega - \nu') \mathrm{d}\omega - \sum_{\bar{\omega} \in S_{\omega_0}^{\lceil N \rfloor}} g(\bar{\omega}) \bar{\mathcal{F}}\left(h(\bar{t}) \cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega}) \bar{\mathcal{F}}\left(h(\bar{t}) \cdot \mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega}) \right|,
\end{aligned}$$

where $\star$ stands for any operation that is distributive with $+$, e.g., matrix product $\star = \cdot$ or tensor product $\star = \otimes$.

Directly applying the above for the original Hamiltonian $\boldsymbol{H}$ suffers from the number of the Bohr frequencies $|B(\boldsymbol{H})|$, which can generally scale with the Hilbert space dimension if the eigenvalue differences are nondegenerate. However, we will see that the above becomes sufficiently stringent if we consider a *rounded* Hamiltonian $\bar{\boldsymbol{H}}$, substantially reducing the number of distinct Bohr frequencies $\left|B(\bar{\boldsymbol{H}})\right|$ while staying close to the original Hamiltonian $\boldsymbol{H}$.

Proof. We use the defining decomposition of the continuous operator Fourier Transform from Proposition 2.5.2

$$\hat{\boldsymbol{A}}_f(\omega) = \sum_{\nu \in B} \hat{f}(\omega - \nu) \boldsymbol{A}_\nu,$$

and its discrete counterpart

$$\hat{\boldsymbol{A}}_h(\bar{\omega}) = \sum_{\nu \in B} \bar{\mathcal{F}}\left(h(\bar{t}) \cdot \mathrm{e}^{\mathrm{i}\nu\bar{t}}\right)(\bar{\omega}) \boldsymbol{A}_\nu,$$

where $\bar{\mathcal{F}}$ denotes the discrete Fourier Transform defined via $\bar{\mathcal{F}}(h(\bar{t}))(\bar{\omega}) = \frac{1}{\sqrt{N}} \sum_{\bar{t} \in S_{t_0}^{\lceil N \rfloor}} h(\bar{t}) \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}}$, where $t_0 = \frac{2\pi}{N\omega_0}$.

Due to the distributivity of $+$ and $\star$ we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} \gamma(\omega) \hat{\boldsymbol{A}}_f(\omega)^\dagger \star \hat{\boldsymbol{A}}_f(\omega) \mathrm{d}\omega - \sum_{\bar{\omega} \in S_{\omega_0}^{\lceil N \rfloor}} g(\bar{\omega}) \hat{\boldsymbol{A}}_h(\bar{\omega})^\dagger \star \hat{\boldsymbol{A}}_h(\bar{\omega}) \\
&\quad = \sum_{\nu, \nu' \in B} (\boldsymbol{A}_\nu)^\dagger \star \boldsymbol{A}_{\nu'} \\
&\left(\int_{-\infty}^{\infty} \gamma(\omega) \hat{f}^*(\omega - \nu) \hat{f}(\omega - \nu') \mathrm{d}\omega - \sum_{\bar{\omega} \in S_{\omega_0}^{\lceil N \rfloor}} g(\bar{\omega}) \bar{\mathcal{F}}\left(h(\bar{t}) \cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega}) \bar{\mathcal{F}}\left(h(\bar{t}) \cdot \mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega}) \right).
\end{aligned}$$

We conclude the proof by using the triangle inequality. ∎

We proceed by controlling the discretization error for the scalar integral. To do so, we also need to regularize the filter function and the transition weight $\gamma(\omega)$ by truncations

$$\begin{aligned}
f_T(t) &:= f(t) \cdot \mathbb{1}(t \in [-T/2, T/2)) \\
\gamma_W(\omega) &:= \gamma(\omega) \cdot \mathbb{1}(\omega \in [-W/2, W/2)) \,.
\end{aligned}$$

Lemma 2.8.2 (Discretization error bounds for the integral)**.** *Let $\bar{\mathcal{F}}$ denote the discrete Fourier Transform with parameters N, ω_0, t_0 and consider $f \in \ell_2(\mathbb{R})$ and $\gamma \in \ell_\infty(\mathbb{R})$ with truncation parameters T, W satisfying $N \geq T/t_0 \in \mathbb{Z}$ and $N \geq W/\omega_0 \in \mathbb{Z}$. Then, for each $\nu, \nu' \in [-K, K]$,*

$$\left| \int_{-\infty}^{\infty} \underbrace{\gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')}_{g(\omega):=} \mathrm{d}\omega - \sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N \rfloor}} \gamma_W(\bar{\omega})\bar{\mathcal{F}}\left(\sqrt{t_0} f_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega})\bar{\mathcal{F}}\left(\sqrt{t_0} f_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega}) \right| \leq \epsilon$$

holds provided the following conditions: [44]

$$\sum_{k\in\mathbb{Z}} \omega_0 \|g(\omega) - g(k\omega_0)\|_{[k\omega_0,(k+1)\omega_0)} \leq \frac{\epsilon}{2}, \tag{2.55}$$

and

$$\sum_{\bar{t}\in S_{t_0}^{\lceil N \rfloor}} t_0 \|f_T(t) - f_T(\bar{t})\|_{[\bar{t},\bar{t}+t_0)} \leq \delta, \qquad t_0 \|f_T(t)\|_\infty T(W+K) \leq \delta, \qquad \|f - f_T\|_1 \leq (\sqrt{2\pi}-2)\delta, \tag{2.56}$$

where $\delta = \min\left(\frac{\epsilon}{4W\|\gamma_W\|_\infty(\|f\|_1+1)}, 2\right)$. [45]

Proof. By (2.55) we have that

$$\begin{aligned} \left| \int_{\mathbb{R}} g(\omega)\mathrm{d}\omega - \sum_{k\in\mathbb{Z}} g(k\omega_0)\omega_0 \right| &\leq \sum_{k\in\mathbb{Z}} \left| \int_{k\omega_0}^{(k+1)\omega_0} g(\omega) - g(k\omega_0)\mathrm{d}\omega \right| \\ &\leq \sum_{k\in\mathbb{Z}} \omega_0 \|g(\omega) - g(k\omega_0)\|_{[k\omega_0,(k+1)\omega_0)} \leq \frac{\epsilon}{2}. \end{aligned} \tag{2.57}$$

[44] Note that the definition of $g(\omega)$ depends on the values ν, ν', however we do not explicitly indicate this dependence for ease of notation.

[45] Note that in case N is odd in order to match the spacing of the grid $S^{\lceil N \rfloor}$, the intervals $[k\omega_0, (k+1)\omega_0)$ should be changed to $[(k-\frac{1}{2})\omega_0, (k+\frac{1}{2})\omega_0)$ and analogously the endpoints of $[\bar{t}, \bar{t}+t_0)$ should be shifted to $\bar{t} \pm t_0/2$.

Also observe that due to (2.56) we have for all $\omega \in \mathbb{R}$ that

$$\begin{aligned}
&\sum_{\bar{t}\in S_{t_0}^{\lceil N\rfloor}} t_0 \|f_T(t)\mathrm{e}^{-\mathrm{i}\omega t} - f_T(\bar{t})\mathrm{e}^{-\mathrm{i}\omega \bar{t}}\|_{[\bar{t},\bar{t}+t_0)} \\
&\qquad \le \sum_{\bar{t}\in S_{t_0}^{\lceil N\rfloor}} t_0 \|f_T(t)\mathrm{e}^{-\mathrm{i}\omega t} - f_T(\bar{t})\mathrm{e}^{-\mathrm{i}\omega t}\|_{[\bar{t},\bar{t}+t_0)} + t_0 \|f_T(\bar{t})\mathrm{e}^{-\mathrm{i}\omega t} - f_T(\bar{t})\mathrm{e}^{-\mathrm{i}\omega \bar{t}}\|_{[\bar{t},\bar{t}+t_0)} \\
&\qquad \le \sum_{\bar{t}\in S_{t_0}^{\lceil N\rfloor}} t_0 \|f_T(t) - f_T(\bar{t})\|_{[\bar{t},\bar{t}+t_0)} + t_0 |f_T(\bar{t})| \|\mathrm{e}^{-\mathrm{i}\omega t} - \mathrm{e}^{-\mathrm{i}\omega \bar{t}}\|_{[\bar{t},\bar{t}+t_0)} \\
&\qquad \le \sum_{\bar{t}\in S_{t_0}^{\lceil N\rfloor}} t_0 \|f_T(t) - f_T(\bar{t})\|_{[\bar{t},\bar{t}+t_0)} + t_0 |f_T(\bar{t})| t_0 |\omega| \\
&\qquad \le \delta + t_0 \|f_T(t)\|_\infty T |\omega|.
\end{aligned}$$

Next we define $\hat{\tilde{f}}_T(\omega) := \frac{t_0}{\sqrt{2\pi}} \sum_{\bar{t}\in S_{t_0}} f_T(\bar{t})\mathrm{e}^{-\mathrm{i}\omega\bar{t}}$. The above two inequalities imply similarly to (2.57) that for all ω satisfying $|\omega| \le W + K$ we have

$$\begin{aligned}
|\hat{\tilde{f}}_T(\omega) - \hat{f}(\omega)| &\le \left|\hat{\tilde{f}}_T(\omega) - \frac{1}{\sqrt{2\pi}} \int f_T(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t\right| + \left|\frac{1}{\sqrt{2\pi}} \int (f_T(t) - f(t))\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t\right| \\
&\le \sum_{\bar{t}\in S_{t_0}} \left|\frac{t_0}{\sqrt{2\pi}} f_T(\bar{t})\mathrm{e}^{-\mathrm{i}\omega\bar{t}} - \frac{1}{\sqrt{2\pi}} \int_{\bar{t}}^{\bar{t}+t_0} f_T(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t\right| + (\sqrt{2\pi} - 2)\frac{\delta}{\sqrt{2\pi}} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{\bar{t}\in S_{t_0}} \left|\int_{\bar{t}}^{\bar{t}+t_0} f_T(\bar{t})\mathrm{e}^{-\mathrm{i}\omega\bar{t}} - f_T(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t\right| + (\sqrt{2\pi} - 2)\frac{\delta}{\sqrt{2\pi}} \\
&\le \frac{2\delta}{\sqrt{2\pi}} + (\sqrt{2\pi} - 2)\frac{\delta}{\sqrt{2\pi}} \\
&= \delta = \min\left(\frac{\epsilon}{4W\|\gamma_W\|_\infty(\|f\|_1 + 1)}, 2\right). \qquad (2.58)
\end{aligned}$$

Let $\tilde{g}(\omega) := \gamma_W(\omega)\hat{\tilde{f}}_T^*(\omega - \nu)\hat{\tilde{f}}_T(\omega - \nu')$. Considering that $|\nu|, |\nu'| \le K$ we get that

for all $\omega \in \mathbb{R}$

$$
\begin{aligned}
&|g(\omega) - \tilde{g}(\omega)| \\
&\leq |\gamma_W(\omega)| \\
&\Big(|\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu') - \hat{f}^*(\omega-\nu)\hat{\tilde{f}}_T(\omega-\nu')| + |\hat{f}^*(\omega-\nu)\hat{\tilde{f}}_T(\omega-\nu') - \hat{\tilde{f}}_T^*(\omega-\nu)\hat{\tilde{f}}_T(\omega-\nu')|\Big) \\
&= |\gamma_W(\omega)|\Big(|\hat{f}^*(\omega-\nu)||\hat{f}(\omega-\nu') - \hat{\tilde{f}}_T(\omega-\nu')| + |\hat{f}^*(\omega-\nu) - \hat{\tilde{f}}_T^*(\omega-\nu)||\hat{\tilde{f}}_T(\omega-\nu')|\Big) \\
&\leq \delta|\gamma_W(\omega)|\Big(|\hat{f}^*(\omega-\nu)| + |\hat{\tilde{f}}_T(\omega-\nu')|\Big) && \text{(by (2.58))} \\
&\leq \delta|\gamma_W(\omega)|\Big(|\hat{f}^*(\omega-\nu)| + |\hat{f}(\omega-\nu')| + |\hat{\tilde{f}}_T(\omega-\nu') - \hat{f}(\omega-\nu')|\Big) \\
&\leq \delta|\gamma_W(\omega)|(2\|f\|_1 + 2) && \text{(by (2.58))} \\
&\leq \frac{\epsilon}{2W}. && \text{(by (2.58))}
\end{aligned}
$$

This implies

$$
\left|\sum_{k\in\mathbb{Z}} \omega_0(g(k\omega_0) - \tilde{g}(k\omega_0))\right| \leq \omega_0 \sum_{k\in\mathbb{Z}} |g(k\omega_0) - \tilde{g}(k\omega_0)| \leq \sum_{\substack{k\in\mathbb{Z} \\ k\omega_0\in[-W/2,W/2)}} \frac{\epsilon\omega_0}{2W} \leq \frac{\epsilon}{2}, \tag{2.59}
$$

showing that

$$
\begin{aligned}
&\left|\int_{\mathbb{R}} g(\omega)\mathrm{d}\omega - \sum_{k\in\mathbb{Z}} \omega_0 \tilde{g}(k\omega_0)\right| \\
&= \left|\int_{\mathbb{R}} \gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega - \sum_{k\in\mathbb{Z}} \omega_0\gamma_W(k\omega_0)\hat{\tilde{f}}_T^*(k\omega_0-\nu)\hat{\tilde{f}}_T(k\omega_0-\nu')\right| \\
&\leq \left|\int_{\mathbb{R}} \gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega - \sum_{k\in\mathbb{Z}} \omega_0\gamma_W(k\omega_0)\hat{f}^*(k\omega_0-\nu)\hat{f}(k\omega_0-\nu')\right| \\
&+ \left|\omega_0 \sum_{k\in\mathbb{Z}} \gamma_W(k\omega_0)\hat{f}^*(k\omega_0-\nu)\hat{f}(k\omega_0-\nu') - \gamma_W(k\omega_0)\hat{\tilde{f}}_T^*(k\omega_0-\nu)\hat{\tilde{f}}_T(k\omega_0-\nu')\right| \\
&\leq \epsilon. && \text{(by (2.57) and (2.59))}
\end{aligned}
$$

We conclude the proof by observing that due to $W \leq N\omega_0$, $T \leq Nt_0$, and $\frac{\omega_0 t_0}{2\pi} = \frac{1}{N}$ we have

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} \omega_0 \tilde{g}(k\omega_0) &= \omega_0 \sum_{\bar{\omega}\in\mathbb{Z}\omega_0} \gamma_W(\bar{\omega}) \hat{\bar{f}}_T^*(\bar{\omega}-\nu) \hat{\bar{f}}_T(\bar{\omega}-\nu') \\
&= \frac{\omega_0 t_0^2}{2\pi} \sum_{\bar{\omega}\in\mathbb{Z}\omega_0} \gamma_W(\bar{\omega}) \left(\sum_{\bar{t}\in\mathbb{Z}t_0} f_T(\bar{t}) \mathrm{e}^{-\mathrm{i}(\bar{\omega}-\nu)\bar{t}} \right)^* \left(\sum_{\bar{t}\in\mathbb{Z}t_0} f_T(\bar{t}) \mathrm{e}^{-\mathrm{i}(\bar{\omega}-\nu')\bar{t}} \right) \\
&= \frac{1}{N} \sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N \rfloor}} \gamma_W(\bar{\omega}) \left(\sum_{\bar{t}\in S_{t_0}^{\lceil N \rfloor}} \sqrt{t_0} f_T(\bar{t}) \mathrm{e}^{\mathrm{i}\nu\bar{t}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} \right)^* \left(\sum_{\bar{t}\in S_{t_0}^{\lceil N \rfloor}} \sqrt{t_0} f_T(\bar{t}) \mathrm{e}^{\mathrm{i}\nu'\bar{t}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} \right) \\
&= \sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N \rfloor}} \gamma_W(\bar{\omega}) \bar{\mathcal{F}}\left(\sqrt{t_0} f_T(\bar{t}) \cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})} \right)^*(\bar{\omega}) \bar{\mathcal{F}}\left(\sqrt{t_0} f_T(\bar{t}) \cdot \mathrm{e}^{(\mathrm{i}\nu'\bar{t})} \right)(\bar{\omega}).
\end{aligned}$$

■

As a sanity check, the above implies that the discretized Lindbladian convergences to the continuum *in the limit*.

Theorem 2.8.1 (Discretizations converge to the continuum)**.** *If $\gamma \in \ell_\infty(\mathbb{R})$, $f \in \ell_2(\mathbb{R})$, and γ, f are continuous almost everywhere (i.e., the set of points of discontinuity has measure zero) while f is bounded on every finite interval, then*

$$\begin{aligned}
&\lim_{W,T\to\infty} \lim_{N\to\infty} \sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N \rfloor}} \gamma_W(\bar{\omega}) \left(\hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega})^\dagger [\cdot] \hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega}) - \frac{1}{2} \{ \hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega})^\dagger \hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega}), \cdot \} \right) \\
&\qquad = \int_{-\infty}^{\infty} \gamma(\omega) \left(\hat{\boldsymbol{A}}_f(\omega)^\dagger [\cdot] \hat{\boldsymbol{A}}_f(\omega) - \frac{1}{2} \{ \hat{\boldsymbol{A}}_f(\omega)^\dagger \hat{\boldsymbol{A}}_f(\omega), \cdot \} \right) \mathrm{d}\omega,
\end{aligned}$$

and if $\gamma \geq 0$, then also

$$\begin{aligned}
&\lim_{W,T\to\infty} \lim_{N\to\infty} \sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N \rfloor}} \sqrt{\gamma_W(\bar{\omega})\gamma_W(-\bar{\omega})} \hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega}) \otimes \hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega})^* \\
&\qquad - \frac{\gamma(\bar{\omega})}{2} \left(\hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega})^\dagger \hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega}) \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega})^{\dagger *} \hat{\boldsymbol{A}}_{\bar{f}_T}(\bar{\omega})^* \right) \\
&\qquad = \int_{-\infty}^{\infty} \sqrt{\gamma(\omega)\gamma(-\omega)} \hat{\boldsymbol{A}}_f(\omega) \otimes \hat{\boldsymbol{A}}_f(\omega)^* \\
&\qquad - \frac{\gamma(\omega)}{2} \left(\hat{\boldsymbol{A}}_f(\omega)^\dagger \hat{\boldsymbol{A}}_f(\omega) \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \hat{\boldsymbol{A}}_f(\omega)^{\dagger *} \hat{\boldsymbol{A}}_f(\omega)^* \right) \mathrm{d}\omega,
\end{aligned}$$

where $\omega_0 = t_0 = \sqrt{2\pi/N}$ and $\bar{f}_T(t) := \sqrt{t_0} f_T(t)$.

Proof. Due to Lemma 2.8.1, it suffices to prove for all $\nu, \nu' \in B$ that

$$\begin{aligned}
&\lim_{W,T\to\infty}\lim_{N\to\infty}\sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N\rfloor}}\gamma_W(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega})\\
&\qquad = \int_{-\infty}^{\infty}\gamma(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega,\\
&\lim_{W,T\to\infty}\lim_{N\to\infty}\sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N\rfloor}}\sqrt{\gamma_W(\bar{\omega})\gamma_W(-\bar{\omega})}\bar{\mathcal{F}}\left(\bar{f}_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)^*(\bar{\omega})\\
&\qquad = \int_{-\infty}^{\infty}\sqrt{\gamma(\omega)\gamma(-\omega)}\hat{f}(\omega-\nu)\hat{f}^*(\omega-\nu')\mathrm{d}\omega. \qquad (2.60)
\end{aligned}$$

Since $f \in \ell_2(\mathbb{R})$, we have $\hat{f} \in \ell_2(\mathbb{R})$ and therefore by Hölder's inequality we get that $\gamma(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu') \in \ell_1(\mathbb{R})$, which then implies

$$\int_{-\infty}^{\infty}\gamma(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega = \lim_{W\to\infty}\int_{-W}^{W}\gamma(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega.$$

Since $f = \lim_{T\to\infty} f_T$, Parseval's Theorem implies $\hat{f} = \lim_{T\to\infty}\hat{f}_T$ and so by Hölder's inequality we get

$$\int_{-W}^{W}\gamma(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega = \lim_{T\to\infty}\int_{-W}^{W}\gamma(\omega)\hat{f}_T^*(\omega-\nu)\hat{f}_T(\omega-\nu')\mathrm{d}\omega.$$

As f_T is bounded, continuous almost everywhere, and has compact support, the Lebesgue-Vitali Theorem [137, 154, Theorem 11.33] implies that it is Riemann integrable. Therefore, $\hat{f}_T$ is bounded and continuous which similarly implies that $g(\omega) := \gamma_W(\omega)\hat{f}_T^*(\omega-\nu)\hat{f}_T(\omega-\nu')$ is Riemann integrable on $[-W, W]$.

Since $f \leftarrow f_T$, $\lim_{N\to 0}\omega_0 = \lim_{N\to 0} t_0 = 0$, and $f_T(t), g(\omega)$ are Riemann integrable (c.f. [48, Theorem 8.26.]), for every $\epsilon > 0$ (2.55)-(2.56) are satisfied for large enough N. Thus Lemma 2.8.2 implies that

$$\int_{-W}^{W}\gamma(\omega)\hat{f}_T^*(\omega-\nu)\hat{f}_T(\omega-\nu')\mathrm{d}\omega = \lim_{N\to\infty}\sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N\rfloor}}\gamma_W(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega}).$$

Equation (2.60) can be analogously proven after replacing $\gamma(\omega)$ with $\sqrt{\gamma(\omega)\gamma(-\omega)}$ throughout the argument. ∎

We believe that a similar result can be shown for any $\gamma \in \ell_\infty(\mathbb{R})$, $f \in \ell_2(\mathbb{R})$ without the other assumptions in Theorem 2.8.1 by applying a further approximation with

the help of mollifiers. While such an asymptotic result is conceptually elegant, for quantum algorithm implementation, we need quantitative, nonasymptotic error bounds for the particular functions we encounter.

Proposition 2.8.1 (Discretization error of Lindbladians and discriminant proxies)**.** *In the setting of Lemma 2.8.2, assume continuity and boundedness assumption for f, γ as in Theorem 2.8.1 with normalization $\|\gamma\|_\infty, \|f\|_2 \le 1$. Consider a single jump operator $\boldsymbol{A}$ with $\|\boldsymbol{A}\| \le 1$, assume that $\gamma(\omega)$ and $\sqrt{\gamma(\omega)\gamma(-\omega)}$ are C-Lipschitz continuous, $\hat{f}$ is $(D \cdot \|\hat{f}\|_\infty)$-Lipschitz continuous, f_T is $(L \cdot \|f_T\|_\infty)$-Lipschitz continuous on $[-T/2, T/2)$, and assume the following conditions:*

$$\|f_T - f\|_2 \le \frac{\epsilon}{64}, \quad \sum_{\bar{\omega} \in S_{t_0}^{\lceil N \rfloor}} t_0 |f(\bar{t}) - f_T(\bar{t})|^2 \le \left(\frac{\epsilon}{64}\right)^2, \quad \sum_{\bar{\omega} \in S_{t_0}^{\lceil N \rfloor}} t_0 |f(\bar{t})|^2 \le 1$$
$$\int_{-\infty}^{\infty} |\hat{f}(\omega) \mathbb{1}(|\omega| \ge W - 2\|\boldsymbol{H}\|)|^2 \mathrm{d}\omega \le \frac{\epsilon}{8\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon} + 1\right)^2},$$

and $\|f - f_T\|_1 \le (\sqrt{2\pi} - 2)\delta$ for $\delta = \min\left(\frac{\epsilon}{32\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon}+1\right)^2 W(\|f\|_1+1)}, 2\right)$, and

$$t_0 \le \frac{\delta}{T\|f_T\|_\infty} \min\left(\frac{1}{W + 2\|\boldsymbol{H}\|}, \frac{1}{L}\right),$$
$$\omega_0 \le \frac{\epsilon}{16} \cdot \min\left(\frac{1}{8T}, \frac{1}{W(C + 2D)\|\hat{f}\|_\infty^2 \left(\frac{256\|\boldsymbol{H}\|T}{\epsilon} + 1\right)^2}\right). \tag{2.61}$$

Then, in the notation of Corollary 2.6.1-2.6.2,

$$\|\mathcal{L}_{(f,\boldsymbol{H})} - \bar{\mathcal{L}}_{(f,\boldsymbol{H})}^{(W)}\|_{1-1} \le \epsilon, \qquad \|\mathcal{D}_{(f,\boldsymbol{H})} - \bar{\mathcal{D}}_{(f,\boldsymbol{H})}^{(W)}\| \le \epsilon,$$

where the discretized Lindbladian and discriminant proxy uses $\gamma_W(\omega)$ instead of $\gamma(\omega)$. Moreover, if

$$\sum_{\bar{\omega} \in S_{\omega_0}^{\lceil N \rfloor}} \left|\bar{\mathcal{F}}\left(\bar{f}(\bar{t}) \mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega}) \mathbb{1}(|\bar{\omega}| \ge W)\right|^2 \le \frac{\epsilon}{8\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon} + 1\right)^2} \quad \text{for each} \quad |\nu| \le 2\|\boldsymbol{H}\|, \tag{2.62}$$

then

$$\|\mathcal{L}_{(f,\boldsymbol{H})} - \bar{\mathcal{L}}_{(f,\boldsymbol{H})}\|_{1-1} \le \epsilon, \qquad \|\mathcal{D}_{(f,\boldsymbol{H})} - \bar{\mathcal{D}}_{(f,\boldsymbol{H})}\| \le \epsilon.$$

Proof. Let $\bar{\boldsymbol{H}}$ be the Hamiltonian obtained by rounding the spectrum of $\boldsymbol{H}$ (down in absolute value) to $\mathbb{Z}\eta$; this is crucial before we invoke Lemma 2.8.1. Our proof proceeds using the following triangle inequalities:

$$\|\mathcal{L}_{(f,\boldsymbol{H})} - \tilde{\mathcal{L}}_{(f,\boldsymbol{H})}\|_{1-1} \le \|\mathcal{L}_{(f,\boldsymbol{H})} - \mathcal{L}_{(f,\bar{\boldsymbol{H}})}\|_{1-1} + \|\mathcal{L}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{L}}_{(f,\bar{\boldsymbol{H}})}\|_{1-1} + \|\tilde{\mathcal{L}}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{L}}_{(f,\boldsymbol{H})}\|_{1-1}$$

$$\|\mathcal{D}_{(f,\boldsymbol{H})} - \tilde{\mathcal{D}}_{(f,\boldsymbol{H})}\| \le \|\mathcal{D}_{(f,\boldsymbol{H})} - \mathcal{D}_{(f,\bar{\boldsymbol{H}})}\| + \|\mathcal{D}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{D}}_{(f,\bar{\boldsymbol{H}})}\| + \|\tilde{\mathcal{D}}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{D}}_{(f,\boldsymbol{H})}\|,$$

where $\tilde{\mathcal{L}}_{(f,\boldsymbol{H})}$ stands for either $\bar{\mathcal{L}}^{(W)}_{(f,\boldsymbol{H})}$ or $\bar{\mathcal{L}}_{(f,\boldsymbol{H})}$ and $\tilde{\mathcal{D}}_{(f,\boldsymbol{H})}$ for either $\bar{\mathcal{D}}^{(W)}_{(f,\boldsymbol{H})}$ or $\bar{\mathcal{D}}_{(f,\boldsymbol{H})}$.

Choosing $\eta := \frac{\epsilon}{64T}$, we get by Corollary 2.6.2 that

$$\begin{aligned}
\|\mathcal{L}_{(f,\boldsymbol{H})} - \mathcal{L}_{(f,\bar{\boldsymbol{H}})}\|_{1-1} &\le \|\mathcal{L}_{(f,\boldsymbol{H})} - \mathcal{L}_{(f_T,\bar{\boldsymbol{H}})}\|_{1-1} + \|\mathcal{L}_{(f_T,\bar{\boldsymbol{H}})} - \mathcal{L}_{(f,\bar{\boldsymbol{H}})}\|_{1-1} \\
&\le 8(T\eta + \|f_T - f\|_2) \le \frac{\epsilon}{8}, \\
\|\mathcal{D}_{(f,\boldsymbol{H})} - \mathcal{D}_{(f,\bar{\boldsymbol{H}})}\| &\le \|\mathcal{D}_{(f,\boldsymbol{H})} - \mathcal{D}_{(f_T,\bar{\boldsymbol{H}})}\| + \|\mathcal{D}_{(f_T,\bar{\boldsymbol{H}})} - \mathcal{D}_{(f,\bar{\boldsymbol{H}})}\| \\
&\le 8(T\eta + \|f_T - f\|_2) \le \frac{\epsilon}{8},
\end{aligned}$$

and similarly by Corollary 2.6.1 that

$$\begin{aligned}
\|\tilde{\mathcal{L}}_{(f,\boldsymbol{H})} - \tilde{\mathcal{L}}_{(f,\bar{\boldsymbol{H}})}\|_{1-1} &\le \|\tilde{\mathcal{L}}_{(f,\boldsymbol{H})} - \tilde{\mathcal{L}}_{(f_T,\bar{\boldsymbol{H}})}\|_{1-1} + \|\tilde{\mathcal{L}}_{(f_T,\bar{\boldsymbol{H}})} - \tilde{\mathcal{L}}_{(f,\bar{\boldsymbol{H}})}\|_{1-1} \le \frac{\epsilon}{8}, \\
\|\tilde{\mathcal{D}}_{(f,\boldsymbol{H})} - \tilde{\mathcal{D}}_{(f,\bar{\boldsymbol{H}})}\| &\le \|\tilde{\mathcal{D}}_{(f,\boldsymbol{H})} - \tilde{\mathcal{D}}_{(f_T,\bar{\boldsymbol{H}})}\| + \|\tilde{\mathcal{D}}_{(f_T,\bar{\boldsymbol{H}})} - \tilde{\mathcal{D}}_{(f,\bar{\boldsymbol{H}})}\| \le \frac{\epsilon}{8}.
\end{aligned}$$

Therefore, it suffices to show that $\|\mathcal{L}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{L}}_{(f,\bar{\boldsymbol{H}})}\|_{1-1}, \|\mathcal{D}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{D}}_{(f,\bar{\boldsymbol{H}})}\| \le \frac{3\epsilon}{4}$. Let $\tilde{\gamma}$ be either γ or γ_W matching the definition of $\tilde{\mathcal{L}}$, and $\tilde{\mathcal{D}}$. We apply Lemma 2.8.1 with $\star = \cdot\boldsymbol{\rho}\cdot$ and $\star = \cdot$, showing that the difference between the discrete and continuous generators $\|\mathcal{L}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{L}}_{(f,\bar{\boldsymbol{H}})}\|_{1-1}$, $\|\mathcal{D}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{D}}_{(f,\bar{\boldsymbol{H}})}\|$ can be bounded by

$$2\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon} + 1\right)^2 \left| \int_{-\infty}^{\infty} \gamma(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega - \sum_{\bar{\omega}\in S^{\lceil N\rfloor}_{\omega_0}} \tilde{\gamma}(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}_T(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega}) \right|, \tag{2.63}$$

since the number of Bohr frequencies for the discretized Hamiltonian $\bar{\boldsymbol{H}}$ satisfies $\left|B(\bar{\boldsymbol{H}})\right| \le \frac{4\|\bar{\boldsymbol{H}}\|}{\eta} + 1 \le \frac{4\|\boldsymbol{H}\|}{\eta} + 1 = \frac{256\|\boldsymbol{H}\|T}{\epsilon} + 1$.

We further bound $\|\mathcal{L}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{L}}_{(f,\bar{\boldsymbol{H}})}\|_{1-1}$ in three steps according to the following triangle inequality:

$$\left|\int_{-\infty}^{\infty} \gamma(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega - \sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N\rfloor}} \tilde{\gamma}(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}(\bar{t})\mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}(\bar{t})\mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega})\right|$$

$$\leq \left|\int_{-\infty}^{\infty} \gamma(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega - \int_{-\infty}^{\infty} \gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega\right| \tag{2.64}$$

$$+\left|\int_{-\infty}^{\infty} \gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')\mathrm{d}\omega - \sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N\rfloor}} \gamma_W(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}(\bar{t})\mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}(\bar{t})\mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega})\right|. \tag{2.65}$$

$$+\left|\sum_{\bar{\omega}\in S_{\omega_0}^{\lceil N\rfloor}} (\gamma_W(\bar{\omega}) - \tilde{\gamma}(\bar{\omega}))\bar{\mathcal{F}}\left(\bar{f}(\bar{t})\mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)^*(\bar{\omega})\bar{\mathcal{F}}\left(\bar{f}(\bar{t})\mathrm{e}^{(\mathrm{i}\nu'\bar{t})}\right)(\bar{\omega})\right|. \tag{2.66}$$

Considering that $|\nu|, |\nu'| \leq 2\|\boldsymbol{H}\|$ and using Hölder's inequality, we can see that truncation at threshold W introduces error $\leq \frac{\epsilon}{8\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon}+1\right)^2} \leq \frac{\epsilon}{8|B(\bar{\boldsymbol{H}})|^2}$ in (2.64), inducing no more than $\frac{\epsilon}{4}$ error in (2.63). If $\tilde{\gamma} := \gamma$, and (2.62) holds, then the analogous argument shows the same bound for (2.66).

We complete our proof by showing that (2.65) is bounded by $\frac{\epsilon}{8\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon}+1\right)^2}$, by applying Lemma 2.8.2, i.e., showing that (2.55)-(2.56) are fulfilled. In Lemma 2.8.2 we set $K = 2\|\boldsymbol{H}\|$ and our accuracy goal $\epsilon \leftarrow \frac{\epsilon}{8\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon}+1\right)^2}$. To bound (2.55) we observe that

$$g(\omega) := \gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}(\omega-\nu')$$

is $(C+2D)\|\hat{f}\|_\infty^2$-Lipschitz continuous:

$$\begin{aligned}
&|g(\omega+\delta) - g(\omega)| \\
&\leq \left|\gamma_W(\omega+\delta)\hat{f}^*(\omega+\delta-\nu)\hat{f}_T(\omega+\delta-\nu') - \gamma_W(\omega)\hat{f}^*(\omega+\delta-\nu)\hat{f}_T(\omega+\delta-\nu')\right| \\
&+ \left|\gamma_W(\omega)\hat{f}^*(\omega+\delta-\nu)\hat{f}_T(\omega+\delta-\nu') - \gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}_T(\omega+\delta-\nu')\right| \\
&+ \left|\gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}_T(\omega+\delta-\nu') - \gamma_W(\omega)\hat{f}^*(\omega-\nu)\hat{f}_T(\omega-\nu')\right| \\
\leq &(C+2D)\|\hat{f}\|_\infty^2 \cdot \delta,
\end{aligned}$$

therefore by (2.61) we can upper bound the left-hand side of (2.55) as

$$W\omega_0(C+2D)\|\hat{f}\|_\infty^2 \leq \frac{\epsilon}{16\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon}+1\right)^2}.$$

Now observe that due to the $(L \cdot \|f_T\|_\infty)$-Lipschitz continuity of f_T we have

$$\sum_{\bar{t}\in S_{t_0}^{\lceil N\rfloor}} t_0 \|f_T(t) - f_T(\bar{t})\|_{[\bar{t},\bar{t}+t_0)} \le \sum_{\bar{t}\in S_{t_0}^{\lceil N\rfloor}\cap[-T/2,T/2)} t_0^2(L \cdot \|f_T\|_\infty) = Tt_0(L \cdot \|f_T\|_\infty) \le \delta.$$

Repeating the above argument replacing $\gamma(\omega)$ by $\sqrt{\gamma(\omega)\gamma(-\omega)}$ proves the same bound for $\|\mathcal{D}_{(f,\bar{\boldsymbol{H}})} - \tilde{\mathcal{D}}_{(f,\bar{\boldsymbol{H}})}\|$. ∎

Lemma 2.8.3. *Both functions $\gamma_G(\omega) := (e^{4\beta\omega}+1)^{-1}$ and $\gamma_M(\omega) := \min(1, e^{-\beta\omega})$ are β-Lipschitz continuous for all $\beta \ge 0$. Moreover, $\sqrt{\gamma_G(\omega)\gamma_G(-\omega)}$ and $\sqrt{\gamma_M(\omega)\gamma_M(-\omega)}$ are $\frac{\beta}{2}$-Lipschitz continuous for all $\beta \ge 0$.*

Proof. A simple calculation shows that the absolute value of the derivative of $(e^{4\beta\omega}+1)^{-1}$ is largest at 0, where it is β, therefore it is β-Lipschitz continuous. Similarly, since $\min(1, e^{-\beta\omega})$ is continuous, and the absolute value of the (right) derivative of $\min(1, e^{-\beta\omega})$ is bounded by β for every $\omega \in \mathbb{R}$, it is also β-Lipschitz continuous. Similar elementary calculation shows the $\frac{\beta}{2}$-Lipschitz continuity of $\sqrt{\gamma_G(\omega)\gamma_G(-\omega)}$ and $\sqrt{\gamma_M(\omega)\gamma_M(-\omega)}$. ∎

Note that since we apply our generic bound Proposition 2.8.1 to the following nice functions, we get rather loose estimates of N, which are certainly off by polynomial factors from the tight values. However, since algorithmically, we only pay (poly-)logarithmic cost in N, the looseness of our bounds probably only results in constant overheads.

Corollary 2.8.1 (Discretization error of "finite-time" Davies generators)**.** *Let $f(t) = \sqrt{\frac{1}{T}}\mathbb{1}(t \in [-T/2, T/2))$ and $\epsilon > 0$. Assuming the normalization condition (2.10), if $\|\gamma\|_\infty \le 1$ and $\sqrt{\gamma(\omega)\gamma(-\omega)}$ are β-Lipschitz continuous (e.g., Metropolis or Glauber), $\|\gamma\|_\infty \le 1$, then setting $N = 2^n$ for $n = \Theta\left(\log\left(\frac{2|A|(\|\boldsymbol{H}\|+1)(T+1)(\beta+1)}{\epsilon}\right)\right)$ with appropriate constants ensures that*

$$\|\mathcal{L}_{(f,\boldsymbol{H})} - \bar{\mathcal{L}}_{(f,\boldsymbol{H})}\|_{1-1} \le \epsilon, \qquad \|\mathcal{D}_{(f,\boldsymbol{H})} - \bar{\mathcal{D}}_{(f,\boldsymbol{H})}\| \le \epsilon.$$

where $\omega_0 = \frac{2}{T}\sqrt{\frac{2\pi}{(C+1)N}}$, $t_0 = \frac{T}{2}\sqrt{\frac{2\pi(C+1)}{N}}$.

Proof. We use Proposition 2.8.1 for each $\boldsymbol{A}^a$ setting $\epsilon' \leftarrow \frac{\epsilon}{|A|}$ to prove the claim. First observe that $f_T = f$, and $\|f\|_2 = 1 = \sum_{\bar{\omega}\in S_{t_0}^{\lceil N\rfloor}} t_0|f(\bar{t})|^2 = 1$. As $\hat{f}(w) = \mathrm{sinc}(T\omega/2)$,

and $|\operatorname{sinc}(T\omega/2)| \le 1$, and $|\frac{\omega}{\mathrm{d}\omega}\operatorname{sinc}(T\omega/2)| \le T/4$ we get that $D \le \frac{T}{4}$ and trivially $L = 0$.

Let us recall the tail bound on the discrete Fourier Transform from Proposition 2.6.6

$$\sum_{|\bar{\omega}|>m\omega_0} \left|\bar{\mathcal{F}}\left(\bar{f}(\bar{t})\cdot \mathrm{e}^{(\mathrm{i}\nu\bar{t})}\right)(\bar{\omega}-\nu)\right|^2 \le \frac{\pi}{Tm\omega_0}.$$

Similar bound holds for the continuous Fourier Transform $\hat{f}(\omega) = \sqrt{\frac{T}{2\pi}}\operatorname{sinc}(T\omega/2)$ of the uniform weight function:

$$\int_{-\infty}^{\infty} |\hat{f}(\omega)\mathbb{1}(|\omega| \ge W)|^2 \mathrm{d}\omega \le \frac{\pi}{TW},$$

so choosing $W := \Theta\left(\left(\frac{(\|\boldsymbol{H}\|+1)(T+1)}{\epsilon'}\right)^3\right) \ge 2\|\boldsymbol{H}\| + 8\pi\frac{T}{\epsilon'}\left(\frac{256\|\boldsymbol{H}\|T}{\epsilon'}+1\right)^2$ ensures the necessary tail bounds. Therefore it is easy to see that setting $N = \Theta\left(\left(|A|\frac{(\|\boldsymbol{H}\|+1)(T+1)(\beta+1)}{\epsilon}\right)^{\mathcal{O}(1)}\right)$ satisfies all requirements of Proposition 2.8.1. ∎

Corollary 2.8.2 (Discretization error of Gaussian Lindbladians and discriminant proxies)**.** *Let* $f(t) = \frac{1}{\sqrt{\sigma\sqrt{2\pi}}}e^{-\frac{t^2}{4\sigma^2}}$ *and* $\epsilon > 0$*. Assuming the normalization condition* (2.10)*, if* $\|\gamma\|_\infty \le 1$ *and* $\sqrt{\gamma(\omega)\gamma(-\omega)}$ *are* β*-Lipschitz continuous (e.g., Metropolis or Glauber), , then setting* $N = 2^n$ *for* $n = \Theta\left(\log\left(\frac{2|A|(\|\boldsymbol{H}\|+1)(\sigma+1/\sigma)(\beta+1)}{\epsilon}\right)\right)$ *with appropriate constants ensures*

$$\|\mathcal{L}_{(f,\boldsymbol{H})} - \bar{\mathcal{L}}_{(f,\boldsymbol{H})}\|_{1-1} \le \epsilon, \qquad \|\mathcal{D}_{(f,\boldsymbol{H})} - \bar{\mathcal{D}}_{(f,\boldsymbol{H})}\| \le \epsilon,$$

where $\omega_0 = \frac{1}{\sigma}\sqrt{\frac{2\pi}{N}}$, $t_0 = \sigma\sqrt{\frac{2\pi}{N}}$.

Proof. We use Proposition 2.8.1 for each $\boldsymbol{A}^a$ setting $\epsilon' \leftarrow \frac{\epsilon}{|A|}$ to prove the claim. First observe that $\|f\|_2 = 1$ and since $t_0 = \mathcal{O}(\sigma)$ we have

$$\sum_{\bar{\omega}\in S_{t_0}^{\lceil N\rfloor}} t_0|f(\bar{t})|^2 = \Theta(1). \tag{2.67}$$

As $\hat{f}(w) = \sqrt{\sigma\sqrt{\frac{2}{\pi}}}e^{-\sigma^2\omega^2}$, we also get $|\frac{\omega}{\mathrm{d}\omega}\hat{f}(w)|/\|\hat{f}\|_\infty \le \sigma$, so $D \le \sigma$. Since $\|f\|_\infty = \|f_T\|_\infty$ and $|\frac{\omega}{\mathrm{d}t}f(t)|/\|f\|_\infty \le \frac{1}{2\sigma}$, we get $L \le \frac{1}{2\sigma}$.

Standard Gaussian tail bound tells us that

$$\|f - f_T\|_2^2 \le \frac{2}{T\sqrt{2\pi}}e^{-\frac{T^2}{2\sigma^2}},$$

from which it follows that

$$\sum_{\bar{\omega}\in S_{t_0}^{\lceil N\rfloor}} t_0|f(\bar{t})-f_T(\bar{t})|^2 \le \left(\frac{2}{T\sqrt{2\pi}}+\frac{2t_0}{\sigma\sqrt{2\pi}}\right)e^{-\frac{T^2}{2\sigma^2}},$$

implying that it suffices to choose $T=\Theta\Big(\sigma\sqrt{\log(1/\epsilon')}+1\Big)$.

By (2.67) and Proposition 2.6.7 the Fourier-transformed tail satisfies

$$\sqrt{\sum_{\substack{\bar{\omega}\in S_{\omega_0}^{\lceil N\rfloor}\\ |\bar{\omega}|\ge W}}\left|\hat{\tilde{f}}(\bar{\omega})\right|^2} \le O\left(\frac{1}{\sqrt{N\omega_0\sigma}}\mathrm{e}^{-N^2\omega_0^2\sigma^2/2}+\frac{1}{\sqrt{Nt_0/\sigma}}\mathrm{e}^{-N^2t_0^2/16\sigma^2}+\frac{1}{\sqrt{W\sigma}}\mathrm{e}^{-W^2\sigma^2}\right),$$

so it suffices to choose $N=\Omega\Big(\Big(\sqrt{\log(1/\epsilon')}+1\Big)\Big)$ and $W=\Theta\Big(\frac{1}{\sigma}\sqrt{\log(1/\epsilon')}+1+2\|\boldsymbol{H}\|\Big)$.

Therefore, it is easy to see that setting $N=\Theta\left(\left(|A|\frac{(\|\boldsymbol{H}\|+1)(\sigma+1/\sigma)(\beta+1)}{\epsilon}\right)^{\mathcal{O}(1)}\right)$ satisfies all requirements of Proposition 2.8.1. ■

2.9 Appendix:Implications for Lindbladians from system-bath interaction

Our algorithmic constructions are closely related to their physical origins, and the analytic framework conversely sheds light on the thermalization of open quantum systems. Under physical assumptions, one can microscopically derive a Lindbladian for a system coupled weakly to a bath (See, e.g., [152]). Among many candidates [56, 58, 139, 150, 170], we mainly focus on the *Coarsed Grained Master Equation* [121, 134] that enjoys transparent nonasymptotic error bounds and nicely connects to our algorithmic construction. Recall

$$\mathcal{L}_{(CGME)}[\boldsymbol{\rho}] := -\mathrm{i}[\boldsymbol{H}_{LS},\boldsymbol{\rho}]+\sum_{a\in A}\int_{-\infty}^{\infty}\gamma(\omega)\left(\hat{\boldsymbol{A}}^a(\omega)\boldsymbol{\rho}\hat{\boldsymbol{A}}^a(\omega)^\dagger-\frac{1}{2}\{\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega),\boldsymbol{\rho}\}\right)\mathrm{d}\omega \tag{2.68}$$

$$=: \mathcal{L}_{uni}[\boldsymbol{\rho}]+\mathcal{L}_{diss}[\boldsymbol{\rho}]$$

and the correlation function $\gamma(\omega)$ satisfying the symmetry $\gamma(\omega)/\gamma(-\omega)=\mathrm{e}^{-\beta\omega}$. The Lamb-shift term [46]

$$\boldsymbol{H}_{LS} := \sum_{a\in A}\frac{\mathrm{i}}{2\sqrt{2\pi}T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2}\mathrm{sgn}(t_1-t_2)c(t_2-t_1)\boldsymbol{A}^{a\dagger}(t_2)\boldsymbol{A}^a(t_1)\mathrm{d}t_2\mathrm{d}t_1 \tag{2.69}$$

[46] Compared with [134, Eq.24], the factor of $\frac{1}{\sqrt{2\pi}}$ is due to our Fourier Transform convention.

depends on the inverse Fourier Transform $c(t)$ of $\gamma(\omega)$, i.e.,

$$c(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \gamma(\omega) \mathrm{e}^{\mathrm{i}\omega t} \mathrm{d}\omega.$$

Let us also impose the normalization convention

$$\|c\|_1 \leq \sqrt{2\pi} \tag{2.70}$$

which also controls the frequency domain by $\|\gamma\|_\infty \leq \frac{\|c\|_1}{\sqrt{2\pi}} \leq 1$. Nicely, the strength of the Lamb-shift term is also suitably normalized:

Lemma 2.9.1 (Norm of Lamb-shift term)**.** *For $\boldsymbol{H}_{LS}$ as in (2.69),*

$$\|\boldsymbol{H}_{LS}\| \leq \frac{\|c\|_1}{2\sqrt{2\pi}} \left\| \sum_{a \in A} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a \right\|.$$

For our normalization conventions (Eq. (2.10), Eq. (2.70)), the RHS would be $\frac{1}{2}$, which is comparable to the super-operator norm of the dissipative part (2.13).

Note that if $c(t)$ is sufficiently smooth (for example, l-Lipschitz continuous with a not too large l), and we can efficiently prepare discretized states proportional to $\sqrt{|c(t)|}$ and $\frac{c(t)}{\sqrt{|c(t)|}}$, then we can get a block-encoding of a good approximation of $\frac{\boldsymbol{H}_{LS}}{\frac{1}{\sqrt{2\pi}} \int_{-T}^{T} |c(t)| \mathrm{d}t}$ by using Hamiltonian simulation time $O(T)$. The key is to prepare states proportional to $\sqrt{|c(t_2 - t_1)|}$ and $\frac{c(t_2 - t_1)}{\sqrt{|c(t_2 - t_1)|}}$ over the domain $[-T/2, T/2] \times [-T/2, T/2]$. This can be done by first preparing a uniform superposition over discretized values of t_1 on the interval $[-T/2, T/2]$, and also a state proportional to $\sqrt{|c(t_2)|}$ on the interval $[-T, T]$. This is a product state, but then we add the first variable to the second $\tilde{t}_2 \leftarrow t_2 + t_1$ (which we implement in superposition on the register containing the discretized values of t_2). The resulting new variables $t_1, \tilde{t}_2$ restricted to the domain $[-T/2, T/2] \times [-T/2, T/2]$ have the desired amplitudes. The case of $\frac{c(t_2 - t_1)}{\sqrt{|c(t_2 - t_1)|}}$ is completely analogous. Now the block-encoding is simple: prepare a state proportional $\frac{c(\bar{t}_2 - \bar{t}_1)}{\sqrt{|c(\bar{t}_2 - \bar{t}_1)|}} |\bar{t}_1\rangle |\bar{t}_2\rangle$, apply $\mathrm{sgn}(\bar{t}_2 - \bar{t}_1) \boldsymbol{A}^a(\bar{t}_2) \boldsymbol{A}^a(\bar{t}_1)$ and finally unprepare the state proportional to $\sqrt{|c(\bar{t}_2 - \bar{t}_1)|} |\bar{t}_1\rangle |\bar{t}_2\rangle$ (run the preparation in reverse). This block-encoding ensures that we can accurately simulate the above Master Equation using Theorem 2.3.2.

Compared with our algorithmic construction, the CGME differs in the following ways. First, it contains a unitary part $\mathcal{L}_{uni}$, especially the Lamb-shift term $\boldsymbol{H}_{LS}$. This requires additional technical tools to handle, so we temporarily drop this term and

postpone its discussion to section 2.9. The second difference is less essential: instead of discrete Fourier Transforms, the Kraus operators are labeled by continuous Bohr frequencies with appropriate normalizations.

$$\hat{\boldsymbol{A}}(\omega) := \sqrt{\frac{1}{2\pi T}} \int_{-T/2}^{T/2} \mathrm{e}^{-\mathrm{i}\omega t} \boldsymbol{A}(t) \mathrm{d}t = \sqrt{\frac{1}{2\pi T}} \sum_{\nu \in B(\boldsymbol{H})} \boldsymbol{A}_\nu \frac{\mathrm{e}^{-\mathrm{i}(\omega-\nu)T/2} - \mathrm{e}^{\mathrm{i}(\omega-\nu)T/2}}{\mathrm{i}\omega}.$$

Note the Fourier Transform convention. The time scale T depends on parameters of the open system, such as the bath correlation function and the coupling strength [134]; in our error bounds, we will keep T as a tunable abstract parameter.

Still, using a similar argument for analyzing the algorithm, we control the fixed point error for the Lindbladian. In fact, the bounds and the presentation simplify as discretization errors vanish. The main idea is to introduce the secular approximation (using notations in section 4.15)

$$\mathcal{L}_{sec} := \sum_{a \in A} \int_{-\infty}^{\infty} \gamma(\omega) \left(\hat{\boldsymbol{S}}^a(\omega)[\cdot] \hat{\boldsymbol{S}}^a(\omega)^\dagger - \frac{1}{2} \{ \hat{\boldsymbol{S}}^a(\omega)^\dagger \hat{\boldsymbol{S}}^a(\omega), \cdot \} \right) \mathrm{d}\omega. \quad (2.71)$$

Theorem 2.9.1 (Fixed point of the dissipative part)**.** *The dissipative part of the CGME Lindbladian* (2.68) *(satisfying normalization and symmetry conditions* (2.10),(2.12),(2.70)*), has an approximate Gibbs fixed point*

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}_{diss}) - \boldsymbol{\rho}_\beta\|_1 \leq O\left(\sqrt{\frac{\beta}{T}} t_{mix}(\mathcal{L}_{diss}) \right).$$

Since the proof structure is analogous, we present the altogether bounds and derive them in the following sections.

Proof. Telescope for the fixed points

$$\begin{aligned} \|\boldsymbol{\rho}_{fix}(\mathcal{L}_{diss}) - \boldsymbol{\rho}_\beta\|_1 &\leq \|\boldsymbol{\rho}_{fix}(\mathcal{L}_{diss}) - \boldsymbol{\rho}_{fix}(\mathcal{L}_{sec})\|_1 + \|\boldsymbol{\rho}_{fix}(\mathcal{L}_{sec}) - \boldsymbol{\rho}_\beta\|_1 \\ &\leq O\left(\left(\|\mathcal{L} - \mathcal{L}_{sec}\|_{1-1} + \left\| \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger \right\|_{2-2} \right) t_{mix}(\mathcal{L}_{diss}) \right) \\ &\leq O\left(\left(\frac{1}{\sqrt{\mu T}} + \beta \sqrt{\frac{\mu}{T}} \right) t_{mix}(\mathcal{L}_{diss}) \right). \end{aligned}$$

The second inequality uses identical arguments as (2.20) from the proof of Theorem 2.1.3. The third inequality plugs in bounds for the secular approximation (the

continuous case is the limit of the discrete case (Lemma 2.6.2)) and approximate detailed balance (Proposition 2.9.1). Optimize the free parameter $\mu = \frac{1}{\beta}$ to conclude the proof. ∎

The mixing time can be bounded by the gap (Proposition 2.2.3) if needed, but for conceptual simplicity, we stuck to the mixing time for the main presentation.

Bounds for approximate detailed balance

Here, we show approximate detailed balance for the Lindbladian of interest $\mathcal{L}_{sec}$.

Proposition 2.9.1 (Approximate detailed balance)**.** *Suppose the secular approximation for $\mathcal{L}_{(CGME)}$ (2.71) is truncated at energy μ. Then,*

$$\frac{1}{2}\left\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^{\dagger}\right\|_{2-2} \leq \mathcal{O}\left(\beta\sqrt{\frac{\mu}{T}}\right).$$

Proof. We simply telescope by inserting the algorithmically constructed discriminant $\mathcal{D}_{sec}$ (Corollary 2.7.1)

$$\begin{aligned}\left\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^{\dagger}\right\|_{2-2} &\leq \|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}) - \mathcal{D}_{sec}\|_{2-2} + \|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^{\dagger}\|_{2-2} \\ &\leq 2\|\mathcal{D}_{sec} - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})\|_{2-2}.\end{aligned}$$

The last inequality uses that $\mathcal{D}_{sec} = \mathcal{D}_{sec}^{\dagger}$ and that $\|\mathcal{A}\|_{2-2} = \|\mathcal{A}^{\dagger}\|_{2-2}$. Note that the algorithmic discussion considered discrete energy labels, so we have to take a continuum limit for the bilinear sum[47]

$$\sum_{\bar{\omega}\in S_{\omega_0}} \gamma(\bar{\omega})\hat{\boldsymbol{S}}^a(\bar{\omega})^{\dagger}\hat{\boldsymbol{S}}^a(\bar{\omega}) \to \int_{-\infty}^{\infty} \gamma(\omega)\hat{\boldsymbol{S}}^{a\dagger}(\omega)\hat{\boldsymbol{S}}^a(\omega)\mathrm{d}\omega.$$

Indeed, the bound does not depend on the discretization scale. ∎

Effects of the Lamb-shift term

In this section, we include the unitary part of the CGME generator. The resulting bounds now depend on two mixing times, and we do not have a desirable conversion between the two mixing times. Still, one can upper bound both via the spectral gap of the Hermitian part of the dissipative part $\mathcal{H}_{diss}$.

[47] Formally speaking, the correctness in this limit can be derived by using our discrete results and taking their limit as in 2.8.1.

Theorem 2.9.2 (Fixed point of CGME)**.** *For the full CGME generator* (2.68)*, which satisfy the symmetry and normalization conditions* (2.10),(2.11),(2.12),(2.70)*, we have that*

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}_{CGME}) - \boldsymbol{\rho}_\beta\|_1 \leq O\left(\sqrt{\frac{\beta}{T}}(t_{mix}(\mathcal{L}_{CGME}) + t_{mix}(\mathcal{L}_{diss}))\right).$$

We have already calculated the errors for the dissipative part; here we only study errors for the Lamb-shift term as in the following sections and combine them at section 2.9. The strategy is similar; we discretize the Hamiltonian, truncate the operator via the secular approximation, and then argue that the resulting operator nearly commutes with the Gibbs state

$$\boldsymbol{H}_{LS} \approx \boldsymbol{H}_{LS,sec} \approx \sqrt{\boldsymbol{\rho}}\boldsymbol{H}_{LS,sec}\sqrt{\boldsymbol{\rho}}^{-1}.$$

Formally, we introduce the intermediate constructs

$$\mathcal{L}_{sec} := \mathcal{L}_{uni,sec} + \mathcal{L}_{diss,sec} \quad \text{and} \quad \mathcal{L}' := \sqrt{\boldsymbol{\rho}}(\mathcal{L}_{diss,sec} - \mathcal{L}_{uni,sec})[\sqrt{\boldsymbol{\rho}}^{-1} \cdot \sqrt{\boldsymbol{\rho}}^{-1}]\sqrt{\boldsymbol{\rho}},$$

$$\text{where} \quad \mathcal{L}_{uni,sec} := -\mathrm{i}[\boldsymbol{H} + \boldsymbol{H}_{LS,sec}, \cdot].$$

Note that in $\mathcal{L}'$ we had to manually flip the sign of the coherent part to ensure $\mathcal{L}_{sec} \approx \mathcal{L}'$. One may add any coherent term $-\mathrm{i}[\boldsymbol{V}, \cdot]$ as long as $[\boldsymbol{V}, \boldsymbol{\rho}] = 0$ (most notably the Hamiltonian $\boldsymbol{H}$); it would not contribute to the error bounds.

Secular approximation for the Lamb-shift

Rewrite the integral by change-of-variable $s := t_2 - t_1$ and apply the secular approximation to the inner integral

$$\boldsymbol{H}_{LS} := \frac{\mathrm{i}}{2T}\int_{-T}^{T} \mathrm{sgn}(-s)C(s)\left(\int_{\max(-T/2,-T/2-s)}^{\min(T/2-s,T/2)} \left(\sum_{a\in A} \boldsymbol{A}^{a\dagger}(s)\boldsymbol{A}^a\right)(t_1)\mathrm{d}t_1\right)\mathrm{d}s$$

The secular approximation in this context differs from the one we used for the bilinear expressions (Lemma 2.6.3). We define and analyze the error as follows, inspired by [40].

Lemma 2.9.2 (Secular approximation for time average)**.** *Consider an operator $\boldsymbol{A}$ and a Hermitian operator $\boldsymbol{H}$. Then, for any unitarily invariant norm $\|\cdot\|_*$ and times t_1, t_2, there exists a secular approximated operator $\boldsymbol{S}_\mu$ such that*

$$\langle\psi_j|\boldsymbol{S}_\mu|\psi_i\rangle = 0 \quad \textit{if} \quad |E_i - E_j| \leq \mu$$

and

$$\left\| \int_{t_1}^{t_2} \mathrm{e}^{\mathrm{i}\boldsymbol{H}s} \boldsymbol{A} \mathrm{e}^{-\mathrm{i}\boldsymbol{H}s} \mathrm{d}s - \boldsymbol{S}_\mu \right\|_* \leq O\left(\|\boldsymbol{A}\|_* \frac{1 + \log(\frac{1}{\mu |t_2 - t_1|})}{\mu} \right).$$

See section 2.9 for the proof. Intuitively, the time average $\int_{t_1}^{t_2} \boldsymbol{A}(s)\mathrm{d}s$ weakens the off-diagonal entries (in the $\boldsymbol{H}$ eigenbasis) with a large Bohr frequency. Dropping them incurs an error depending on the truncation value μ.

Applying the secular approximation for the Lamb-shift Hamiltonian yields the following bound.

Corollary 2.9.1 (Secular approximation for the Lamb-shift term)**.** *In the setting of Theorem 2.9.2, the exists a Hermitian operator $\boldsymbol{H}_{LS,sec}$ such that*

$$\left\langle \psi_j \middle| \boldsymbol{H}_{LS,sec} \middle| \psi_i \right\rangle = 0 \quad \textit{if} \quad \left| E_i - E_j \right| \leq \mu$$

and

$$\|\boldsymbol{H}_{LS} - \boldsymbol{H}_{LS,sec}\| = \tilde{O}\left(\frac{\|\sum_{a \in A} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a\|}{\mu T} \int_{-T}^{T} |C(s)| \mathrm{d}s \right) = \tilde{O}\left(\frac{1}{\mu T} \right).$$

Proof. Apply secular approximation to the inner integral (which depends on s) to obtain

$$\boldsymbol{H}_{LS,sec} := \frac{\mathrm{i}}{2T} \int_{-T}^{T} \mathrm{sgn}(-s) C(s) \boldsymbol{S}_\mu^{(s)} \mathrm{d}s$$

and calculate

$$\|\boldsymbol{H}_{LS} - \boldsymbol{H}_{LS,sec}\| \leq \frac{1}{2T} \int_{-T}^{T} |C(s)| \left\| \int_{\max(-T/2, -T/2-s)}^{\min(T/2-s, T/2)} \left(\sum_{a \in A} \boldsymbol{A}^{a\dagger}(s) \boldsymbol{A}^a \right)(t_1) \mathrm{d}t_1 - \boldsymbol{S}_\mu^{(s)} \right\| \mathrm{d}s.$$

Use the secular approximation (Lemma 2.9.2) for the integral over t_1 to conclude the proof. ∎

Approximate detailed balance

Thirdly, we also control the error for approximate detailed balance.

Lemma 2.9.3 (Apprximate detailed balance for the unitary part)**.** *In the setting of Theorem 2.9.2, if $\beta\mu \leq 1$,*

$$\left\| \mathcal{L}_{uni}^\dagger + \sqrt{\boldsymbol{\rho}} \mathcal{L}_{uni} [\sqrt{\boldsymbol{\rho}}^{-1} \cdot \sqrt{\boldsymbol{\rho}}^{-1}] \sqrt{\boldsymbol{\rho}} \right\|_{2-2} \leq \tilde{O}\left(\beta\mu \int_{-T}^{T} |C(s)| \, \mathrm{d}s \right) = \tilde{O}(\beta\mu).$$

We present the superoperator form to feed into our existing fixed-point analysis. Still, the calculation essentially reduces to the operator norm. We will need the following proposition, whose proof is reminiscent of the arguments (section 2.7) analyzing the $\delta\mathcal{R}$ part of Lemma 2.7.3.

Proposition 2.9.2. *Suppose an operator $\boldsymbol{A}$ satisfies*

$$\langle E_i|\boldsymbol{A}|E_j\rangle = 0 \quad \textit{whenever} \quad |E_i - E_j| \leq \mu$$

and $\beta\mu \leq 1$, then

$$\|\boldsymbol{A} - \sqrt{\boldsymbol{\rho}}\boldsymbol{A}\sqrt{\boldsymbol{\rho}}^{-1}\| = O(\|\boldsymbol{A}\|\beta\mu).$$

Proof. Consider nearby energy projectors at energy resolution μ.

$$\boldsymbol{I} = \sum_{a\in\mathbb{Z}} \boldsymbol{P}_{a\mu} \quad \text{where} \quad \boldsymbol{P}_{a\mu} := \sum_{(a+\frac{1}{2})\mu > E \geq (a-\frac{1}{2})\mu} \boldsymbol{P}_E.$$

Then, the matrix $\boldsymbol{A}$ is tri-block-diagonal $\boldsymbol{A} = \boldsymbol{U} + \boldsymbol{L} + \boldsymbol{D}$ with blocks labeled by integer multiples of μ. For the lower-diagonal-blocks $\boldsymbol{L} = \sum_a \boldsymbol{P}_{(a+1)\mu}\boldsymbol{A}\boldsymbol{P}_{a\mu}$, we evaluate the commutator for each term

$$\boldsymbol{P}_{(a+1)\mu}\boldsymbol{A}\boldsymbol{P}_{a\mu} - \sqrt{\boldsymbol{\rho}}\boldsymbol{P}_{(a+1)\mu}\boldsymbol{A}\boldsymbol{P}_{a\mu}\sqrt{\boldsymbol{\rho}}^{-1} = \boldsymbol{P}_{(a+1)\mu}\boldsymbol{A}\boldsymbol{P}_{a\mu} - \mathrm{e}^{-\beta\boldsymbol{H}'/2}\boldsymbol{P}_{(a+1)\mu}\boldsymbol{A}\boldsymbol{P}_{a\mu}\mathrm{e}^{\beta\boldsymbol{H}'/2},$$

where

$$\boldsymbol{H}' = \boldsymbol{H} - (a+\frac{1}{2})\boldsymbol{I} \quad \text{such that} \quad \|\boldsymbol{H}'\| \leq \mu.$$

Therefore,

$$\begin{aligned}
\|\boldsymbol{L} - \sqrt{\boldsymbol{\rho}}\boldsymbol{L}\sqrt{\boldsymbol{\rho}}^{-1}\| &\leq \max_a \|\boldsymbol{P}_{(a+1)\mu}\boldsymbol{A}\boldsymbol{P}_{a\mu} - \sqrt{\boldsymbol{\rho}}\boldsymbol{P}_{(a+1)\mu}\boldsymbol{A}\boldsymbol{P}_{a\mu}\sqrt{\boldsymbol{\rho}}^{-1}\| \\
&\qquad\qquad \text{(By Lemma 2.7.1)} \\
&= O(\beta\mu). \qquad \text{(By } |\mathrm{e}^x - 1| \leq 2|x| \text{ for } |x| \leq 1)
\end{aligned}$$

The bounds on $\boldsymbol{D}$ and $\boldsymbol{U}$ are analogous. ∎

Proof of Lemma 2.9.3. Expand

$$\begin{aligned}
&\mathcal{L}_{uni}^{\dagger}[\boldsymbol{A}] + \sqrt{\boldsymbol{\rho}}\mathcal{L}_{uni}[\sqrt{\boldsymbol{\rho}}^{-1}\boldsymbol{A}\sqrt{\boldsymbol{\rho}}^{-1}]\sqrt{\boldsymbol{\rho}} \\
&= \mathrm{i}\boldsymbol{H}_{LS,sec}\boldsymbol{A} - \mathrm{i}\boldsymbol{A}\boldsymbol{H}_{LS,sec} - \mathrm{i}\sqrt{\boldsymbol{\rho}}\boldsymbol{H}_{LS,sec}\sqrt{\boldsymbol{\rho}}^{-1}\boldsymbol{A} + \mathrm{i}\boldsymbol{A}\sqrt{\boldsymbol{\rho}}^{-1}\boldsymbol{H}_{LS,sec}\sqrt{\boldsymbol{\rho}} \\
&= \mathrm{i}\Big(\boldsymbol{H}_{LS,sec} - \sqrt{\boldsymbol{\rho}}\boldsymbol{H}_{LS,sec}\sqrt{\boldsymbol{\rho}}^{-1}\Big)\boldsymbol{A} - \mathrm{i}\boldsymbol{A}\Big(\boldsymbol{H}_{LS,sec} - \sqrt{\boldsymbol{\rho}}^{-1}\boldsymbol{H}_{LS,sec}\sqrt{\boldsymbol{\rho}}\Big).
\end{aligned}$$

The Hamiltonian term H disappears because it commutes with the Gibbs state. Now, Holder's inequality reduces the superoperator norm to the operator norm, which can be controlled by Proposition 2.9.2

$$\left\|\boldsymbol{H}_{LS,sec} - \sqrt{\boldsymbol{\rho}}\boldsymbol{H}_{LS,sec}\sqrt{\boldsymbol{\rho}}^{-1}\right\| \leq O(\|\boldsymbol{H}_{LS,sec}\|\beta\mu) = O\big((\|\boldsymbol{H}_{LS}\| + \|\boldsymbol{H}_{LS} - \boldsymbol{H}_{LS,sec}\|)\beta\mu\big)$$

Use Corollary 2.9.1 and Lemma 2.9.2 to conclude the proof. ∎

Altogether: Proof of fixed point correctness (Theorem 2.9.2)
We now put together the error bounds for the full CGME Lindbladian.

Proof of Theorem 2.9.2. Recall the bound on the fixed point error

$$\begin{aligned}
&\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\|_1 \\
&= \|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}_{fix}(\mathcal{L}_{sec})\|_1 + \|\boldsymbol{\rho}_{fix}(\mathcal{L}_{sec}) - \boldsymbol{\rho}_{fix}(\mathcal{L}')\|_1 \\
&\leq O\left(\|\mathcal{L} - \mathcal{L}_{sec}\|_{1-1} t_{mix}(\mathcal{L}) + \frac{\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}')\|_{2-2}}{\varsigma_{-2}(\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}))}\right) \\
&\leq O\Bigg(\|\mathcal{L} - \mathcal{L}_{sec}\|_{1-1} t_{mix}(\mathcal{L}) + \left\|\mathcal{L}_{diss} - \mathcal{L}_{diss,sec}\right\|_{1-1} t_{mix}(\mathcal{L}_{diss}) \\
&+ \left(\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec}) + \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec})^\dagger\|_{2-2} + \|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{diss,sec}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{diss,sec})^\dagger\|_{2-2}\right) t_{mix}(\mathcal{L}_{diss})\Bigg),
\end{aligned} \tag{2.72}$$

where we compare the fixed points of $\mathcal{L}_{sec}$ and $\mathcal{L}'$ by eigenvector perturbation (Proposition 2.10.2, noting that $\mathcal{L}_{sec}$ contains an eigenvalue zero as it generates a CPTP map; $\mathcal{L}'$ has the Gibbs state as its fixed point, which has eigenvalue zero). The third inequality bound the singular value by the mixing time: apply Fan-Hoffman [21, Proposition III.5.1] and use perturbation bounds for sorted singular values

$$\begin{aligned}
2\varsigma_{-2}(\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})) &\geq \varsigma_{-2}\left(\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec}) + \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{sec})^\dagger\right) \\
&\geq \lambda_2\left(\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{diss,sec}) + \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{diss,sec})^\dagger\right) - \left\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec}) + \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec})^\dagger\right\|_{2-2} \\
&\geq \Omega\left(\frac{1}{t_{mix}(\mathcal{L}_{diss,sec})}\right) - \left\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec}) + \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec})^\dagger\right\|_{2-2} \\
&\geq \Omega\left(\frac{1}{t_{mix}(\mathcal{L}_{diss})}\right) - \left\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec}) + \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec})^\dagger\right\|_{2-2}
\end{aligned} \tag{2.73}$$

The rest are analogous to (2.20), except that we have to manually include the term $\left\|\mathcal{L}_{diss} - \mathcal{L}_{diss,sec}\right\|_{1-1} t_{mix}(\mathcal{L}_{diss})$ to ensure Eqn. (2.73) holds.

By secular approximation with truncation energy μ and μ' (which will be set to different values to minimize the error bounds),

$$\|\boldsymbol{H}_{LS} - \boldsymbol{H}_{LS,sec}\| = \tilde{O}(\frac{1}{\mu T}) \quad \text{and} \quad \|\mathcal{L}_{diss} - \mathcal{L}_{diss,sec}\|_{1-1} = O(\frac{1}{\sqrt{\mu' T}})$$

which combines to

$$\begin{aligned}\|\mathcal{L} - \mathcal{L}_{sec}\|_{1-1} &= \left\|\mathrm{i}[\boldsymbol{H}_{LS}, \cdot] + \mathcal{L}_{diss} - \mathrm{i}[\boldsymbol{H}_{LS,sec}, \cdot] - \mathcal{L}_{diss,sec}\right\|_{1-1} \\ &= \tilde{O}\left(\frac{1}{T\mu} + \frac{1}{\sqrt{\mu' T}}\right).\end{aligned}$$

The second inequality reduces the superoperator norm $\|\cdot\|_{1-1}$ to operator norm by $\|\boldsymbol{A}\boldsymbol{\rho}\|_1 \leq \|\boldsymbol{A}\|\|\boldsymbol{\rho}\|_1$. Next, we combine the approximate detailed balance-type errors from the Lamb-shift term (Lemma 2.9.3) and the dissipative term (Proposition 2.9.1)

$$\|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec}) + \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{uni,sec})^\dagger\|_{2-2} + \|\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{diss,sec}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L}_{diss,sec})^\dagger\|_{2-2}$$

$$= \tilde{O}\left(\beta\mu + \beta\sqrt{\frac{\mu'}{T}}\right).$$

Altogether, choose $\mu = \sqrt{\frac{1}{\beta T}} \leq \frac{1}{\beta}$ and $\mu' = \frac{1}{\beta}$ so that

$$(2.72) \leq \tilde{O}\left(\max(t_{mix}(\mathcal{L}), t_{mix}(\mathcal{L}_{diss})) \cdot \sqrt{\frac{\beta}{T}}\right),$$

which concludes the proof.

■

Proof for secular approximation for time average (Lemma 2.9.2)

Intuitively, we want to truncate the Bohr frequency far from zero. Unfortunately, the sharp truncation from section 4.15 does not seem to work here because the truncation error is related to the 1-norm $\|f\|_1$ (instead of 2-norm $\|f\|_2$). The 1-norm is more delicate to handle, forcing us to *smoothly* truncate the tail and explicitly evaluate the Fourier Transform in the time domain. Pictorially, the time domain function becomes a smeared version of the sharp window function $\mathbb{1}(|t| \leq T)$ where the discontinuity is smoothed out due to convolution with a smooth bump function.

Proof of Lemma 2.9.2. Without loss of generality, we can conjugate with time evolution to shift the integral so that $t_2 = -t_1 = T/2$. Let $f(t) = \mathbb{1}(|t| \leq T/2)$, then

$$\frac{1}{\sqrt{2\pi}} \int_{-T/2}^{T/2} \boldsymbol{A}(s)\mathrm{d}s = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s)\boldsymbol{A}(s)\mathrm{d}s = \sum_{\nu} \boldsymbol{A}_\nu \hat{f}(-\nu)$$

with $\hat{f}(\omega) = \frac{\mathrm{e}^{-\mathrm{i}\omega T/2} - \mathrm{e}^{\mathrm{i}\omega T/2}}{\sqrt{2\pi}\omega}$. Let us truncate the frequency domain function

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (b * f)(s)\boldsymbol{A}(s)\mathrm{d}s = \sum_{\nu} \boldsymbol{A}_\nu \hat{b} \cdot \hat{f}(-\nu) =: \boldsymbol{S}_\mu$$

by multiplying with a carefully chosen smooth bump function

$$\hat{b}(x) := \begin{cases} 0 & \text{if} \quad |x| \geq \mu \\ 1 & \text{if} \quad x = 0 \\ \leq 1 & \text{else} \end{cases}.$$

Then, for any unitarily invariant norm,

$$\begin{aligned} &\left\| \int_{-T/2}^{T/2} \boldsymbol{A}(s)\mathrm{d}s - \boldsymbol{S}_\mu \right\|_* \\ &\leq \frac{1}{\sqrt{2\pi}} \|f - b * f\|_1 \\ &= \frac{1}{\sqrt{2\pi}} \left(\int_{W_\epsilon^c} (f - b * f)(t)\mathrm{d}t + \int_{W_\epsilon} (f - b * f)(t)\mathrm{d}t \right) \\ &\leq O\left(\frac{1}{\mu} + \frac{1 + \log(\frac{1}{\mu T})}{\mu} \right). \end{aligned} \tag{2.74}$$

The second equality separately evaluates the integral around ϵ-balls near $\pm T$

$$W_\epsilon := [\frac{T}{2} - \epsilon, \frac{T}{2} + \epsilon] \cup [-\frac{T}{2} - \epsilon, -\frac{T}{2} + \epsilon] \quad \text{for} \quad \epsilon = \frac{1}{\mu}.$$

For each $t \in W_\epsilon^c$, the convolution is point-wise close to the original value

$$\int_{-\infty}^{\infty} f(t-s)b(s)\mathrm{d}s - f(t) = f(t)\left(\int_{-\infty}^{\infty} b(s)\mathrm{d}s - 1 \right) + r(\mu|t - T|)$$

up to an error $r(x)$ falling super-polynomially with $|x|$. Thus, the integral over W_ϵ^c (2.74) is then bounded by

$$2 \int_{|t-T| \geq \mu} |r(\mu|t - T|)|\mathrm{d}t = \frac{4}{\mu} \int_1^{\infty} |r(x)|\mathrm{d}x = O(\frac{1}{\mu}).$$

The main error arises from the sharp edge at $\pm T$; we invoke general norm bounds $\|f\|_\infty = 1$ and

$$\begin{aligned}\|b * f\|_\infty &\le \frac{1}{\sqrt{2\pi}}\|\hat{b}\cdot\hat{f}\|_1 \\ &\le \frac{1}{\sqrt{2\pi}}\left(\int_{-1/T}^{1/T}\left|\hat{f}(\omega)\right|\mathrm{d}\omega + \left(\int_{1/T}^{\mu} + \int_{-\mu}^{-1/T}\right)\left|\hat{f}(\omega)\right|\mathrm{d}\omega\right) \quad (\text{By } \|\hat{b}\| \le 1\ ) \\ &= O\left(1 + \log\left(\frac{1}{\mu T}\right)\right) \qquad\qquad (\text{By } \left|\hat{f}(\omega)\right| \le \min\left(\tfrac{T}{\sqrt{2\pi}}, \tfrac{1}{\sqrt{2\pi}\omega}\right)\ )\end{aligned}$$

and integrate over W_ϵ to obtain the bound. ■

2.10 Appendix:Spectral bounds and mixing times

In this section, we present missing proofs for lemmas and propositions. While some arguments are standard and included merely for completeness, controlling the spectrum of nearly Hermitian matrices requires a substantial linear algebraic argument. We begin with eigenvalue and eigenvector perturbation theory (section 2.10), which is crucial for establishing mixing time bounds (section 2.10) and the correctness of fixed points (section 2.10).

Perturbation bounds for eigenvalues and eigenvectors

In this section, we present some useful bounds for eigenvalue and eigenvector perturbation.

Proposition 2.10.1 (Bauer-Fike Theorem with multiplicity, cf. [21, Theorem VI.3.3 & Problem VI.8.6])**.** *Perturb a normal matrix $\boldsymbol{N}$ by an arbitrary matrix $\boldsymbol{A}$. Then, the spectrum of $\boldsymbol{N}$ and $\boldsymbol{N}+\boldsymbol{A}$ are $\|\boldsymbol{A}\|$-close to each other:*

$$\begin{aligned}&Spec(\boldsymbol{N}+\boldsymbol{A}) \subset \cup_{s\in Spec(\boldsymbol{N})} D(s, \|\boldsymbol{A}\|), \qquad \text{and} \\ &\quad Spec(\boldsymbol{N}) \subset \cup_{s\in Spec(\boldsymbol{N}+\boldsymbol{A})} D(s, \|\boldsymbol{A}\|), \quad \text{where} \quad D(s,\epsilon) = \{z\in\mathbb{C}\colon |z-s|\le\epsilon\}.\end{aligned}$$

Moreover, the connected components of $\cup_{s\in Spec(\boldsymbol{N})} D(s, \|\boldsymbol{A}\|)$ contain an equal number of eigeinvalues of $\boldsymbol{N}$ and $\boldsymbol{N}+\boldsymbol{A}$ when counted with algebraic multiplicity.

Proof. The first half of the statement is the Bauer-Fike Theorem [21, Theorem VI.3.3]. To study the number of eigenvalues per connected component, we consider an interpolation path

$$\boldsymbol{X}(t) = \boldsymbol{N} + t\boldsymbol{A} \quad \text{for} \quad 0 \le t \le 1.$$

The Bauer-Fike Theorem applied to $\boldsymbol{N} + t\boldsymbol{A}$ implies for every $0 \leq t \leq 1$ that

$$\mathrm{Spec}(\boldsymbol{X}(t)) \quad \subset \quad \cup_{s\in\mathrm{Spec}(\boldsymbol{N})} D(s, t\|\boldsymbol{A}\|) \quad \subset \quad \cup_{s\in\mathrm{Spec}(\boldsymbol{N})} D(s, \|\boldsymbol{A}\|).$$

Then, by continuity of eigenvalues along the path, cf. [21, Corollary VI.1.6], no eigenvalues enter or exit the connected components of $\cup_{s\in\mathrm{Spec}(\boldsymbol{N})} D(s, \|\boldsymbol{A}\|)$, therefore their number (counted with algebraic multiplicity) is the same for $\boldsymbol{X}(0) = \boldsymbol{N}$ and $\boldsymbol{X}(1) = \boldsymbol{N} + \boldsymbol{A}$. ∎

Corollary 2.10.1 (Eigenvalue perturbation for discriminants)**.** *If $\mathcal{D}$ is Hermitian and has norm bounded by $\|\mathcal{D}\| \leq 1$, and $\mathcal{D}'$ has a right eigenvector $|\psi\rangle$ with eigenvalue 1, then the top eigenvalue of $\mathcal{D}$ satisfies*

$$|\lambda_1(\mathcal{D}) - 1| \leq \|\mathcal{D} - \mathcal{D}'\|.$$

Eigenvector perturbation bounds

Intuitively, perturbing a matrix yields small changes in eigenvectors with well-isolated eigenvalues — we prove this below rigorously under suitable but quite general conditions using a simple linear algebraic argument.

Proposition 2.10.2 (Eigenvector perturbation)**.** *Perturb a matrix $\boldsymbol{M}$ by another matrix $\boldsymbol{A}$. Let $|v\rangle$ be a normalized right eigenvector $\boldsymbol{M}|v\rangle = \lambda|v\rangle$, and λ' an eigenvalue of $\boldsymbol{M} + \boldsymbol{A}$. Then the corresponding right eigenvector $(\boldsymbol{M} + \boldsymbol{A})|v'\rangle = \lambda'|v'\rangle$ can be normalized such that*

$$\langle v|v\rangle = \langle v'|v'\rangle = 1 \quad \textit{and} \quad \||v'\rangle - |v\rangle\| \leq \frac{2\sqrt{2}(\|\boldsymbol{A}\| + |\lambda' - \lambda|)}{\varsigma_{-2}(\boldsymbol{M} - \lambda\boldsymbol{I})},$$

where $\varsigma_{-2}(\cdot)$ denotes the second-smallest singular value (with multiplicity). Due to Fan-Hoffman [21, Proposition III.5.1] the singular value $\varsigma_{-2}(\boldsymbol{M} - \lambda\boldsymbol{I}) \geq -\lambda_2\left(\frac{\boldsymbol{M}+\boldsymbol{M}^\dagger}{2} - \mathrm{Re}(\lambda)\boldsymbol{I}\right)$ can be bounded in terms of the Hermitian part.

Proof. We can assume without loss of generality that $|v'\rangle \propto |v\rangle + \epsilon|v^\perp\rangle$ for $\epsilon \in [0, \infty)$ and $\langle v^\perp|v^\perp\rangle = 1$, yielding

$$\begin{aligned}(\boldsymbol{M} + \boldsymbol{A})(|v\rangle + \epsilon\big|v^\perp\big\rangle) &= \lambda'\boldsymbol{I}(|v\rangle + \epsilon\big|v^\perp\big\rangle)\\ \Updownarrow&\\ \boldsymbol{A}(|v\rangle + \epsilon\big|v^\perp\big\rangle) + (\boldsymbol{M} - \lambda'\boldsymbol{I})|v\rangle &= \epsilon(\lambda'\boldsymbol{I} - \boldsymbol{M})\big|v^\perp\big\rangle.\end{aligned}$$

Taking the norms above on both sides and defining $\kappa := |\lambda' - \lambda|$ we get

$$\begin{aligned}\epsilon = \frac{\|\boldsymbol{A}(|v\rangle + \epsilon|v^\perp\rangle) + (\lambda - \lambda')|v\rangle\|}{\|(\boldsymbol{M} - \lambda'\boldsymbol{I})|v^\perp\rangle\|} &\leq \frac{(1+\epsilon)\|\boldsymbol{A}\| + \kappa}{\varsigma_{-2}(\boldsymbol{M} - \lambda\boldsymbol{I}) - \kappa} \\ \Downarrow \\ \epsilon &\leq \frac{\|\boldsymbol{A}\| + \kappa}{\varsigma_{-2}(\boldsymbol{M} - \lambda\boldsymbol{I}) - \|\boldsymbol{A}\| - \kappa} \\ &= \frac{\|\boldsymbol{A}\| + \kappa}{\varsigma_{-2}(\boldsymbol{M} - \lambda\boldsymbol{I})} \cdot \frac{1}{1 - \frac{\|\boldsymbol{A}\|+\kappa}{\varsigma_{-2}(\boldsymbol{M}-\lambda\boldsymbol{I})}}. \end{aligned} \tag{2.75}$$

The last inequality is a rearrangement. The first inequality uses the triangle inequality for the numerator, and for the denominator that $\|(\boldsymbol{M} - \lambda'\boldsymbol{I})|v^\perp\rangle\| \geq \|(\boldsymbol{M} - \lambda\boldsymbol{I})|v^\perp\rangle\| - \|(\lambda' - \lambda)\boldsymbol{I}|v^\perp\rangle\| = \|(\boldsymbol{M} - \lambda\boldsymbol{I})|v^\perp\rangle\| - \kappa$ and

$$\|(\boldsymbol{M} - \lambda\boldsymbol{I})\big|v^\perp\big\rangle\| \geq \varsigma_{-2}(\boldsymbol{M} - \lambda\boldsymbol{I}).$$

We conclude by setting the appropriate normalization $|v'\rangle = \frac{1}{\sqrt{1+\epsilon^2}}(|v\rangle + \epsilon|v^\perp\rangle)$ and utilizing the above bound (2.75):

$$\begin{aligned}\||v\rangle - |v'\rangle\| &= \frac{1}{\sqrt{1+\epsilon^2}}\|\sqrt{1+\epsilon^2}|v\rangle - (|v\rangle + \epsilon\big|v^\perp\big\rangle)\| \\ &= \frac{\sqrt{(\sqrt{1+\epsilon^2} - 1)^2 + \epsilon^2}}{\sqrt{1+\epsilon^2}} \leq \sqrt{2}\epsilon \leq \min\left(\sqrt{2}, \frac{2\sqrt{2}(\|\boldsymbol{A}\| + \kappa)}{\varsigma_{-2}(\boldsymbol{M} - \lambda'\boldsymbol{I})}\right).\end{aligned}$$

The first inequality uses $\sqrt{1+\epsilon^2} - 1 \leq \epsilon$. The last inequality uses that the bound is vacuous at $\||v\rangle - |v'\rangle\| \leq \sqrt{2}$ and combines (2.75) with the elementary estimate $\forall x \in [0, 1/2]: \sqrt{2}\frac{x}{1-x} \leq 2\sqrt{2}x$ after substituting $x \leftarrow \frac{\|\boldsymbol{A}\|+\kappa}{\varsigma_{-2}(\boldsymbol{M}-\lambda\boldsymbol{I})}$. ∎

When we apply the above perturbation bound for a Hermitian $\boldsymbol{M}$, we can set λ to be the top eigenvalue such that $\varsigma_{-2}(\boldsymbol{M} - \lambda\boldsymbol{I})$ will be the gap $\lambda_{gap}(\boldsymbol{M}) = \lambda_1(\boldsymbol{M}) - \lambda_2(\boldsymbol{M})$. In that case, Proposition 2.10.1 guarantees the existence of a nearby eigenvalue λ' of $\boldsymbol{M} + \boldsymbol{A}$ such that $|\lambda' - \lambda| \leq \|\boldsymbol{A}\|$. When we handle the Lamb-shift term, we need to consider a nonHermitian $\boldsymbol{M}$. There, bounding eigenvalue perturbation is not generally obvious, so we will simply assume that there is a nearby eigenvalue λ' of $\boldsymbol{M} + \boldsymbol{A}$.

Approximate detailed balance implies approximately correct fixed point

When detailed balance holds approximately for $\boldsymbol{\rho}$, we still expect the fixed point to be *approximately* $\boldsymbol{\rho}$; we provide a proof of this in this section, which relies on the matrix perturbation results (section 2.10). Recall that in section 2.2, we defined the

Hermitian and anti-Hermitian parts (under similarity transformation) as follows:

$$\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}) = \boldsymbol{\rho}^{-1/4}\mathcal{L}[\boldsymbol{\rho}^{1/4} \cdot \boldsymbol{\rho}^{1/4}]\boldsymbol{\rho}^{-1/4} = \mathcal{H} + \mathcal{A},$$
$$\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger = \boldsymbol{\rho}^{1/4}\mathcal{L}^\dagger[\boldsymbol{\rho}^{-1/4} \cdot \boldsymbol{\rho}^{-1/4}]\boldsymbol{\rho}^{1/4} = \mathcal{H} - \mathcal{A}.$$

Observe that $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger[\sqrt{\boldsymbol{\rho}}] = 0$, but it needs not be the case for $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})$, which we care about.

We recall some facts: every Lindbladian satisfies that $\mathrm{Spec}(\mathcal{L}) \subseteq \{z \in \mathbb{C} \colon \mathrm{Re}(z) \leq 0\}$ [181, Proposition 6.1]. As $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger[\sqrt{\boldsymbol{\rho}}] = 0$ we also have that 0 is an element of the spectrums $\mathrm{Spec}(\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger)$, $\mathrm{Spec}(\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}))$. Since $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})$ is defined by a similarity transformation we have that $\mathrm{Spec}(\mathcal{L}) = \mathrm{Spec}(\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}))$ and due to Proposition 2.10.1 this implies

$$|\lambda_1(\mathcal{H})| \leq \|\mathcal{A}\|_{2\text{–}2}. \tag{2.76}$$

Proposition 2.10.3 (Fixed point accuracy)**.** *Suppose a Lindbladian $\mathcal{L}$ satisfies the ϵ-approximate $\boldsymbol{\rho}$-detailed balance condition. If $\lambda_{gap}(\mathcal{H}) > 2\epsilon$, then there is a unique state as its fixed point $\boldsymbol{\rho}_{fix}(\mathcal{L}) > 0$ and its deviation from $\boldsymbol{\rho}$ is bounded by*

$$\left\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\right\|_1 \leq \frac{14\epsilon}{\lambda_{gap}(\mathcal{H})}.$$

The RHS indicates that the fixed point accuracy may deteriorate if the map has a large anti-Hermitian component or if the gap closes.

Proof. Every CPTP map has at least a stationary state [181, Theorem 6.11], and thus there is a fixed point $\boldsymbol{\rho}_{fix}(\mathcal{L}) \geq 0$ of unit trace. The condition $\lambda_{gap}(\mathcal{H}) > 2\|\mathcal{A}\|_{2\text{–}2}$ translates to $\lambda_2(\mathcal{H}) < -\|\mathcal{A}\|_{2\text{–}2}$ implying that 0 has algebraic multiplicity 1 in $\mathrm{Spec}(\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}))$ due to Proposition 2.10.1, which then proves the uniqueness of the fixed point.

By our eigenvector perturbation bound (Proposition 2.10.2), we get that there is a matrix $\boldsymbol{R}$ of unit Frobenius norm in the kernel of $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})$ such that

$$\|\sqrt{\boldsymbol{\rho}} - \boldsymbol{R}\|_2 \leq \frac{4\sqrt{2}\|\mathcal{A}\|_{2\text{–}2}}{-\lambda_2(\mathcal{H})},$$

where we used that the Frobenius norm of a matrix is equal to the Euclidean norm of its vectorization $\|\boldsymbol{A}\|_2 = \||\boldsymbol{A}\rangle\|$. This in turn means that $\boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4}$ is in the kernel

of $\mathcal{L}$, thus $\boldsymbol{\rho}_{fix}(\mathcal{L}) = \boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4}/\mathrm{Tr}(\boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4})$, moreover

$$\begin{aligned}\|\boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4} - \boldsymbol{\rho}_{fix}(\mathcal{L})\|_1 &= \left|\mathrm{Tr}(\boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4}) - 1\right| \|\boldsymbol{\rho}_{fix}(\mathcal{L})\|_1 \\ &= \left|\mathrm{Tr}(\boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4} - \boldsymbol{\rho})\right| \leq \|\boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4} - \boldsymbol{\rho}\|_1,\end{aligned}$$

where in the last step we used the trace-norm inequality $|\mathrm{Tr}(\boldsymbol{A})| \leq \|\boldsymbol{A}\|_1$. We can further bound

$$\|\boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4} - \boldsymbol{\rho}\|_1 = \|\boldsymbol{\rho}^{1/4}(\boldsymbol{R} - \sqrt{\boldsymbol{\rho}})\boldsymbol{\rho}^{1/4}\|_1 \leq \|\boldsymbol{\rho}^{1/4}\|_4^2 \cdot \left\|\boldsymbol{R} - \sqrt{\boldsymbol{\rho}}\right\|_2 = \|\boldsymbol{R} - \sqrt{\boldsymbol{\rho}}\|_2,$$

where we used Hölder's inequality $\|\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}\|_1 \leq \|\boldsymbol{B}\|_4^2\|\boldsymbol{A}\|_2$. Combining the above three inequalities, we get

$$\begin{aligned}\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\|_1 &\leq \|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4}\|_1 + \|\boldsymbol{\rho}^{1/4}\boldsymbol{R}\boldsymbol{\rho}^{1/4} - \boldsymbol{\rho}\|_1 \\ &\leq 2\|\boldsymbol{R} - \sqrt{\boldsymbol{\rho}}\|_2 \leq \frac{8\sqrt{2}\|\mathcal{A}\|_{2\to 2}}{-\lambda_2(\mathcal{H})}.\end{aligned}$$

Finally, we convert λ_2 to λ_{gap} in the above bound. Due to (2.76) we have $-\lambda_2(\mathcal{H}) = \lambda_{gap}(\mathcal{H}) - \lambda_1(\mathcal{H}) \geq \lambda_{gap}(\mathcal{H}) - \|\mathcal{A}\|_{2\to 2}$ so we can further bound the above by $8\sqrt{2}\|\mathcal{A}\|_{2\to 2}/(\lambda_{gap}(\mathcal{H}) - \|\mathcal{A}\|_{2\to 2})$. But this bound is vacuous at $\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\|_1 \leq 2$, i.e., when $\lambda_{gap}(\mathcal{H}) < (4\sqrt{2}+1)\|\mathcal{A}\|_{2\to 2}$. If $\lambda_{gap}(\mathcal{H}) \geq (4\sqrt{2}+1)\|\mathcal{A}\|_{2\to 2}$, then $\lambda_{gap}(\mathcal{H}) - \|\mathcal{A}\|_{2\to 2} \geq \frac{4\sqrt{2}}{4\sqrt{2}+1}\lambda_{gap}(\mathcal{H})$, yielding

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\|_1 \leq \frac{(8\sqrt{2}+2)\|\mathcal{A}\|_{2\to 2}}{\lambda_{gap}(\mathcal{H})}. \qquad \blacksquare$$

Corollary 2.2.1 (Fixed point accuracy)**.** *If a Lindbladian $\mathcal{L}$ satisfies the ϵ-approximate $\boldsymbol{\rho}$-detailed balance condition, then its fixed point $\boldsymbol{\rho}_{fix}(\mathcal{L})$ deviates from $\boldsymbol{\rho}$ by at most*

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\|_1 \leq 20 t_{mix}(\mathcal{L})\epsilon.$$

Proof. In the proof of Proposition 2.10.3 we got the ultimate bound by converting λ_2 to λ_{gap} by the observation $-\lambda_2(\mathcal{H}) \geq \lambda_{gap}(\mathcal{H}) - \|\mathcal{A}\|_{2\to 2}$. Due to Proposition 2.10.1 the same bound $-\lambda_2(\mathcal{H}) \geq \lambda_{\mathrm{Re}(gap)}(\mathcal{L}) - \|\mathcal{A}\|_{2\to 2}$ also holds for $\lambda_{\mathrm{Re}(gap)}(\mathcal{L})$ (c.f. Proposition 2.10.5) since 0 is an eigenvalue of $\mathcal{L}$. Hence the conversion combined with Proposition 2.10.5 shows

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}) - \boldsymbol{\rho}\|_1 \leq \frac{(8\sqrt{2}+2)\|\mathcal{A}\|_{2\to 2}}{\lambda_{\mathrm{Re}(gap)}(\mathcal{L})} \leq \frac{(8\sqrt{2}+2)\|\mathcal{A}\|_{2\to 2}}{\ln(2)} t_{mix}(\mathcal{L}). \qquad \blacksquare$$

Perturbation bounds for Lindbladians regarding gaps and mixing times

This section provides proof for scattered statements circling spectral gaps and mixing time. Most results are standard, except maybe the most technical result (Lemma 2.10.2).

Lemma 2.2.1 (Fixed point difference)**.** *For any two Lindbladians $\mathcal{L}_1$ and $\mathcal{L}_2$, the difference of their fixed points (in the Schrödinger picture) is bounded by*

$$\|\boldsymbol{\rho}_{fix}(\mathcal{L}_1) - \boldsymbol{\rho}_{fix}(\mathcal{L}_2)\|_1 \leq 4\|\mathcal{L}_1 - \mathcal{L}_2\|_{1-1} \cdot t_{mix}(\mathcal{L}_1).$$

Proof. We begin by recalling Duhamel's identity. We use its integral form derived in, e.g., [81, Eq. (40)]:

$$\mathrm{e}^{\boldsymbol{A}} - \mathrm{e}^{\boldsymbol{B}} = \int_0^1 \mathrm{e}^{s\boldsymbol{A}}(\boldsymbol{A} - \boldsymbol{B})\mathrm{e}^{(1-s)\boldsymbol{B}}\mathrm{d}s.$$

We apply the above identity with $\boldsymbol{A} \leftarrow t\mathcal{L}_1$, $\boldsymbol{B} \leftarrow t\mathcal{L}_2$ and take the $1-1$ operator norm on both sides

$$\begin{aligned}\|\mathrm{e}^{t\mathcal{L}_1} - \mathrm{e}^{t\mathcal{L}_2}\|_{1-1} &\leq \int_0^1 \|\mathrm{e}^{st\mathcal{L}_1}(t\mathcal{L}_1 - t\mathcal{L}_2)\mathrm{e}^{(1-s)t\mathcal{L}_2}\|_{1-1}\mathrm{d}s \\ &\leq t\int_0^1 \|\mathrm{e}^{st\mathcal{L}_1}\|_{1-1}\|t\mathcal{L}_1 - t\mathcal{L}_2\|_{1-1}\|\mathrm{e}^{(1-s)t\mathcal{L}_2}\|_{1-1}\mathrm{d}s = t\|\mathcal{L}_1 - \mathcal{L}_2\|_{1-1}.\end{aligned} \tag{2.77}$$

The last equality uses that $\mathrm{e}^{x\mathcal{L}}$ is a CPTP map and so $\|\mathrm{e}^{x\mathcal{L}}\|_{1-1} = 1$ [181, Theorem 8.16].[48]

Now, let $\boldsymbol{\rho}_1 := \boldsymbol{\rho}_{fix}(\mathcal{L}_1)$, $\boldsymbol{\rho}_2 := \boldsymbol{\rho}_{fix}(\mathcal{L}_2)$, and $t_{mix} := t_{mix}(\mathcal{L}_1^\dagger)$. Then

$$\begin{aligned}\|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\|_1 &= \left\|\mathrm{e}^{\mathcal{L}_1 \cdot t_{mix}}[\boldsymbol{\rho}_1] - \mathrm{e}^{\mathcal{L}_2 t_{mix}}[\boldsymbol{\rho}_2]\right\|_1 \\ &\leq \left\|\mathrm{e}^{\mathcal{L}_1 \cdot t_{mix}}[\boldsymbol{\rho}_1] - \mathrm{e}^{\mathcal{L}_1 t_{mix}}[\boldsymbol{\rho}_2]\right\|_1 + \left\|\mathrm{e}^{\mathcal{L}_1 t_{mix}}[\boldsymbol{\rho}_2] - \mathrm{e}^{\mathcal{L}_2 t_{mix}}[\boldsymbol{\rho}_2]\right\|_1 \\ &\leq \frac{1}{2} \cdot \|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\|_1 + 2t_{mix}\|\mathcal{L}_1 - \mathcal{L}_2\|_{1-1}.\end{aligned} \tag{2.78}$$

The second inequality follows from (2.77) and that $\|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\|_1 \leq 2$. Rearrange (2.78) to conclude the proof. ∎

Proposition 2.10.4 (Mixing time difference)**.** *The mixing times of two Lindbladians $\mathcal{L}_1$, $\mathcal{L}_2$ are related by*

$$t_{mix}(\mathcal{L}_2) \leq t_{mix}(\mathcal{L}_1)\left\lceil \frac{\ln(1/2)}{\ln(1/2 + t_{mix}(\mathcal{L}_1)\|\mathcal{L}_1 - \mathcal{L}_2\|_{1-1})} \right\rceil \quad \textit{if} \quad t_{mix}(\mathcal{L}_1)\|\mathcal{L}_1 - \mathcal{L}_2\|_{1-1} < \frac{1}{2}.$$

[48] Here, by the $1-1$ norm, we mean the $1-1$ norm of the operators restricted to the subspace of Hermitian matrices.

Proof. Let $t_{mix} := t_{mix}(\mathcal{L}_1)$ and $\boldsymbol{R}$ be a traceless Hermitian matrix with $\|\boldsymbol{R}\|_1 = 1$ maximizing $\|e^{\mathcal{L}_2 t_{mix}}[\boldsymbol{R}]\|_1$, then

$$\begin{aligned}\|e^{\mathcal{L}_2 t_{mix}}[\boldsymbol{R}]\|_1 &\leq \|e^{\mathcal{L}_1 t_{mix}}[\boldsymbol{R}]\|_1 + \|e^{\mathcal{L}_2 t_{mix}}[\boldsymbol{R}] - e^{\mathcal{L}_1 t_{mix}}[\boldsymbol{R}]\|_1 \\ &\leq \frac{1}{2}\|\boldsymbol{R}\|_1 + t_{mix}\|\mathcal{L}_1 - \mathcal{L}_2\|_{1-1}\|\boldsymbol{R}\|_1 = \left(\frac{1}{2} + t_{mix}\|\mathcal{L}_1 - \mathcal{L}_2\|_{1-1}\right)\|\boldsymbol{R}\|_1.\end{aligned}$$

The second inequality follows from (2.77). Set $\boldsymbol{R} = \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2$, rearrange, and take the logarithm to conclude. ∎

Relating the mixing time to the spectral gap using exact detailed balance

The mixing time of a general Lindbladian may be difficult to analyze. Fortunately, many handy bounds exist, especially circling the spectral gap when detailed balance holds.

Proposition 2.2.2 (Mixing time from spectral gap [97])**.** *If a Lindbladian $\mathcal{L}$ satisfies $\boldsymbol{\rho}$-detailed balance, then*

$$t_{mix}(\mathcal{L}) \leq \frac{\ln(2\|\boldsymbol{\rho}^{-1/2}\|)}{\lambda_{gap}(\mathcal{L})},$$

where $\lambda_{gap}(\mathcal{L})$ is the eigenvalue gap of the Lindbladian, and the mixing time t_{mix} is the smallest time for which

$$\left\|e^{\mathcal{L} t_{mix}}[\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2]\right\|_1 \leq \frac{1}{2}\|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\|_1 \quad \textit{for any states} \quad \boldsymbol{\rho}_1, \boldsymbol{\rho}_2.$$

Proof. Write $\boldsymbol{R} = \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2$, then

$$\begin{aligned}\left\|e^{\mathcal{L}t}[\boldsymbol{R}]\right\|_1 &= \left\|\boldsymbol{\rho}^{1/4} e^{\mathcal{D}t}[\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}]\boldsymbol{\rho}^{1/4}\right\|_1 \\ &\leq \left\|\boldsymbol{\rho}^{1/4}\right\|_4 \cdot \left\|e^{\mathcal{D}t}[\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}]\right\|_2 \cdot \left\|\boldsymbol{\rho}^{1/4}\right\|_4 \\ &\leq e^{-\lambda_{gap}(\mathcal{H})t}\left\|\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}\right\|_2 \\ &\leq e^{-\lambda_{gap}(\mathcal{H})t}\|\boldsymbol{\rho}^{-1/4}\|^2\|\boldsymbol{R}\|_2 \\ &\leq e^{-\lambda_{gap}(\mathcal{H})t}\|\boldsymbol{\rho}^{-1/4}\|^2\|\boldsymbol{R}\|_1 \\ &= e^{-\lambda_{gap}(\mathcal{L}^\dagger)t}\|\boldsymbol{\rho}^{-1/2}\|\|\boldsymbol{R}\|_1.\end{aligned}$$

The first inequality uses Hölder's inequality. The second inequality uses the orthogonality to the leading eigenvector such that $\mathrm{Tr}[\sqrt{\boldsymbol{\rho}}\cdot\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}] = \mathrm{Tr}[\boldsymbol{R}] = 0$. Take the logarithm to conclude the proof. ∎

Relating the mixing time to the Hermitian gap (and approximate detailed balance)

Proposition 2.10.5 (Spectral gap from mixing time)**.** *For any Lindbladian $\mathcal{L}$, let $-\lambda_{\mathrm{Re}(gap)}(\mathcal{L})$ be the second largest real part in its spectrum (counted by algebraic multiplicity), then*

$$\lambda_{gap}(\mathcal{H}) + 2\|\mathcal{A}\|_{2\text{–}2} \geq \|\mathcal{A}\|_{2\text{–}2} - \lambda_2(\mathcal{H}) \geq \lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \geq \frac{\ln(2)}{t_{mix}(\mathcal{L})}.$$

Moreover, if $\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \geq 2\|\mathcal{A}\|_{2\text{–}2}$, then there is unique eigenvalue $\lambda_1(\mathcal{H}) \geq -\|\mathcal{A}\|_{2\text{–}2}$ and

$$\lambda_{gap}(\mathcal{H}) + 2\|\mathcal{A}\|_{2\text{–}2} \geq \|\mathcal{A}\|_{2\text{–}2} - \lambda_2(\mathcal{H}) \geq \lambda_{\mathrm{Re}(gap)}(\mathcal{L}). \tag{2.79}$$

Proof. We know any Lindbladian has at least a stationary state of eigenvalue 0, and each eigenvalue which has no real part has a trivial Jordan block [181, Theorem 6.11 & Proposition 6.2]. Therefore, if $\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) = 0$, then $t_{mix}(\mathcal{L}) = \infty$. If $\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) > 0$, take any eigenvalue λ such that $-\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) = \mathrm{Re}(\lambda)$, and let $\boldsymbol{R}$ be a corresponding right eigenvector of $\mathcal{L}$, which is then necessarily traceless. We can assume without loss of generality that the Hermitian part $\boldsymbol{R}_H := (\boldsymbol{R} + \boldsymbol{R}^\dagger)/2$ is nonzero (otherwise, we can just take $\boldsymbol{R} \leftarrow \mathrm{i}\boldsymbol{R}$). Since $\mathcal{L}$ is Hermiticity preserving, we get that $\boldsymbol{R}_H$ is also a right eigenvector with eigenvalue λ. For $t < \frac{\ln(2)}{\lambda_{\mathrm{Re}(gap)}(\mathcal{L})}$ we have

$$\|\mathrm{e}^{\mathcal{L}t}[\boldsymbol{R}_H]\|_1 = \left|\mathrm{e}^{\lambda t}\right| \cdot \|\boldsymbol{R}_H\|_1 > \frac{1}{2}\|\boldsymbol{R}_H\|_1$$

implying that

$$t_{mix}(\mathcal{L}) \geq \frac{\ln(2)}{\lambda_{\mathrm{Re}(gap)}(\mathcal{L})}.$$

We conclude using Bauer-Fike Proposition 2.10.1 to show $\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \leq -\lambda_2(\mathcal{H}) + \|\mathcal{A}\|_{2\text{–}2}$, combined with (2.76). ∎

What if the detailed balance condition is violated? In the worst case, the conversion from spectral gap to mixing time can be poor. However, the Lindbladians we consider are "sufficiently" detailed balanced so that essentially the same consequences hold. The following two scenarios cover our main use cases

$$\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}) = \mathcal{H} + \mathcal{A} \quad \text{where} \quad \|\mathcal{A}\|_{2\text{–}2} \ll \lambda_{gap}(\mathcal{H}), \qquad \text{(approximate detailed balance)}$$

$$\text{or} \quad \lambda_1(\mathcal{H}) \ll \lambda_{gap}(\mathcal{H}). \qquad \text{(nonperturbative } \mathcal{A}\text{)}$$

The first case should be understood as nonHermitian eigenvalue perturbation, consistent with the framework of approximate detailed balance. The second case is less intuitive, as the anti-Hermitian part can be arbitrarily large. Intriguingly, the spectral properties of the Hermitian part $\mathcal{H}$ suffice to control convergence even in the presence of a large perturbation $\mathcal{H} + \mathcal{A}$. In fact, the second case is strictly more general since $\lambda_1(\mathcal{H}) \leq \|\mathcal{A}\|_{2-2}$ (2.76), therefore we will only analyze the second scenario. Intuitively, when the Lindbladian is exactly detailed balanced $\mathcal{A} = 0$, we have that $\lambda_1(\mathcal{H}) = 0$; the $\lambda_1(\mathcal{H}) \ll \lambda_{gap}(\mathcal{H})$ condition is essentially the requirement that $\lambda_1(\mathcal{H}) \approx 0$ in spite of a large anti-hermitian component $\|\mathcal{A}\|_{2-2} \gg 0$. Before proving our result, we need a few lemmas.

Lemma 2.10.1 (Norm of matrix exponential)**.** *For a Hermitian matrix $\boldsymbol{H}$ and an anti-Hermitian matrix $\boldsymbol{B}$, we have*

$$\|\mathrm{e}^{(\boldsymbol{H}+\boldsymbol{B})t}\| \leq \mathrm{e}^{\lambda_1(\boldsymbol{H})t} \quad \textit{for each} \quad t \geq 0.$$

If $\boldsymbol{B}$ is an arbitrary matrix, we alternatively get $\|\mathrm{e}^{(\boldsymbol{H}+\boldsymbol{B})t}\| \leq \mathrm{e}^{(\lambda_1(\boldsymbol{H})+\|\boldsymbol{B}\|)t}$.

Proof. The claim follows from the Trotter representation of exponential and triangle inequality

$$\begin{aligned}\|\mathrm{e}^{(\boldsymbol{H}+\boldsymbol{B})t}\| = \|\lim_{r\to\infty}\left(\mathrm{e}^{\boldsymbol{H}/r}\mathrm{e}^{\boldsymbol{B}/r}\right)^{rt}\| = \lim_{r\to\infty}\|\left(\mathrm{e}^{\boldsymbol{H}/r}\mathrm{e}^{\boldsymbol{B}/r}\right)^{rt}\| &\leq \liminf_{r\to\infty}\left(\|\mathrm{e}^{\boldsymbol{H}/r}\|\|\mathrm{e}^{\boldsymbol{B}/r}\|\right)^{rt} \\ &= \liminf_{r\to\infty}\left(\|\mathrm{e}^{\boldsymbol{H}/r}\|\right)^{rt} = \mathrm{e}^{\lambda_1(\boldsymbol{H})t},\end{aligned}$$

where we used that $\mathrm{e}^{\boldsymbol{B}/r}$ is unitary so that $\|\mathrm{e}^{\boldsymbol{B}/r}\| = 1$. The second claim follows from isolating the Hermitian part $\frac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^\dagger)$, whose norm is bounded by $\|\boldsymbol{B}\|$. ∎

Lemma 2.10.2 (Hermitian gap controls decay)**.** *Consider a Hermitian matrix $\boldsymbol{H}$ and an anti-Hermitian $\boldsymbol{A}$. If*

$$r := \frac{\lambda_1(\boldsymbol{H})}{\lambda_{gap}(\boldsymbol{H})} \leq \frac{1}{100},$$

and $\boldsymbol{H} + \boldsymbol{A}$ has an eigenvalue 0 with left and right eigenvectors

$$\langle L|(\boldsymbol{H} + \boldsymbol{A}) = 0 \quad \textit{and} \quad (\boldsymbol{H} + \boldsymbol{A})|R\rangle = 0,$$

then

$$\left\|\mathrm{e}^{(\boldsymbol{H}+\boldsymbol{A})t}(\boldsymbol{I} - \boldsymbol{P}_0)\right\| \leq \frac{3}{2}\exp\left(\frac{\lambda_2(\boldsymbol{H})t}{2}\right) \quad \textit{where} \quad \boldsymbol{P}_0 := \frac{1}{\langle L|R\rangle}|R\rangle\langle L|.$$

Intuitively speaking, the conditions above ensure that even if the anti-Hermitian $\boldsymbol{A}$ is large, it mainly introduces "rotations" and the spectral properties of $\boldsymbol{H}$ still guarantee fast convergence to $\boldsymbol{P}_0$.

Proof. Let $|\psi_1\rangle$ be the eigenvector of $\boldsymbol{H}$ corresponding to its top eigenvalue $\lambda_1(\boldsymbol{H})$. We can assume without loss of generality that $\langle L|$, $|R\rangle$, and $|\psi_1\rangle$ have unit norm

$$\||\psi_1\rangle\| = \|\langle L|\| = \||R\rangle\| = 1,$$

and we are free to choose the phase of these vectors as the projector $\boldsymbol{P}_0$ is invariant under changing the phase of $\langle L|$, $|R\rangle$. The guiding intuition behind the proof is that if r is small, then (up to a phase)

$$|L\rangle \approx |\psi_1\rangle \approx |R\rangle.$$

To show this, take the real part of $\langle R|(\boldsymbol{H}+\boldsymbol{A})|R\rangle = 0$ to obtain[49]

$$\begin{aligned}
\langle R|\boldsymbol{H}|R\rangle = 0 &= \lambda_1(\boldsymbol{H})|\langle R|\psi_1\rangle|^2 + \sum_{i=2}\lambda_i(\boldsymbol{H})|\langle R|\psi_i\rangle|^2 \\
&\le \lambda_1|\langle R|\psi_1\rangle|^2 + \lambda_2(1-|\langle R|\psi_1\rangle|^2), \\
\text{implying}\quad 1-|\langle R|\psi_1\rangle|^2 &\le \frac{\lambda_1(\boldsymbol{H})}{\lambda_{gap}(\boldsymbol{H})} = r, \\
\text{and thus}\quad \||R\rangle - |\psi_1\rangle\| &\le 1.01\sqrt{r},
\end{aligned}$$

where in the last step, we assumed without loss of generality that the phase of $|R\rangle$ is such that $\langle R|\psi_1\rangle$ is nonnegative real, so we get $\langle R|\psi_1\rangle = \cos(\theta)$ for some $\theta \in [0, \pi/2]$. Then $\sin(\theta) \le \sqrt{r}$, and since for every $\theta \in [0, \arcsin(1/10)]$ we have $1-\cos(\theta) \le \sin(\theta)/10$, we get the norm bound $\frac{101}{100}\sin(\theta) \le \frac{101}{100}\sqrt{r}$. Similarly,

$$\langle L|(\boldsymbol{H}+\boldsymbol{A})|L\rangle = 0 \quad \text{implies} \quad \||L\rangle - |\psi_1\rangle\| \le 1.01\sqrt{r}.$$

As a direct consequence, using that $\sqrt{r} \le 1/10$, we get

$$\begin{aligned}
|\langle L|R\rangle - 1| = |\langle L|R\rangle - \langle\psi_1|\psi_1\rangle| &\le |\langle L|R\rangle - \langle L|\psi_1\rangle| + |\langle L|\psi_1\rangle - \langle\psi_1|\psi_1\rangle| \\
&\le \||R\rangle - |\psi_1\rangle\| + \||L\rangle - |\psi_1\rangle\| \le 2.02\sqrt{r} \le \frac{1}{4},
\end{aligned}$$

and similarly that

$$\left|\frac{1}{\langle L|R\rangle} - 1\right| = \left|\frac{1-\langle L|R\rangle}{\langle L|R\rangle}\right| \le \frac{4}{3}|1-\langle L|R\rangle| \le 2.73\sqrt{r}.$$

[49] Note that $\Re(\langle R|(\boldsymbol{H}+\boldsymbol{A})|R\rangle) = 0$ implies $\lambda_1(\boldsymbol{H}) \ge 0$, and therefore $r \ge 0$. This then further implies $\lambda_2(\boldsymbol{H}) \le 0$ due to $r \le 0.01$.

Therefore,

$$\|\boldsymbol{P}_0 - |\psi_1\rangle\langle\psi_1|\| \le \|\frac{|R\rangle\langle L|}{\langle L|R\rangle} - |R\rangle\langle L|\| + \||R\rangle\langle L| - |R\rangle\langle\psi_1|\| + \||R\rangle\langle\psi_1| - |\psi_1\rangle\langle\psi_1|\|$$

$$= \left|\frac{1}{\langle L|R\rangle} - 1\right| + \||L\rangle - |\psi_1\rangle\| + \||R\rangle - |\psi_1\rangle\| \le 4.75\sqrt{r}. \qquad (2.80)$$

We will use the following properties of the projector $\boldsymbol{P}_0$; one can easily see that it commutes with $(\boldsymbol{H}+\boldsymbol{A})$:

$$\boldsymbol{P}_0(\boldsymbol{H}+\boldsymbol{A}) = \frac{1}{\langle L|R\rangle}|R\rangle\langle L|(\boldsymbol{H}+\boldsymbol{A}) = 0 = (\boldsymbol{H}+\boldsymbol{A})\boldsymbol{P}_0, \qquad (2.81)$$

and since it is a projector, it satisfies the algebraic identity

$$\boldsymbol{P}_0(\boldsymbol{I}-\boldsymbol{P}_0) = 0 \quad \text{so that} \quad \mathrm{e}^{x\boldsymbol{P}_0}(\boldsymbol{I}-\boldsymbol{P}_0) = \boldsymbol{I}-\boldsymbol{P}_0 \quad \text{for any } x\in\mathbb{C}. \qquad (2.82)$$

The above properties streamline the rest of the proof of our bound

$$\begin{aligned}
\|\mathrm{e}^{(\boldsymbol{H}+\boldsymbol{A})t}(\boldsymbol{I}-\boldsymbol{P}_0)\| &= \|\mathrm{e}^{(\boldsymbol{H}+\boldsymbol{A})t}\cdot\mathrm{e}^{-\lambda_{gap}(\boldsymbol{H})\boldsymbol{P}_0 t}(\boldsymbol{I}-\boldsymbol{P}_0)\| && \text{(by (2.82))}\\
&= \|\mathrm{e}^{(\boldsymbol{H}+\boldsymbol{A}-\lambda_{gap}(\boldsymbol{H})\boldsymbol{P}_0)t}(\boldsymbol{I}-\boldsymbol{P}_0)\| && \text{(by (2.81))}\\
&\le \|\mathrm{e}^{(\boldsymbol{H}+\boldsymbol{A}-\lambda_{gap}(\boldsymbol{H})\boldsymbol{P}_0)t}\|\cdot\|\boldsymbol{I}-\boldsymbol{P}_0\|\\
&= \left\|\exp\left(\overbrace{(\boldsymbol{H}-\lambda_{gap}(\boldsymbol{H})|\psi_1\rangle\langle\psi_1|)t}^{\boldsymbol{H}':=} + \boldsymbol{A}t + \overbrace{\lambda_{gap}(\boldsymbol{H})(|\psi_1\rangle\langle\psi_1|-\boldsymbol{P}_0)t}^{\boldsymbol{B}:=}\right)\right\|\cdot\|\boldsymbol{I}-\boldsymbol{P}_0\|\\
&\le \mathrm{e}^{(\lambda_1(\boldsymbol{H}')+\|\boldsymbol{B}\|)t}(\|\boldsymbol{I}-|\psi_1\rangle\langle\psi_1|\| + \||\psi_1\rangle\langle\psi_1|-\boldsymbol{P}_0\|)\\
&\qquad\qquad \text{(by Lemma 2.10.1)}\\
&\le \exp\left(\lambda_2(\boldsymbol{H})(1-\frac{4.75\sqrt{r}}{1-r})t\right)(1+4.75\sqrt{r}). && \text{(by Eq. (2.80))}\\
&\le \frac{3}{2}\cdot\exp\left(\frac{1}{2}\lambda_2(\boldsymbol{H})t\right). && \text{(by } r\le 1/100) \quad \blacksquare
\end{aligned}$$

Now, we can specialize the above to the case of Lindbladians to obtain mixing times.

Proposition 2.2.3 (Mixing time from Hermitian gap)**.** *For any Lindbladian $\mathcal{L}$ and a full-rank state $\boldsymbol{\rho}$, suppose the self-adjoint component $\mathcal{H} = \mathcal{H}(\boldsymbol{\rho},\mathcal{L})$ satisfies*

$$\frac{\lambda_1(\mathcal{H})}{\lambda_{gap}(\mathcal{H})} \le \frac{1}{100}, \quad \textit{then} \quad t_{mix}(\mathcal{L}) \le 3\frac{\ln(3\|\boldsymbol{\rho}^{-1/2}\|)}{\lambda_{gap}(\mathcal{H})}.$$

Proof. We proceed as in the proof of Proposition 2.2.2. For any traceless Hermitian $\boldsymbol{R}$, we have

$$\begin{aligned}
\left\|\mathrm{e}^{\mathcal{L}t}[\boldsymbol{R}]\right\|_1 &= \left\|\boldsymbol{\rho}^{1/4}\mathrm{e}^{t(\mathcal{H}+\mathcal{A})}[\boldsymbol{\rho}^{-1/4}(\boldsymbol{R})\boldsymbol{\rho}^{-1/4}]\boldsymbol{\rho}^{1/4}\right\|_1 && \text{(by definition)}\\
&\le \left\|\boldsymbol{\rho}^{1/4}\right\|_4 \cdot \left\|\mathrm{e}^{t(\mathcal{H}+\mathcal{A})}[\boldsymbol{\rho}^{-1/4}(\boldsymbol{R})\boldsymbol{\rho}^{-1/4}]\right\|_2 \cdot \left\|\boldsymbol{\rho}^{1/4}\right\|_4 \\
& && \text{(by Hölder's inequality)}\\
&\le \frac{3}{2}\mathrm{e}^{\lambda_2(\boldsymbol{H})t/2}\left\|\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}\right\|_2 && \text{(by Lemma 2.10.2)}\\
&\le \frac{3}{2}\mathrm{e}^{\lambda_2(\boldsymbol{H})t/2}\|\boldsymbol{\rho}^{-1/4}\|^2\|\boldsymbol{R}\|_2 && \text{(by Hölder's inequality)}\\
&\le \frac{3}{2}\mathrm{e}^{\lambda_2(\boldsymbol{H})t/2}\|\boldsymbol{\rho}^{-1/2}\|\|\boldsymbol{R}\|_1. && \text{(since } \boldsymbol{\rho} \succ 0 \text{ and } \|\cdot\|_2 \le \|\cdot\|_1)
\end{aligned}$$

The second inequality uses Lemma 2.10.2 since $\lim_{t\to\infty} \mathrm{e}^{\mathcal{L}t}[\boldsymbol{R}] = 0$ (or in the notation of Lemma 2.10.2, $\boldsymbol{P}_0|\boldsymbol{R}\rangle = 0$). By assumption, $\lambda_2(\mathcal{H}) = \lambda_1(\mathcal{H}) - \lambda_{gap}(\mathcal{H}) \le -0.99\lambda_{gap}(\mathcal{H})$; take the logarithm to conclude the proof. ∎

2.11 Appendix:Improved incoherent Lindbladian simulation

The improved algorithm builds on the circuit in Figure 2.10, which is similar to that of our weak measurement scheme in Figure 2.3, with a key technical difference: here, a specific ancilla state $\boldsymbol{Y}_\delta|0\rangle = \sqrt{1-\delta}|0\rangle + \sqrt{\delta}|1\rangle$ "triggers" the appropriate weak measurement. Thus, the circuit is "idle" if we remove the $\boldsymbol{Y}_\delta$ gate, which is essential for our "compression" argument. However, this also makes the circuit post-selective, which could exponentially decrease the probability of success if we were to run for longer times naively. Thus, for $t > 2$ we decompose the simulation to $\lceil t/2\rceil$ equal segments, each of which can be amplified with constant (coherent) repetitions by oblivious amplitude amplification.

The main conceptual ingredient leading to the substantial improvement is *compression* [49]: instead of naively running r-repetitions of the circuit from Figure 2.10 as outlined in Figure 2.11, we *compress* the circuit such that it uses only $h \sim t\log(t/\epsilon) \ll r = \Theta(t^2/\epsilon)$ repetitions. The compression technique relies on understanding the joint initial state $\left(\left|0^{c+1}\right\rangle\boldsymbol{Y}_\delta|0\rangle\right)^{\otimes r}$ of the circuit in Figure 2.11 after the single-qubit gates. The upshot is that $\boldsymbol{Y}_\delta^{\otimes r}|0^r\rangle$ is *concentrated* on strings with Hamming weight $\le h$, thus h repetitions will suffice for the mass of the amplitudes.

Before diving into the proof, we explain the intuitive compression strategy in more detail; the proof closely follows this, but with technical changes. Let $t \le 2$ and let $\boldsymbol{C}'$ be the circuit that we get from $\boldsymbol{C}$ by removing the two single qubit gates $\boldsymbol{Y}_\delta, \boldsymbol{Y}_{\delta/4}^\dagger$

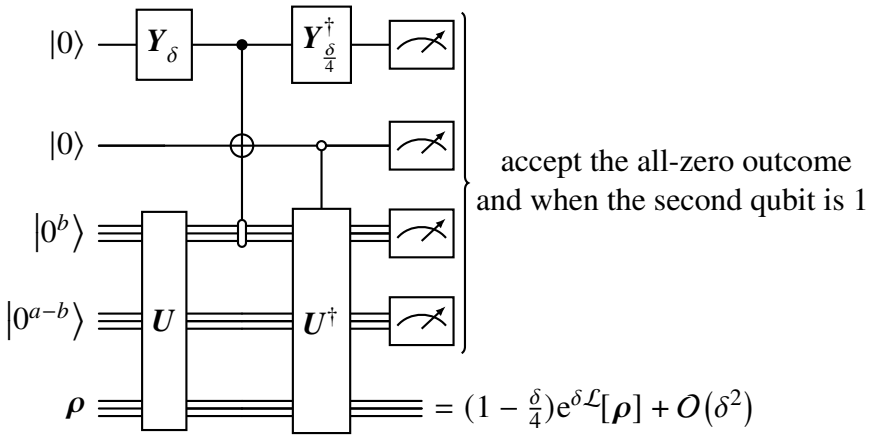


Figure 2.10: Alternative quantum circuit implementation of an approximate δ-time step via a postselective weak measurement scheme. Let $\boldsymbol{C}'$ be the circuit that we get by removing the two single qubit rotation gates $\boldsymbol{Y}_\delta, \boldsymbol{Y}^\dagger_{\delta/4}$ from the first qubit. For our compression argument it is of paramount importance that $\boldsymbol{C}' \cdot |0^{a+2}\rangle\langle 0^{a+2}| \otimes \boldsymbol{I} = |0^{a+2}\rangle\langle 0^{a+2}| \otimes \boldsymbol{I}$.

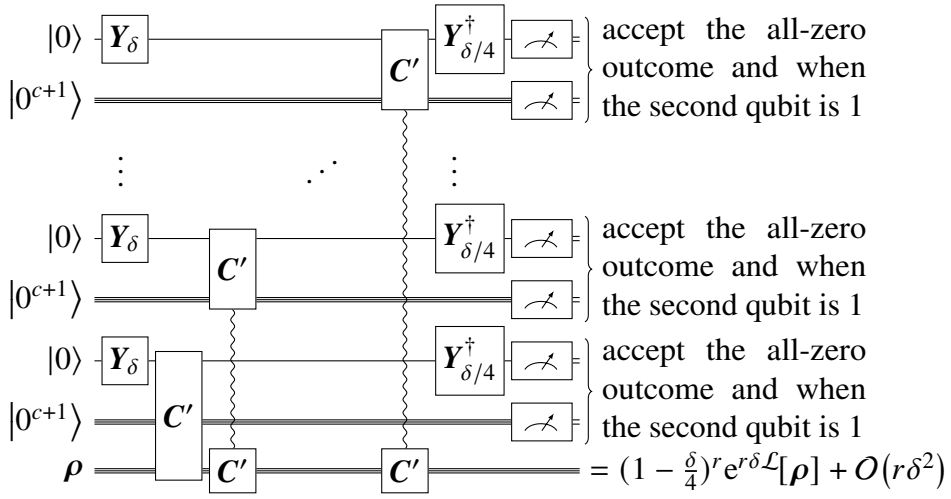


Figure 2.11: r subsequent repetitions of the circuit $\boldsymbol{C}'$ from Figure 2.10. The circuits $\boldsymbol{C}'$ act on potentially nonadjacent qubits, which is indicated by the vertical curly connection between the visually split "halves" of the affected $\boldsymbol{C}'$ circuits.

from the first qubit in Figure 2.10. Let X_i be the random variable representing the measurement outcome of a computational basis measurement of the i-th qubit in $\boldsymbol{Y}_\delta^{\otimes r}|0^r\rangle$. Then, the Chernoff bound tells us that $\Pr\left(\sum_{i=1}^r X_i > (1+y)t\right) \leq \left(\frac{e^y}{(1+y)^{1+y}}\right)^t$ so that the probability that the Hamming weight of $\left(\sqrt{1-\delta}|0\rangle + \sqrt{\delta}|1\rangle\right)^{\otimes r}$ is greater

than $h := (1+y)t$ is at most $\frac{e^{h-t}t^h}{h^h} \leq (\frac{et}{h})^h \leq (\frac{2e}{h})^h$. In particular, choosing

$$h = \Theta\left(\frac{\log(1/\epsilon)}{\log\log(1/\epsilon)}\right)$$

ensures that this probability is at most $O(\epsilon^2)$. Therefore, the initial state can be replaced by its (normalized) projection $|\phi_0\rangle$ to the subspace of Hamming-weight $\leq h$ states while inflicting an error that is bounded by

$$\|\boldsymbol{Y}_\delta^{\otimes r}|0^r\rangle - |\phi_0\rangle\| \leq \epsilon.$$

This bound on Hamming-weights translates into a reduction of applications of the circuit $\boldsymbol{C}'$. Since $|\phi_0\rangle$ is a superposition of bitstrings of Hamming-weights at most h, in all branches of the superposition all but h applications of $\boldsymbol{C}'$ can be neglected, crucially because it acts trivially when the ancilla register is in state $\left|0^{a+2}\right\rangle$:

$$\boldsymbol{C}' \cdot |0^{a+2}\rangle\langle 0^{a+2}| \otimes \boldsymbol{I} = |0^{a+2}\rangle\langle 0^{a+2}| \otimes \boldsymbol{I}. \tag{2.83}$$

Now we define a compression scheme for the r ancilla registers, each containing $(a+2)$ qubits[50] in Figure 2.11, inspired by [18, 49]. The compression scheme can represent the $\leq h$ Hamming-weight states of the ancilla registers on just $h \cdot (\log(r+1)+a+2)$ qubits (concerning the r registers to be compressed, by Hamming weight, we mean the number of registers that do not contain the state $\left|0^{a+2}\right\rangle$). Marking the register state $\left|0^{a+2}\right\rangle$ by 0 and the content of the i-th nonzero register by d_i the encoding works as follows:

$$\left(\{0,1\}^{a+2}\right)^r \ni 0^{s_1}d_1 0^{s_2}d_2 \ldots \to (s_1, s_2, \ldots) \times (d_1, d_2, \ldots) \in \{0,1,2,\ldots,r\}^h \times \left(\{0,1\}^{a+2}\right)^h. \tag{2.84}$$

The compressed representation's first "compression" register contains h blocks of $\log(r+1)$ qubits, designated to store a sequence $s \in \{0,1,2,\ldots,r\}^h$, where s_i is the number of consecutive ancilla registers containing $\left|0^{a+2}\right\rangle$ before the i-th ancilla register that is not in state 0; if i exceeds the Hamming weight, then we set $s_i = r$. The second "data" register consists of h blocks of $(a+2)$ qubits, where the i-th block represents the qubits of the i-th nonzero register of the uncompressed state; if the Hamming weight is less than i, then the block is set to 0. The property (2.83) means that the $\boldsymbol{C}'$ gates can be applied "transversally" on the second "data" register of the encoded scheme because they do not change the location of the nonzero registers of the uncompressed state.

[50] Note that the first qubit is redundant in this encoding, but we add it here for clarity of the presentation.

Theorem 2.3.2 (Compressed incoherent Lindbladian simulation algorithm)**.** *Suppose* $\boldsymbol{U}$ *is a block-encoding of the Lindblad operators of a purely irreversible Lindbladian* $\mathcal{L}$ *as in Definition 2.1.2. Let* $\epsilon \leq 1/2$*, then we can simulate the action of the superoperator* $\mathrm{e}^{t\mathcal{L}}$ *to precision* ϵ *in diamond norm using*

$$\begin{aligned} &O((c+\log((t+1)/\epsilon))\log((t+1)/\epsilon)) && \textit{(resettable) ancilla qubits,}\\ &O\left((t+1)\frac{\log((t+1)/\epsilon)}{\log\log((t+1)/\epsilon)}\right) && \textit{(controlled) uses of } \boldsymbol{U} \textit{ and } \boldsymbol{U}^\dagger,\\ \textit{and}\quad &O((t+1)(c+1)\mathrm{polylog}((t+1)/\epsilon)) && \textit{other two-qubit gates.}\end{aligned}$$

If the Lindbladian has a coherent part $-\mathrm{i}[\boldsymbol{H},\boldsymbol{\rho}]$*, and we have access to a block-encoding of* $\boldsymbol{H} = (\langle 0^c|\otimes \boldsymbol{I})\boldsymbol{V}(|0^c\rangle\otimes\boldsymbol{I})$*, then we can simulate* $\mathrm{e}^{t\mathcal{L}}$ *with* $O\left((t+1)\frac{\log((t+1)/\epsilon)}{\log\log((t+1)/\epsilon)}\right)$ *additional (controlled) uses of* $\boldsymbol{V}$ *and* $\boldsymbol{V}^\dagger$*.*

Proof of Theorem 2.3.2. We begin with analyzing the (modified) weak-measurement scheme using similar calculations to Theorem 2.3.1, and then compress it. We focus on the purely irreversible scenario and, at the end, handle the general case.

(Postselected weak-measurement.) The circuit $\boldsymbol{C}$ from Figure 2.10 on a pure input state $|\psi\rangle$ acts as

$$\begin{aligned} |0^{a+2}\rangle|\psi\rangle &\overset{(1)}{\to} (\sqrt{1-\delta}|0\rangle + \sqrt{\delta}|1\rangle)|0\rangle\boldsymbol{U}|0^c\rangle|\psi\rangle\\ &\overset{(2)}{\to} \sqrt{1-\delta}|00\rangle\boldsymbol{U}|0^c\rangle|\psi\rangle + \sqrt{\delta}|11\rangle\Big(|0^b\rangle\langle 0^b|\otimes\boldsymbol{I}\Big)\boldsymbol{U}|0^c\rangle|\psi\rangle\\ &\quad + \sqrt{\delta}|10\rangle(\boldsymbol{I}-|0^b\rangle\langle 0^b|\otimes\boldsymbol{I})\boldsymbol{U}|0^c\rangle|\psi\rangle\\ &= (\sqrt{1-\delta}|00\rangle + \sqrt{\delta}|10\rangle)\boldsymbol{U}|0^c\rangle|\psi\rangle + \sqrt{\delta}|11\rangle|0^b\rangle\underbrace{(\langle 0^b|\otimes\boldsymbol{I})\boldsymbol{U}|0^c\rangle|\psi\rangle}_{|\psi_0'\rangle:=}\\ &\quad - \sqrt{\delta}|10\rangle\Big(|0^b\rangle\langle 0^b|\otimes\boldsymbol{I}\Big)\boldsymbol{U}|0^c\rangle|\psi\rangle\\ &\overset{(3)}{\to} \left(\boldsymbol{Y}^\dagger_{\frac{\delta}{4}}\otimes\boldsymbol{I}\right)\Big((\sqrt{1-\delta}|00\rangle + \sqrt{\delta}|10\rangle)|0^c\rangle|\psi\rangle + \sqrt{\delta}|11\rangle|0^b\rangle|\psi_0'\rangle\\ &\quad - \sqrt{\delta}|10\rangle\boldsymbol{U}^\dagger\Big(|0^b\rangle\langle 0^b|\otimes\boldsymbol{I}\Big)\boldsymbol{U}|0^c\rangle|\psi\rangle\Big) \end{aligned} \tag{2.85}$$

Now, let us compute the part of $\boldsymbol{C}\left|0^{a+2}\right\rangle|\psi\rangle$ starting with $\left|0^{a+2}\right\rangle$

$$\begin{aligned}
&(\left\langle 0^{a+2}\right| \otimes \boldsymbol{I})\boldsymbol{C}\left|0^{a+2}\right\rangle|\psi\rangle \\
&= \underbrace{\left(\sqrt{\left(1-\frac{\delta}{4}\right)(1-\delta)}+\frac{\delta}{2}\right)}_{=1-\frac{\delta}{8}+O\left(\delta^2\right)=\sqrt{1-\frac{\delta}{4}}+O\left(\delta^2\right)}|\psi\rangle - \frac{\delta}{2}\underbrace{(\langle 0^c| \otimes \boldsymbol{I})\boldsymbol{U}^\dagger\left(|0^b\rangle\langle 0^b| \otimes \boldsymbol{I}\right)\boldsymbol{U}|0^c\rangle|\psi\rangle}_{\sum_{j\in J}\boldsymbol{L}_j^\dagger \boldsymbol{L}_j|\psi\rangle} \\
&= \sqrt{1-\frac{\delta}{4}}\left(\left(\boldsymbol{I}-\frac{\delta}{2}\sum_{j\in J}\boldsymbol{L}_j^\dagger \boldsymbol{L}_j\right)|\psi\rangle + O\left(\delta^2\right)\right),
\end{aligned}$$

and the part where second qubit is $|1\rangle$

$$(\boldsymbol{I} \otimes \langle 1| \otimes \boldsymbol{I})\boldsymbol{C}\left|0^{a+2}\right\rangle|\psi\rangle = \sqrt{\delta}\boldsymbol{Y}_{\frac{\delta}{4}}^\dagger|1\rangle\left|0^b\right\rangle\left|\psi_0'\right\rangle = \sqrt{\delta}\boldsymbol{Y}_{\frac{\delta}{4}}^\dagger|1\rangle\left|0^b\right\rangle\sum_{j\in J}|j\rangle\boldsymbol{L}_j|\psi\rangle.$$

Let $\widetilde{\boldsymbol{\Pi}} := \frac{1}{\sqrt{1-\frac{\delta}{4}}}\left(|0^{a+2}\rangle\langle 0^{a+2}| \otimes \boldsymbol{I} + \boldsymbol{I} \otimes |1\rangle\langle 1| \otimes \boldsymbol{I}\right)$, the above implies that

$$\begin{aligned}
&\mathrm{Tr}_{a+2}\left(\widetilde{\boldsymbol{\Pi}}C\left(|0^{a+2}\rangle\langle 0^{a+2}| \otimes |\psi\rangle\langle\psi|\right)C^\dagger\widetilde{\boldsymbol{\Pi}}\right) \\
&= \left(\boldsymbol{I}-\frac{\delta}{2}\sum_{j\in J}\boldsymbol{L}_j^\dagger \boldsymbol{L}_j\right)|\psi\rangle\langle\psi|\left(\boldsymbol{I}-\frac{\delta}{2}\sum_{j\in J}\boldsymbol{L}_j^\dagger \boldsymbol{L}_j\right) + \delta\sum_{j\in J}\boldsymbol{L}_j|\psi\rangle\langle\psi|\boldsymbol{L}_j^\dagger + O\left(\delta^2\right) \\
&= \mathrm{e}^{\delta\mathcal{L}}[|\psi\rangle\langle\psi|] + O\left(\delta^2\right).
\end{aligned}$$

Similarly to the proof of Theorem 2.3.1, it is easy to see that this implies

$$\left\|\mathrm{Tr}_{a+2}\left(\widetilde{\boldsymbol{\Pi}}C\left(|0^{a+2}\rangle\langle 0^{a+2}| \otimes [\cdot]\right)C^\dagger\widetilde{\boldsymbol{\Pi}}\right) - \mathrm{e}^{\delta\mathcal{L}}[\cdot]\right\|_\diamond = O\left(\delta^2\right).$$

Like in Theorem 2.3.1, choosing $\delta = \Theta(\epsilon/t)$ and repeating the process $r := t/\delta$ times (every time using $a+2$ fresh ancillas) yields an ϵ-accurate simulation (Figure 2.11). This gives similar circuit complexity as Theorem 2.3.1, except that the resulting postselective protocol has success probability about $(1-\frac{\epsilon}{t})^{\frac{t^2}{\epsilon}} = \exp(-\Theta(t))$. More precisely, the square of the subnormalization factor is $(1-\frac{\delta}{4})^{\frac{t}{\delta}}$, which is at least $\frac{1}{2}$ for $t \le 2$ and $\delta \in (0,1]$ since $(1-\frac{\delta}{4})^{\frac{t}{\delta}} \ge (1-\frac{\delta}{4})^{\frac{2}{\delta}} \ge (\frac{3}{4})^2 > \frac{1}{2}$.

(Compression.) While the above discussion is largely similar to Theorem 2.3.1, we now further compress to obtain the desired complexity. We analyze the case when $t \le 2$ and later show how to bootstrap the results for arbitrary large t.

Instead of running the verbose circuit of Figure 2.11, we directly prepare the compressed version of the state $|\phi_0\rangle$ using the techniques of [18]. Using the algorithm of [18], we can directly prepare the "compression" register corresponding

to $|\phi_0\rangle$, then we can initialize the "data" register by looping through all blocks and applying an $\boldsymbol{X}$ gate on the first qubit conditioned on the corresponding "compression" block containing a number less than r. Then we apply the $\boldsymbol{C}'$ gates "transversally", looping through each of the h blocks of the "data" register of the compressed state in increasing order starting from the first block.

The last and technically most challenging difficulty that we face is to evaluate the post-selection criterion in Figure 2.11. We could, of course, uncompress the state, apply the single qubit gates $(\boldsymbol{Y}_{\delta/4}^{\dagger})^{\otimes r}$ and perform the measurement literally as depicted on Figure 2.11, however it is possible to evaluate this criterion while keeping the compressed representation.[51]

We use a slightly modified variant of the compressed measurement scheme outlined in [18, Section 5] that enables us to perform a measurement of the form $\left(\boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|, \boldsymbol{I} - \boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|\right)^{\otimes r}$, where $\boldsymbol{P}_0 = \boldsymbol{Y}_{\delta/4}|0\rangle\langle 0|\boldsymbol{Y}_{\delta/4}^{\dagger}$. This compressed measurement procedure reports the result also in a compressed form by listing the (uncompressed) indices $(i_1, i_2, \ldots, i_\ell)$ where the measurement outcome is $\boldsymbol{I} - \boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|$.

The initial observation of [18] is that the measurement $\left(\left(\boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|\right)^{\otimes r}, \boldsymbol{I} - \left(\boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|\right)^{\otimes r}\right)$ can be approximately performed by using the compressed state preparation circuit approximately preparing the compressed version of the state $\left(|0^{c+1}\rangle\boldsymbol{Y}_{\delta/4}|0\rangle\right)^{\otimes r}$. Indeed, we can just run this compressed state preparation in reverse and verify that we get the all-zero state.

A similar procedure can be devised for performing the compressed measurement

$$\left(\left(\boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|\right)^{\otimes \ell}, \boldsymbol{I} - \left(\boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|\right)^{\otimes \ell}\right) \tag{2.86}$$

[51] The compression and compact verification procedure for checking whether all segments were applied successfully in the encoded scheme (and the resulting reflection operator) does not appear to work as described in [49]. The issue is manifest in their description of the reflection operator about accepted outcomes: "Therefore, the corresponding operation in the encoded representation is first applying $\mathrm{e}^{\dagger}$, then applying the reflection about the encoded state $|0^c\rangle|0^b\rangle$ on the first two registers, and last applying E."

However, this approach does not seem to work, as noted in [18] (here U_m is the analog of E in the above quote from [49]): "At first glance, one might imagine that applying U_m in place of $R^{\otimes m}$ would yield a succinct representation of the final outcome state, so measuring in the computational basis would provide the correct result. Unfortunately, this does not accurately simulate the final measurement except in the case where the all-zero string is obtained."

For completeness, we include and analyze our modified compression / verification scheme. There may be a simpler fix for the above issue [176], or alternatively, the techniques of the recent work [112] — circumventing compression — could also be compatible with our improvements [176], which would in turn probably also simplify our circuits.

for any given consecutive ℓ (uncompressed) registers $d_i, d_{i+1}, \ldots, d_{i+\ell-1}$. Towards this, observe that we can efficiently convert an encoded string $e_{1;r}$ of the form (2.84) to a new form $(e_{1;i-1}, e_{i;i+\ell-1}, e_{i+\ell;r})$ where $e_{j;k}$ is an encoding of the uncompressed block-string $u_j, u_{j+1}, \ldots, u_k$. The measurement is then performed by applying this conversion $e_{1;r} \to (e_{1;i-1}, e_{i;i+\ell-1}, e_{i+\ell;r})$ in superposition, then applying the "full" measurement (2.86) on $\left|e_{i;i+\ell-1}\right\rangle$, and finally reversing the conversion $(e_{1;i-1}, e_{i;i+\ell-1}, e_{i+\ell;r}) \to e_{1;r}$. This way, using binary search, we can, e.g., locate the first (uncompressed) index where the measurement result is $\boldsymbol{I} - \boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|$ using $\log(r)$ such compressed measurements. With very high probability at most $O(h)$ indices will result in outcome $\boldsymbol{I} - \boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|$, so this binary-search-based compressed measurement scheme will terminate after $O(h \log(r))$ steps with very high probability. We refer the reader to [18, Section 5] for further details about the precise error and run-time bounds.

Once we obtained the list of indices $(i_1, i_2, \ldots, i_\ell)$ where the uncompressed measurement would have resulted in $\boldsymbol{I} - \boldsymbol{P}_0 \otimes |0^{c+1}\rangle\langle 0^{c+1}|$, we need to check the alternative acceptance condition, i.e., whether the second qubit is in state 1 or not, completing the verification whether all circuit segments were applied successfully. Given such an index i we first use the first "compression" register of the compressed encoding to identify how many nonzero (uncompressed) registers are before i, and then look up the corresponding block in the "data" register to check whether the second qubit is in state 1; if the i-th uncompressed register contains $\left|0^{a+2}\right\rangle$ according to the "compression" register then we conclude that the second qubit is in state 0 without looking at the "data" register. All of these operations can be performed in time that is polynomial in the size of the "compression" register, which is $O(\text{polylog}(1/\epsilon))$.

This completes the description of how to simulate $\mathrm{e}^{t\mathcal{L}}[\cdot]$ to precision $O(1/\epsilon)$ with success probability $\geq 1/2$ when $t \leq 2$ using

$$\begin{aligned} O(h(a + \log((r+1)))) = O(\log(1/\epsilon)(a + \log(1/\epsilon))) \quad & \text{ancilla qubits,} \\ h = O\left(\frac{\log(1/\epsilon)}{\log\log(1/\epsilon)}\right) \quad & \text{(controlled) uses of } \boldsymbol{U} \text{ and } \boldsymbol{U}^\dagger, \\ O((a+1)\text{polylog}(1/\epsilon)) \quad & \text{other two-qubit gates.} \end{aligned}$$

The success probability can be improved to 1 while keeping the precision $O(1/\epsilon)$, using 3-steps of oblivious amplitude amplification, cf. [49, 73]. For $t > 2$, we divide up the evolution to $\lceil t/2 \rceil$ equal segments and repeat the process $\lceil t/2 \rceil$ times, setting the precision to ϵ/t in each segment. This gives the stated final complexity.

With a slight modification, we can make the above algorithm work for general Lidbladians as well, assuming that we have an (at most) a-qubit block-encoding $\boldsymbol{V}$ of the driving Hamiltonian term $\boldsymbol{H}$. One just needs to modify the circuit of Figure 2.10, sketched in Figure 2.12.

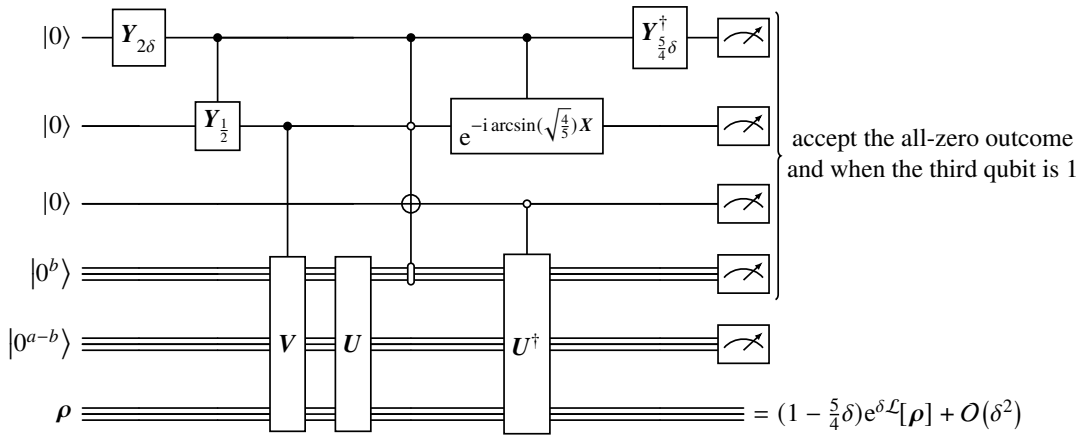


Figure 2.12: Alternative quantum circuit implementation of an approximate δ-time step via a postselective weak measurement scheme including the coherence term $-i[\boldsymbol{H}, \boldsymbol{\rho}]$ for the block-encoded Hamiltonian $\boldsymbol{H} = (\langle 0^c| \otimes \boldsymbol{I})\boldsymbol{V}(|0^c\rangle \otimes \boldsymbol{I})$.

For completeness, we include the analysis of the circuit $\boldsymbol{C_H}$ from Figure 2.12. Let us define

$$\boldsymbol{R} := (\boldsymbol{Y}^\dagger_{\frac{5}{4}\delta} \otimes \boldsymbol{I}) \cdot \left(|0\rangle\langle 0| \otimes \boldsymbol{I} + |1\rangle\langle 1| \otimes \mathrm{e}^{-\mathrm{i}\arcsin(\sqrt{\frac{4}{5}})\boldsymbol{X}} \right).$$

Similarly to (2.85) we analyze the action of $\boldsymbol{C_H}$ on a pure state $|\psi\rangle$:

$$
\begin{aligned}
&\left|0^{a+3}\right\rangle|\psi\rangle \\
&\overset{(1)}{\to} (\sqrt{1-2\delta}|0\rangle + \sqrt{2\delta}|1\rangle)|00\rangle|0^c\rangle|\psi\rangle \\
&\overset{(2)}{\to} (\sqrt{1-2\delta}|00\rangle + \sqrt{\delta}|10\rangle + \sqrt{\delta}|11\rangle)|0\rangle|0^c\rangle|\psi\rangle \\
&\overset{(3)}{\to} (\sqrt{1-2\delta}|000\rangle + \sqrt{\delta}|100\rangle)|0^c\rangle|\psi\rangle + \sqrt{\delta}|110\rangle\boldsymbol{V}|0^c\rangle|\psi\rangle \\
&\overset{(4)}{\to} (\sqrt{1-2\delta}|000\rangle + \sqrt{\delta}|100\rangle)\boldsymbol{U}|0^c\rangle|\psi\rangle + \sqrt{\delta}|110\rangle\boldsymbol{U}\boldsymbol{V}|0^c\rangle|\psi\rangle \\
&\overset{(5)}{\to} \sqrt{1-2\delta}|000\rangle\boldsymbol{U}|0^c\rangle|\psi\rangle + \sqrt{\delta}|101\rangle\Big(|0^b\rangle\langle 0^b| \otimes \boldsymbol{I}\Big)\boldsymbol{U}|0^c\rangle|\psi\rangle \\
&+ \sqrt{\delta}|100\rangle(\boldsymbol{I} - |0^b\rangle\langle 0^b| \otimes \boldsymbol{I})\boldsymbol{U}|0^c\rangle|\psi\rangle\sqrt{\delta}|110\rangle\boldsymbol{U}\boldsymbol{V}|0^c\rangle|\psi\rangle \\
&= (\sqrt{1-2\delta}|000\rangle + \sqrt{\delta}|100\rangle)\boldsymbol{U}|0^c\rangle|\psi\rangle + \sqrt{\delta}|101\rangle\left|0^b\right\rangle\underbrace{(\left\langle 0^b\right| \otimes \boldsymbol{I})\boldsymbol{U}|0^c\rangle|\psi\rangle}_{|\psi_0'\rangle :=} \\
&- \sqrt{\delta}|100\rangle\Big(|0^b\rangle\langle 0^b| \otimes \boldsymbol{I}\Big)\boldsymbol{U}|0^c\rangle|\psi\rangle + \sqrt{\delta}|110\rangle\boldsymbol{U}\boldsymbol{V}|0^c\rangle|\psi\rangle \\
&\overset{(6-7)}{\to} (\boldsymbol{R} \otimes \boldsymbol{I})\Big((\sqrt{1-\delta}|000\rangle + \sqrt{\delta}|100\rangle)|0^c\rangle|\psi\rangle + \sqrt{\delta}|101\rangle\left|0^b\right\rangle\left|\psi_0'\right\rangle \\
&-\sqrt{\delta}|100\rangle\boldsymbol{U}^\dagger\Big(|0^b\rangle\langle 0^b| \otimes \boldsymbol{I}\Big)\boldsymbol{U}|0^c\rangle|\psi\rangle\Big) + (\boldsymbol{R} \otimes \boldsymbol{I})\sqrt{\delta}|110\rangle\boldsymbol{V}|0^c\rangle|\psi\rangle.
\end{aligned}
$$

Considering that

$$
\langle 000|(\boldsymbol{R} \otimes I) = \left(\boldsymbol{R}^\dagger|00\rangle\right)^\dagger\langle 0| = \left(\sqrt{1-\frac{5}{4}\delta}\langle 000| + \sqrt{\frac{\delta}{4}}\langle 100| - i\sqrt{\delta}\langle 110|\right),
$$

let us compute the part of $\boldsymbol{C_H}\left|0^{a+3}\right\rangle|\psi\rangle$ starting with $\left|0^{a+3}\right\rangle$:

$$
\begin{aligned}
&(\left\langle 0^{a+3}\right| \otimes \boldsymbol{I})\boldsymbol{C_H}\left|0^{a+3}\right\rangle|\psi\rangle \\
&= \underbrace{\left(\sqrt{\left(1-\frac{5}{4}\delta\right)(1-\delta)} + \frac{\delta}{2}\right)}_{=1-\frac{5}{8}\delta+O(\delta^2)=\sqrt{1-\frac{5}{4}\delta}+O(\delta^2)}|\psi\rangle - \frac{\delta}{2}\underbrace{(\langle 0^c| \otimes \boldsymbol{I})\boldsymbol{U}^\dagger\Big(|0^b\rangle\langle 0^b| \otimes \boldsymbol{I}\Big)\boldsymbol{U}|0^c\rangle|\psi\rangle}_{\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j|\psi\rangle} \\
&\qquad\qquad - i\delta(\langle 0^c| \otimes \boldsymbol{I})\boldsymbol{V}|0^c\rangle|\psi\rangle \\
&= \sqrt{1-\frac{5}{4}\delta}\left(\left(\boldsymbol{I} - i\delta\boldsymbol{H} - \frac{\delta}{2}\sum_{j\in J}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\right)|\psi\rangle + O\left(\delta^2\right)\right),
\end{aligned}
$$

and the part where third qubit is $|1\rangle$:

$$
(\boldsymbol{I} \otimes \langle 1| \otimes \boldsymbol{I})\boldsymbol{C_H}\left|0^{a+3}\right\rangle|\psi\rangle = \sqrt{\delta}\boldsymbol{R}|10\rangle\left|0^b\right\rangle\left|\psi_0'\right\rangle = \sqrt{\delta}\boldsymbol{R}|10\rangle\left|0^b\right\rangle\sum_{j\in J}|j\rangle\boldsymbol{L}_j|\psi\rangle.
$$

Let $\widetilde{\boldsymbol{\Pi}} := \frac{1}{\sqrt{1-\frac{5}{4}\delta}}\left(|0^{a+3}\rangle\langle 0^{a+3}| \otimes \boldsymbol{I} + \boldsymbol{I} \otimes |1\rangle\langle 1| \otimes \boldsymbol{I}\right)$, the above implies that

$$\begin{aligned}
&\mathrm{Tr}_{a+3}\left(\widetilde{\boldsymbol{\Pi}} C\left(|0^{a+3}\rangle\langle 0^{a+3}| \otimes |\psi\rangle\langle\psi|\right) C^\dagger \widetilde{\boldsymbol{\Pi}}\right) \\
&= \left(\boldsymbol{I} - i\delta \boldsymbol{H} - \frac{\delta}{2}\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j\right)|\psi\rangle\langle\psi|\left(\boldsymbol{I} + i\delta \boldsymbol{H} - \frac{\delta}{2}\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j\right) \\
&+ \delta \sum_{j\in J} \boldsymbol{L}_j |\psi\rangle\langle\psi| \boldsymbol{L}_j^\dagger + \mathcal{O}\left(\delta^2\right) \\
&= \mathrm{e}^{\delta \mathcal{L}}[|\psi\rangle\langle\psi|] + \mathcal{O}\left(\delta^2\right).
\end{aligned}$$

Similarly to the proof of Theorem 2.3.1, it is easy to see that this implies

$$\|\mathrm{Tr}_{a+2}\left(\widetilde{\boldsymbol{\Pi}} C\left(|0^{a+3}\rangle\langle 0^{a+3}| \otimes [\cdot]\right) C^\dagger \widetilde{\boldsymbol{\Pi}}\right) - \mathrm{e}^{\delta \mathcal{L}}[\cdot]\|_\diamond = \mathcal{O}\left(\delta^2\right). \qquad \blacksquare$$

2.12 Appendix:Quantum simulated annealing

A subroutine for a coherent Gibbs sampler is to prepare the top eigenvector of the discriminant. In semi-group settings, one simply iterates the map for an arbitrary initial state to find its fixed point; given coherent access to some discriminant $\mathcal{D}$ (which is not quite a CPTP map), the standard approach is *quantum simulated annealing* [24, 179, 186]. To keep this section self-contained, in the following, we assume coherent access to some discriminants $\mathcal{D}_{\beta_j}$. First, we use QSVT to boost the gap; this is the origin of the quadratic speedup.

Proposition 2.12.1 (Quadratic speedup [73, 118])**.** *Given a block-encoding $\boldsymbol{U}_{\mathcal{D}}$ of a Hermitian matrix $\boldsymbol{I} + \mathcal{D}$ with eigenvalue gap $\lambda_{gap}(\mathcal{D})$ and $\lambda_1(\mathcal{D}) \geq -\lambda_{gap}(\mathcal{D})$, we can construct a unitary $\boldsymbol{U}'$ block-encoding a matrix $p(\boldsymbol{I} + \mathcal{D})$ that has the same top eigenvector as $\mathcal{D}$ but with $O(1)$ eigenvalue gap, with $O(\frac{1}{\sqrt{\lambda_{gap}(\mathcal{D})}})$ uses of $\boldsymbol{U}_{\mathcal{D}}$ and $\boldsymbol{U}_{\mathcal{D}}^\dagger$.*

Second, following [24], we consider the discretized adiabatic paths through temperatures

$$\begin{aligned}
&\left|\lambda_1(\mathcal{D}_{\beta_0})\right\rangle \to \cdots \left|\lambda_1(\mathcal{D}_{\beta_j})\right\rangle \to \cdots \left|\lambda_1(\mathcal{D}_{\beta_k})\right\rangle \\
&\boldsymbol{U}_{\mathcal{D}_{\beta_0}} \to \cdots \boldsymbol{U}_{\mathcal{D}_{\beta_j}} \to \cdots \boldsymbol{U}_{\mathcal{D}_{\beta_k}} \quad \text{where} \quad \beta_j = \frac{j}{k}\beta.
\end{aligned} \tag{2.87}$$

While a more refined annealing schedule is possible, we consider the above linear schedule for simplicity. In particular, the initial state is the maximally entangled state $\beta = 0$. In the following, we show that choosing $k = \Theta(\beta\|\boldsymbol{H}\|)$ ensures that the

consecutive overlaps remain constant large, allowing us to jump between consecutive states using a few steps of (fixed-point) amplitude amplification.

Proposition 2.12.2 (Consecutive overlaps)**.** *Suppose the discriminants have a top eigenvector close to the purified Gibbs states*

$$\| \left|\lambda_1(\mathcal{D}_{\beta_j})\right\rangle - \left|\sqrt{\boldsymbol{\rho}_{\beta_j}}\right\rangle \| \leq \frac{1}{10}. \tag{2.88}$$

Let $\delta\beta := \beta_{j+1} - \beta_j$, then the consecutive overlaps are large

$$\left|\langle \lambda_1(\mathcal{D}_{\beta_j}) | \lambda_1(\mathcal{D}_{\beta_{j+1}}) \rangle\right|^2 \geq \frac{7}{10} - O((\delta\beta)^2 \|\boldsymbol{H}^2 \mathrm{e}^{-\delta\beta \boldsymbol{H}}\|). \tag{2.89}$$

Proof. Let us evaluate the overlap between the ideal Gibbs states and rewrite using the Hilbert-Schmidt inner product

$$\begin{aligned}\left|\langle \sqrt{\boldsymbol{\rho}_{\beta_j}} | \sqrt{\boldsymbol{\rho}_{\beta_{j+1}}} \rangle\right|^2 = \frac{\mathrm{Tr}\left[\mathrm{e}^{-\beta_j \boldsymbol{H}/2} \mathrm{e}^{-\beta_j \boldsymbol{H}/2} \mathrm{e}^{-\delta\beta \boldsymbol{H}/2}\right]^2}{\mathrm{Tr}[\mathrm{e}^{-\beta_j \boldsymbol{H}}] \mathrm{Tr}[\mathrm{e}^{-\beta_j \boldsymbol{H}} \mathrm{e}^{-\delta\beta \boldsymbol{H}}]} &= \frac{\langle \mathrm{e}^{-\delta\beta \boldsymbol{H}/2} \rangle_{\beta_j}^2}{\langle \boldsymbol{I} \rangle_{\beta_j} \langle \mathrm{e}^{-\delta\beta \boldsymbol{H}} \rangle_{\beta_j}} \\ &= 1 - O((\delta\beta)^2 \langle \boldsymbol{H}^2 \mathrm{e}^{-\delta\beta \boldsymbol{H}} \rangle_{\beta_j}) \\ &\geq 1 - O((\delta\beta)^2 \|\boldsymbol{H}^2 \mathrm{e}^{-\delta\beta \boldsymbol{H}}\|),\end{aligned}$$

where we denote the thermal expectation by $\langle \boldsymbol{A} \rangle_\beta = \mathrm{Tr}[\boldsymbol{\rho}_\beta \boldsymbol{A}]$. The last equality expands the exponential

$$\mathrm{e}^{-\delta\beta \boldsymbol{H}/2} = \boldsymbol{I} - \delta\beta \boldsymbol{H}/2 + (\mathrm{e}^{-\delta\beta \boldsymbol{H}/2} - \boldsymbol{I} + \delta\beta \boldsymbol{H}/2).$$

The small discrepancy (2.88) between $\left|\lambda_1(\mathcal{D}_{\beta_j})\right\rangle$ and $\left|\sqrt{\boldsymbol{\rho}_{\beta_j}}\right\rangle$ only induces a minor change in the overlaps thus proving (2.89). ∎

Proposition 2.12.3 (Simulated annealing)**.** *In the setting of Proposition 2.12.2, following the discretized adiabatic path (2.87) with $k = \lceil c\beta \|\boldsymbol{H}\| \rceil$ prepares a state $|\psi\rangle$ such that $\| |\psi\rangle - \left|\lambda_1(\mathcal{D}_{\beta_{j+1}})\right\rangle \| \leq \delta$ using*

$$O\left(\frac{\beta \|\boldsymbol{H}\|}{\min_j \sqrt{\lambda_{gap}(\mathcal{D}_{\beta_j})}} \log^2(\beta \|\boldsymbol{H}\| / \delta) \right).$$

total calls for the oracle for discriminants $\mathcal{D}_{\beta_j}$.

Proof. Use fixed-point amplitude amplification [185] to "jump" between the eigenvectors $\left|\lambda_1(\mathcal{D}_{\beta_j})\right\rangle$. We can implement a $\frac{\sqrt{\delta}}{k}$-approximate projector to each eigenvector with $O(\frac{\log(k/\delta)}{\sqrt{\lambda_{gap}(\mathcal{D}_{\beta_j})}})$ calls to the block-encoded discriminants $\mathcal{D}_{\beta_j}$. To ensure that all the k jumps are all approximated to error $\leq \frac{\delta}{k}$, all fixed-point amplitude amplification consists of $\log(k/\delta)$ rounds, each calling the block-encoded (approximate) projectors. ∎

A simple lower bound on β dependence

In this section, we prove a simple lower-bound for the temperature dependence in the sense of implementing a reflection about the purified Gibbs state.

Proposition 2.12.4 (Lower-bound on simulation time)**.** *A circuit implementing the reflection operator*

$$\boldsymbol{R}_{\beta,\boldsymbol{H}} := \boldsymbol{I} - \left|\sqrt{\boldsymbol{\rho}_{\beta,\boldsymbol{H}}}\right\rangle\!\left\langle\sqrt{\boldsymbol{\rho}_{\beta,\boldsymbol{H}}}\right|$$

using Hamiltonian simulation for $\boldsymbol{H}$ as a black-box must use Hamiltonian simulation time $T = \Omega(\beta)$.

Proof. The idea is to argue that the reflection operator is sensitive to β and the Hamiltonian $\boldsymbol{H}$, so the Hamiltonian simulation time T cannot be too short. First, we control the norm change of the reflection operator. Let $\boldsymbol{\rho}_{\beta,\boldsymbol{H}} =: \boldsymbol{\rho}$ and $\boldsymbol{\rho}_{\beta,\boldsymbol{H}'} =: \boldsymbol{\rho}'$, then

$$\begin{aligned}
\|\boldsymbol{R}_{\beta,\boldsymbol{H}} - \boldsymbol{R}_{\beta,\boldsymbol{H}'}\| &= \||\sqrt{\boldsymbol{\rho}}\rangle\!\langle\sqrt{\boldsymbol{\rho}}| - |\sqrt{\boldsymbol{\rho}'}\rangle\!\langle\sqrt{\boldsymbol{\rho}'}|\| \\
&\geq \left|\langle\sqrt{\boldsymbol{\rho}}|\Big(|\sqrt{\boldsymbol{\rho}}\rangle\!\langle\sqrt{\boldsymbol{\rho}}| - |\sqrt{\boldsymbol{\rho}'}\rangle\!\langle\sqrt{\boldsymbol{\rho}'}|\Big)|\sqrt{\boldsymbol{\rho}}^{\perp}\rangle\right| \\
&= \left|\langle\sqrt{\boldsymbol{\rho}}|\sqrt{\boldsymbol{\rho}'}\rangle\langle\sqrt{\boldsymbol{\rho}'}|\sqrt{\boldsymbol{\rho}}^{\perp}\rangle\right| \\
&= \sqrt{1-\theta^2}\cdot\theta,
\end{aligned}$$

where $\theta := \sqrt{1 - \left|\langle\sqrt{\boldsymbol{\rho}}|\sqrt{\boldsymbol{\rho}'}\rangle\right|^2}$ and

$$\left|\sqrt{\boldsymbol{\rho}}^{\perp}\right\rangle \in Span\{\left|\sqrt{\boldsymbol{\rho}}\right\rangle, \left|\sqrt{\boldsymbol{\rho}'}\right\rangle\} \quad \text{such that} \quad \langle\sqrt{\boldsymbol{\rho}}^{\perp}|\sqrt{\boldsymbol{\rho}}\rangle = 0.$$

Now, for infinitesimal $\epsilon \to 0$, let

$$\begin{aligned}
\boldsymbol{H} &= |0\rangle\!\langle 0| + |1\rangle\!\langle 1| \\
\boldsymbol{H}' &= |0\rangle\!\langle 0| + (1+\epsilon)|1\rangle\!\langle 1|
\end{aligned}$$

such that

$$\begin{aligned}
\left|\sqrt{\boldsymbol{\rho}}\right\rangle &\propto |00\rangle + |11\rangle \\
\left|\sqrt{\boldsymbol{\rho}'}\right\rangle &\propto |00\rangle + (1+\beta\epsilon/2)|11\rangle + O(\epsilon^2).
\end{aligned}$$

Direct calculation gives $\theta = \Omega(\beta\epsilon)$ for small $\epsilon \to 0$. To conclude the proof, suppose the block-box circuit uses only Hamiltonian simulation time T. Then, the resulting

circuits for $\boldsymbol{H}$ and $\boldsymbol{H}'$ can only differ by $\|\boldsymbol{H} - \boldsymbol{H}'\|T = \epsilon T$. Therefore, for small $\epsilon \to 0$,

$$T\epsilon \geq \sqrt{1-\theta^2} \cdot \theta = \Omega(\beta\epsilon),$$

proving the advertised result. ∎

The above sensitivity argument similarly applies to algorithms preparing the Gibbs state using black-box Hamiltonian simulation.

Proposition 2.12.5 (Lower-bound on simulation time)**.** *A circuit preparing the Gibbs state $\boldsymbol{\rho}_\beta \propto \mathrm{e}^{-\beta \boldsymbol{H}}$ using Hamiltonian simulation for $\boldsymbol{H}$ as a black-box must use Hamiltonian simulation time $T = \Omega(\beta)$.*

Proof. Again, consider $\boldsymbol{H}$ and $\boldsymbol{H}'$ as above and their Gibbs states

$$\begin{aligned}\boldsymbol{\rho} &\propto |0\rangle\langle 0| + |1\rangle\langle 1| \\ \boldsymbol{\rho}' &\propto |0\rangle\langle 0| + (1+\beta\epsilon)|1\rangle\langle 1| + O(\epsilon^2).\end{aligned}$$

Then, for infinitesimal $\epsilon \to 0$,

$$\|\boldsymbol{\rho} - \boldsymbol{\rho}'\|_1 \geq \Omega(\beta\epsilon),$$

which implies $T = \Omega(\beta)$ as advertised. ∎

2.13 Appendix:Impossibility of boosted shift-invariant in-place phase estimation

In this section, we include the proof that certain "boosted shift-invariant in place phase estimation" utilized in [167] is impossible. The impossibility result was developed in parallel with this work; we reproduce the main argument here with the permission of András Gilyén and Dávid Matolcsi until their manuscript becomes publicly available.

We begin by reviewing the phase estimation assumptions made by [167]. First, they assume [167, Eqn.(11), Supplemental Information] the phase estimation map is shift-invariant in the sense that

$$\Phi := \sum_{\bar{\nu}} \sum_{\bar{\mu}} \boldsymbol{M}_{\bar{\nu}}^{\bar{\mu}} \otimes |\bar{\nu}\rangle\langle\bar{\mu}| \quad \text{where} \quad \boldsymbol{M}_{\bar{\nu}}^{\bar{\mu}} := \sum_i \alpha(E_i, \bar{\nu} - \bar{\mu}) |\psi_i\rangle\langle\psi_i| \quad \text{and} \quad \bar{\nu}, \bar{\mu} \in \{\mathbb{Z}\bar{\omega}\}.$$

In [167], the shift-invariance was shown to hold for an unboosted phase estimation unitary [167, Eqn.(12), Supplemental Information], which has a slowly decaying tail when E_i deviates substantially from $\bar{\nu} - \bar{\mu}$.

Second, to prove the correctness of the fixed point, they impose [167, Eqn.(10), Supplemental Information] that the profile $\alpha(E_i, \bar{\nu} - \bar{\mu})$ can be boosted:[52] the only nonzero matrix elements are such that

$$|\bar{\nu} - \bar{\mu} - E_i| < \bar{\omega},$$

where $\bar{\omega}$ is the energy resolution for the phase estimation readout registers. Unfortunately, these two assumptions are not compatible with each other, as argued by the following.

Proposition 2.13.1 (Impossibility for shift-invariant boosting)**.** *There exists no continuous family of "boosted shift-invariant in place phase estimation" unitaries. More precisely, for every constant k for large enough N there exists no profile $\alpha\colon \mathbb{R} \times \frac{2\pi}{N} \cdot \{0, 1, 2, \ldots, N-1\} \to \mathbb{C}$ that simultaneously satisfies:*

- *(almost) unitarity:* $\boldsymbol{M}_E := \sum_{i,j=0}^{N-1} \alpha(E, \frac{2\pi}{N}(i-j \mod N))|i\rangle\langle j|$ *is close to some unitary* $\boldsymbol{U}_E$ *for all* $E \in \mathbb{R}$*:* $\|\boldsymbol{M}_E - \boldsymbol{U}_E\| \leq \frac{1}{2}$
- *boosting:* $\alpha(E, \bar{\nu} - \bar{\mu}) = 0$ *if* $|\bar{\nu} - \bar{\mu} - E \mod 2\pi| > k\frac{2\pi}{N}$
- *continuity:* $\alpha(E, \bar{\nu})$ *depends continuously on* E

Proof. We prove the statement by contradiction. Let us assume that such a profile exists for $N \gg k^3$. We will track how the profile changes as we increase the energy from $E_0 = k\frac{2\pi}{N}$ to $E_1 = (3k+1)\frac{2\pi}{N}$.

Now, consider the polynomial $p_E(z) = \sum_{j=0}^{4k+1} \alpha(E, j\frac{2\pi}{N}) z^j$ (whose physical meaning will become clear). Due to boosting the polynomial $p_{E_0}(z)$ has a degree at most $2k$, so it has at most $2k$ roots, and in particular, at most $2k$ roots are situated within the complex unit circle. Due to continuity, the polynomial $p_E(z)$ is continuously transformed to $p_{E_1}(z)$ whose smallest nonzero coefficient comes with a power of z at least $2k+1$ due to boosting. This implies that 0 is a root with multiplicity at least $2k+1$, and in particular, we have at least $2k+1$ roots within the unit

[52] This is explicitly stated in the last paragraph of page 13 [167, Supplemental Information]: "According to (10) we can replace the function $f(E_j, k-p)$ by its enhanced counterpart $\alpha_{E_j}(k-p)$, which acts as a binary amplitude for the two closest r-bit integers to the actual energy E_j."

circle. Since the polynomial changes continuously, its multi-set of roots also changes continuously, which means that at some point, a root must enter the unit circle[53] (here we acknowledge that some roots enter from infinity when the degree of the polynomial increases but that does not affect our argument — one can make this precise by tracking roots on the surface of the Riemann sphere). Thus, there is some energy E' for which a complex unit number $z = \mathrm{e}^{2\pi i\varphi}$ is a root of the corresponding polynomial $p_{E'}(z)$.

We show that this implies that the plane wave with quasi-momentum φ is (almost) in the kernel of the shift-invariant matrix $\boldsymbol{M}_{E'}$: for every $i \le N - 4k - 2$ we have

$$\langle i|\boldsymbol{M}_{E'} \sum_{j=0}^{N-1} |j\rangle z^{-j} = \sum_{j=0}^{N-1} \alpha(E', \frac{2\pi}{N}(i-j \mod N)) z^{-j} = z^{-i} p_{E'}(z) = 0,$$

implying

$$\|\boldsymbol{M}_{E'} \sum_{j=0}^{N-1} |j\rangle z^{-j}\|^2 = O\left(k^3\right) \ll N = \|\boldsymbol{U}_E \sum_{j=0}^{N-1} |j\rangle z^{-j}\|^2,$$

a contradiction. ∎

[53] The idea of tracking the roots of this polynomial is due to Dávid Matolcsi.

Table 2.1: A comparison of existing thermal state preparation algorithms[7].

Algorithms	Ham. sim. time	assumptions	potential caveats
Quantum Metropolis [167]: a quantum version of rejection sampling via QPE	$\mathrm{Poly}(\epsilon^{-1}, \beta, n) t_{QPE} \cdot t_{mix}$	shift-invariant, boosted QPE (provably impossible)	without the assumption, the QPE runtime is uncontrolled
[149]: simulate a Lindbladian with forbidden energy transitions	$\tilde{O}(\frac{1}{\gamma_{att}} \frac{\beta^3 \tilde{t}_{mix}}{\epsilon^7})$[1]	-	overhead for randomized rounding; forbidding transitions impacts $\tilde{t}_{mix}$
[160]: system-bath evolution at weak-coupling	$O(\frac{\beta t_{mix}^3}{\epsilon})$	controllable, refreshable bath and ETH[2]	large bath; no guarantees without ETH.
[38]: system-bath evolution at weak-coupling	$\mathrm{Poly}(\epsilon^{-1}, \beta, n, t_{mix})$	controllable, refreshable bath	large bath and large overheads
Theorem 2.1.3: simulate a Lindbladian via operator FT	$\tilde{O}(\frac{\beta t_{mix}}{\epsilon} \cdot t_{mix})$	-	-
Quantum[2] Metropolis [186]: a coherent version of Quantum Metropolis via QPE	$\tilde{O}(t_{QPE} \cdot \frac{\beta^2 \langle \boldsymbol{H}^2 \rangle_{\beta=0}}{\epsilon \sqrt{\lambda_{gap}}})$[3]	perfect QPE	needs nondegenerate energies,[4] $e^{\Omega(n)}$ runtime for QPE
[180]:a coherent version of Lindbladian via QPE	$\tilde{O}(t_{QPE} \cdot \frac{\beta \lVert \boldsymbol{H} \rVert}{\sqrt{\lambda_{gap}}})^3$	rounding promise for $\boldsymbol{H}$	unknown error for unrounded Hamiltonians
Theorem 2.1.4: a coherent version of Lindbladian via operator FT	$\tilde{O}(\frac{\beta^2 \lVert \boldsymbol{H} \rVert}{\lambda_{gap}^{3/2}} + \frac{\beta}{\epsilon \lambda_{gap}^{3/2}})$	-	-
	Gate complexity		
[26]: patching via recovery maps	$e^{O(C^d \ln^d(n/\epsilon))}$[5]	Markov and clustering with length scale C	nonconstructive, quasi-local recovery maps
QITE [132]: imaginary time evolution via local tomography	$e^{\tilde{O}(C^d \ln^d(\beta n/\epsilon))}$	correlation length C	costly at low-temperature, strong correlation, or high accuracy
[133]: drive transition by measurements	$\tilde{O}(t_{mix}\beta)$	certain measurement operator	the measurement must have "good overlap" with energy basis[6]
[90]:a perturbative approach	$e^{O(\beta \lVert \boldsymbol{V} \rVert)}$	$\boldsymbol{H} = \boldsymbol{H}_0 + \boldsymbol{V}$ for small $\boldsymbol{V}$ and simple $\boldsymbol{H}_0$	costly at low-temperature or nonperturbative regime
QSVT [73]: directly implementing $e^{-\beta \boldsymbol{H}}$	$e^{O(\beta \lVert \boldsymbol{H} \rVert)}$	-	not scalable

Chapter 3

AN EFFICIENT AND EXACTLY DETAILED-BALANCED QUANTUM GIBBS SAMPLER

3.1 Introduction

One of the leading candidate applications of quantum computers [54] is to simulate quantum systems [68]. In particular, the preparation of thermal states or ground states for materials and molecules has received significant attention [14, 30, 37, 107]. Surprisingly, there has not been a consensus on the "go-to" quantum algorithm for this task due to a lack of provable guarantees or empirical evidence [108]. Recently, several Monte Carlo-style, nonunitary quantum algorithms have been proposed [40, 61, 149, 167, 180, 186]. While their efficacy has only been validated using small-scale numerics and under strong theoretical assumptions [38, 61, 160], there are reasons for optimism. On physical grounds, these algorithms resemble naturally occurring system-bath dynamics [134]; if a system rapidly cools in a refrigerator, the same plausibly applies to a "cooling algorithm" that emulates this process. Alternatively, from a computer science perspective, these algorithms are cousins of classical Markov chain Monte Carlo (MCMC) methods but with quantum mechanical effects and complications. This work sets out to complete this line of thought and construct an ideal quantum MCMC algorithm where the robustness, simplicity, and empirical success of the classical case may be transferrable.

The cornerstone of classical Markov chain Monte Carlo methods is *detailed balance* (see, e.g., [110]): given a target state π, we impose a certain symmetry of the Markov chain $\boldsymbol{M}$

$$\boldsymbol{M}_{s's}\pi_s = \pi_{s'}\boldsymbol{M}_{s's} \quad \text{for each configuration} \quad s, s', \quad \text{ensuring stationarity} \quad \boldsymbol{M}[\pi] = \pi.$$

This simple recipe for the stationary state has been crucial in constructing and analyzing the Metropolis-Hastings algorithm and related Markov chains. Notably, detailed balance gives a conceptually simple picture of convergence via the spectral gap of $\boldsymbol{M}$, a quantity amenable to numerical and analytic bounds. If the problem at hand has a local structure, detailed balance can often be imposed locally and efficiently, relegating the algorithm's complexity to the *mixing time*, the time scale of convergence towards stationarity. While the mixing time may be challenging to

analyze, MCMC methods can often be employed heuristically. In particular, we are often interested in sampling the *Gibbs distribution* $\pi_\beta \propto e^{-\beta H}$ of a certain energy functional H at temperature $1/\beta$. Analogously, the central idea of *quantum Gibbs sampling* is to construct a detailed-balanced quantum process where the quantum Gibbs state is stationary. In this work, we focus on designing a Lindbladian $\mathcal{L}_\beta$ (the quantum analog of a continuous-time Markov chain generator) such that

$$e^{\mathcal{L}_\beta t}[\boldsymbol{\rho}_\beta] = \boldsymbol{\rho}_\beta \quad \text{where} \quad \boldsymbol{\rho}_\beta := e^{-\beta \boldsymbol{H}}/\mathrm{Tr}(e^{-\beta \boldsymbol{H}}),$$

for *any* target quantum Hamiltonian $\boldsymbol{H}$. As in the classical case, we can prepare samples of quantum Gibbs states if the Lindbladian evolution can be *efficiently* implemented and the state converges *rapidly* to the Gibbs state.

The main issue with existing quantum Gibbs sampling algorithms is that *quantum detailed balance* (Figure 3.1) only holds *approximately* unless we can distinguish individual energy eigenstates exactly, which is generally intractable except for fast-forwardable Hamiltonians (e.g., Hamiltonians with commuting terms). Consequently, we either lose accuracy guarantees for the stationary state or the efficiency for the individual steps of the Gibbs sampling algorithm, leading to significant aggregated complexity plaguing various constructions; see Ref. [40] for a comprehensive catalog. The algorithmic challenge in enforcing quantum detailed balance is the energy-time uncertainty principle rooted in *metrology*: for each energy estimate, the uncertainty scales inversely proportional to the Hamiltonian simulation time. Indeed, all existing quantum MCMC algorithms attempt to attain detailed balance via an "energy estimation" subroutine (quantum phase estimation [149, 167, 180] or operator Fourier Transform [40, Appendix A]). Consequently, this error propagates to the desired Gibbs state and impacts the implementation cost.

To our knowledge, the best general lower-bound on the Hamiltonian simulation time is $\Omega(\beta)$ per Gibbs sample [40, Proposition G.5]. This comes from a sensitivity argument that the Gibbs state is a smooth matrix function of $\boldsymbol{H}$ with derivatives bounded by $O(\beta)$. This conceptual gap motivates our guiding question:

Can we design an efficiently implementable
yet exactly detailed-balanced quantum Gibbs sampler?

If so, we may recover both the simplicity and versatility of classical MCMC algorithms. In this work, we answer this question in the affirmative by explicitly constructing an exactly detailed-balanced Lindbladian at a moderate cost: $\tilde{O}(\beta)$-Hamiltonian

simulation cost per *unit time* of Linbladian evolution $e^{\mathcal{L}}$. (The unit-time evolution for a continuous-time generator should be regarded as "one step," corresponding to one discrete Markov chain update.) Furthermore, for lattice Hamiltonians (with local jumps), our Lindbladian is (quasi-)local with locality scaling as $\tilde{O}(\beta)$. Thus, one step of the algorithm only needs to simulate localized Hamiltonian patches; this starkly contrasts with previous works, whose cost per unit time step generally scales with the system size due to simulating the *global* Hamiltonian. The mathematical and conceptual simplicity of our result immediately initiates a list of new directions, which we discuss in detail in section 3.4.

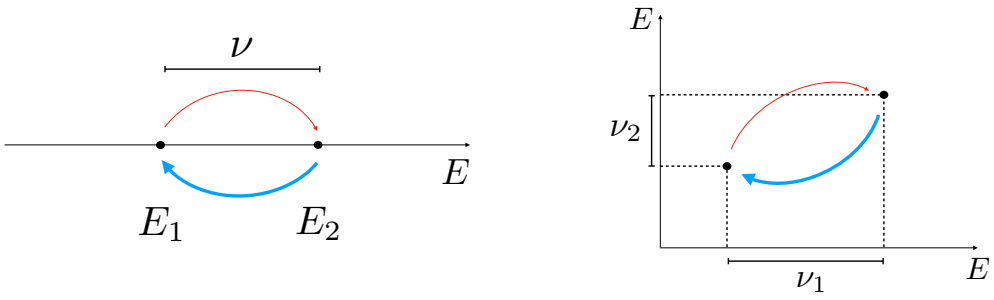
The key revelation behind our construction is that quantum detailed balance can be enforced *smoothly* without ever knowing the energy. Indeed, the standard *metrology* lower bound $\sim \Omega(\frac{1}{\epsilon})$ is not an obstruction because having access to a detailed-balanced Lindbladian (or the Gibbs state) does not give energy estimates. We have seen that Quantum Signal Processing [117] or Quantum Singular Value Transform (QSVT) [73] allows one to directly access smooth (polynomial) functions of Hamiltonians without a phase-estimation subroutine; the costs often scale linearly with the largest derivatives and only *logarithmically* with the precision. However, Lindbladians, as superoperators, are more restrictive to manipulate than Hermitian matrices. A key design ingredient is a carefully chosen *coherent* term in our Lindbladian

$$\mathcal{L}_\beta[\boldsymbol{\rho}] = \underbrace{-\mathrm{i}[\boldsymbol{B}, \boldsymbol{\rho}]}_{\text{"coherent"}} + (\text{"dissipative"}),$$

which appears necessary to *coherently* and *exactly* cancel out certain unwanted errors from the dissipative part.

As a by-product, purifying our Lindbladian yields a temperature-dependent family of "parent Hamiltonians" whose zero-eigenstate is a canonical purification of the Gibbs state.[1] Similarly to how classical Markov chains can be "quantized" to prepare the purified stationary state, here we prepare the purified Gibbs state by following a prescribed adiabatic path (called *quantum simulated annealing* [164, 186]), drawing a surprisingly simple connection between thermal dissipation and adiabatic evolution. In particular, for lattice Hamiltonians, the parent Hamiltonian inherits the (quasi-)locality, which curiously connects the purified Gibbs state to the ground state of (quasi-)local Hamiltonians.

[1] This purification coincides with the *Thermal Field Double* state featured in recent quantum gravity discussions. See, e.g., [122].



Classical Quantum

Figure 3.1: (Left) For the classical Gibbs distribution, the detailed balance condition is a pairwise relation between heating (red) and cooling (blue) transition rates, depending on the energy difference ν of states. (Right) For the quantum Gibbs state, the detailed balance condition refers to pairs of matrix elements of the density operator (expanded in the energy basis), where each matrix element is described by a pair of energies (of the basis elements in the ket and bra respectively), and therefore the relation depends on both of the respective energy differences ν_1 and ν_2.

Main results

Our main results, based on the algorithmic framework of [40, Section III], consider the following Lindbladian in the Schrodinger picture

$$\mathcal{L}_\beta[\cdot] := \underbrace{-\mathrm{i}[\boldsymbol{B}, \cdot]}_{\text{“coherent”}} + \sum_{a \in A} \int_{-\infty}^{\infty} \gamma(\omega) \left(\underbrace{\hat{\boldsymbol{A}}^a(\omega)(\cdot)\hat{\boldsymbol{A}}^a(\omega)^\dagger}_{\text{“transition”}} - \underbrace{\frac{1}{2}\{\hat{\boldsymbol{A}}^a(\omega)^\dagger \hat{\boldsymbol{A}}^a(\omega), \cdot\}}_{\text{“decay”}} \right) \mathrm{d}\omega \tag{3.1}$$

which is parameterized by the following terms (with convenient normalization conditions [40, Section I.B]):

- The distinct *jump operators* $\boldsymbol{A}^a$ “drive” the transitions, and can be chosen arbitrarily as long as their adjoints are included

$$\{\boldsymbol{A}^a : a \in A\} = \{\boldsymbol{A}^{a\dagger} : a \in A\} \quad \text{and} \quad \left\| \sum_{a \in A} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a \right\| \leq 1. \tag{3.2}$$

  For better mixing and ergodicity, the jumps should be “scrambling” and not commute with the Hamiltonian (e.g., breaking the symmetries of the Hamiltonian). For lattice Hamiltonians, the jump operators may be chosen simply to be the single-site Pauli operators, but global jumps could also be helpful in some cases as in the classical case (e.g., cluster updates).

- The *Operator Fourier Transform* (OFT)[2] (section 3.2) weighted by a Gaussian *filter* with a tunable width $\sim \sigma_E^{-1}$

$$\hat{\boldsymbol{A}}^a(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^a \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} \mathrm{e}^{-\mathrm{i}\omega t} f(t) \mathrm{d}t \tag{3.3}$$

$$\text{where} \quad f(t) := \mathrm{e}^{-\sigma_E^2 t^2} \sqrt{\sigma_E \sqrt{2/\pi}} = \frac{\mathrm{e}^{-t^2/\beta^2}}{\sqrt{\beta\sqrt{\pi/2}}} \quad \text{if} \quad \sigma_E = \frac{1}{\beta}.$$

  In particular, the Gaussian is normalized $\int_{-\infty}^{\infty} |f(t)|^2 \mathrm{d}t = 1$. Naturally, the Heisenberg evolution $\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^a \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}$ diagnoses the energy difference ω before and after the jump. Integrating over time $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\cdot) \mathrm{e}^{-\mathrm{i}\omega t} f(t) \mathrm{d}t$ yields the operator Fourier Transform $\hat{\boldsymbol{A}}^a(\omega)$, which selects the transitions of $\boldsymbol{A}^a$ that *increase* the energy by roughly $\sim \omega \pm O(\sigma_E)$. At first glance, the Gaussian filter seems to merely ensure good concentration for both the frequency and time domain, but it turns out to have a more intimate connection [133] to quantum detailed balance (see also section 3.8 for an alternative justification).

- The *transition weight* $\gamma(\omega)$ follows (yet another) Gaussian with a tunable variance $\sigma_\gamma > 0$:

$$\gamma(\omega) = \exp\left(-\frac{(\omega+\omega_\gamma)^2}{2\sigma_\gamma^2}\right) \quad \text{with variance} \quad \sigma_\gamma^2 := \frac{2\omega_\gamma}{\beta} - \sigma_E^2 \tag{3.4}$$

$$= \exp\left(-\frac{(\beta\omega+1)^2}{2}\right) \quad \text{if} \quad \sigma_E = \sigma_\gamma = \omega_\gamma = \frac{1}{\beta}. \tag{3.5}$$

  The normalization is such that $\|\gamma(\omega)\|_\infty \leq 1$, and the maximum is attained at $\omega = -\omega_\gamma$.

- A coherent (i.e., nondissipative) term generated by a fine-tuned Hermitian matrix $\boldsymbol{B}$. The expression depends on $\omega_\gamma, \sigma_E, \beta$ in the general case (see Corollary 3.3.1), but it simplifies to

$$\boldsymbol{B} := \sum_{a\in A} \int_{-\infty}^{\infty} b_1(t) \mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t} \left(\int_{-\infty}^{\infty} b_2(t') \mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t'} \boldsymbol{A}^{a\dagger} \mathrm{e}^{-2\mathrm{i}\beta\boldsymbol{H}t'} \boldsymbol{A}^a \mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t'} \mathrm{d}t' \right) \mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t} \mathrm{d}t$$

$$\text{if} \quad \omega_\gamma = \sigma_E = \sigma_\gamma = \frac{1}{\beta} \tag{3.6}$$

  for some carefully chosen smooth and rapidly decaying functions b_1, b_2 normalized by $\|b_1\|_1, \|b_2\|_1 \leq 1$. The coherent term $\boldsymbol{B}$ may appear intimidating

[2] Note the sign convention, which might differ from that of other works in the literature.

but plays a crucial role in ensuring *quantum detailed balance* for the Gibbs state $\boldsymbol{\rho}_\beta$, defined as

$$\mathcal{L}_\beta^\dagger[\cdot] = \sqrt{\boldsymbol{\rho}_\beta}^{-1}\mathcal{L}_\beta[\sqrt{\boldsymbol{\rho}_\beta}\cdot\sqrt{\boldsymbol{\rho}_\beta}]\sqrt{\boldsymbol{\rho}_\beta}^{-1} \quad \text{for fixed} \quad \beta, \boldsymbol{H}. \tag{3.7}$$

It implies the stationarity of Gibbs state exactly (see Definition 3.2.1).

The Gaussian transition weight (3.4) is inspired by an observation of [133]:[3] the functional form of Gaussians is naturally compatible with exact detailed balance[4] if we make conscious choices of ω_γ, σ_E

$$\exp\left(-\frac{(\omega+\omega_\gamma)^2}{2\sigma^2}\right) = \exp(-\frac{2\omega_\gamma}{\sigma^2}\cdot\omega)\exp\left(-\frac{(-\omega+\omega_\gamma)^2}{2\sigma^2}\right).$$

Compared with the usual step-function-like Metropolis weight $\min(1, \mathrm{e}^{-\beta\omega})$, the Gaussian weight is more selective, only allowing energy transitions $-\omega_\gamma \pm O(\sigma_\gamma)$; this narrower window could potentially freeze the dynamics, leading to long mixing time.

Fortunately, quantum detailed balance 3.7 is preserved under *linear combination* of Lindbladians; hence, choosing a linear combination of γ covering a range of different width σ_γ can remove the heavy restriction on energy transitions. Surprisingly, a suitable linear combination recovers Metropolis-like transition weights, which we focus on as the representative. To obtain the corresponding exactly detailed-balanced Lindbladian, the only change compared to (3.4)-(3.6) is the choice of transition weight

$$\text{(Metropolis-Style)} \quad \gamma^M(\omega) := \exp\left(-\beta\max\left(\omega+\frac{1}{2\beta}, 0\right)\right) \quad \text{if} \quad \sigma_E = \frac{1}{\beta} \tag{3.8}$$

with the corresponding coherent term $\boldsymbol{B}^M$ parameterized by another function $b_2^M(t)$ (the function $b_1^M(t) = b_1(t)$ remains the same as in (3.6)).[5]

Now, we present the first main result: the Gibbs state is an *exact* stationary state of the advertised Lindbladian (see section 3.2 for the proof). Although we have

[3] Their algorithm [133] seems qualitatively different from Monte Carlo style quantum algorithms [40, 149, 167, 180, 186] and closer to performing phase estimation on trial states; see the discussion in [133, Page 5].

[4] We thank Jonathan Moussa for pointing us to his paper and raising the question of whether detailed balance can hold exactly in the precursor of this work [40].

[5] The generalized function (distribution) $b_2^M(t)$ should be interpreted as the Cauchy principal value $\lim_{\eta\to 0+}\mathbb{1}(|t|>\eta)b_2^M(t)$. In case $[\boldsymbol{H}, \sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a] \neq 0$ an additional correction term $\frac{1}{16\sqrt{2}\pi}\delta(t)$ should be added.

mainly focused on the Gibbs state, we can formally invoke Gibbs sampling for $\beta \boldsymbol{H} = \log(\boldsymbol{\rho}_{fix})$ for arbitrary target stationary state $\boldsymbol{\rho}_{fix}$, albeit with potential overhead from implementing the matrix logarithm.

Theorem 3.1.1 (Gibbs state is stationary)**.** *For any $\beta \geq 0$, the Lindbladian (3.1)-(3.3) with $\sigma_E = \frac{1}{\beta}$, Gaussian transition wieght (3.5), and the coherent term $\boldsymbol{B}$ (3.6) satisfies detailed balance (3.7) exactly. Therefore, the Gibbs state is stationary*

$$\mathcal{L}_\beta[\boldsymbol{\rho}_\beta] = 0.$$

The same applies for the Metropolis transition weight $\gamma^M(\omega)$ (3.8) with the corresponding coherent term $\boldsymbol{B}^M$.

Furthermore, building on the algorithmic machinery developed in Ref. [40, Section III], the Lindbladian can be efficiently simulated at a moderate cost (see section 3.3 for the proof).

Theorem 3.1.2 (Efficient implementation)**.** *Instantiate the Lindbladian parameters of Theorem 3.1.1 for either the Gaussian $\gamma(\omega)$ or Metropolis $\gamma^M(\omega)$ transition weight. Then, the Lindbladian evolution*

$$e^{\mathcal{L}_\beta t} \quad \textit{for each} \quad t \geq 1$$

can be implemented efficiently in ϵ-diamond distance with cost:

$$\begin{aligned}
&\tilde{O}(t \cdot \beta) \quad \textit{(total Hamiltonian simulation time)} \\
&\tilde{O}(1) \quad \textit{(resettable ancilla)} \\
&\tilde{O}(t) \quad \textit{(block-encodings for the jumps } \sum_{a \in A} |a\rangle \otimes \boldsymbol{A}^a \textit{)} \\
&\tilde{O}(t) \quad \textit{(other two-qubit gates)}.
\end{aligned}$$

The $\tilde{O}(\cdot)$ notation absorbs logarithmic dependencies on $t, \beta, \|\boldsymbol{H}\|, n, 1/\epsilon, |A|$.

Here, our Lindbladian is normalized (3.2), (3.3), (3.4), (3.6) (or the Metropolis-like weight (3.8) with its coherent term $\boldsymbol{B}^M$) such that

$$\|\mathcal{L}_\beta\|_{1-1} = \tilde{O}(1).$$

Therefore, evolving for unit time $t = 1$ corresponds to a $\tilde{O}(1)$-strength update $e^{\mathcal{L}_\beta}$ and only requires a characteristic Hamiltonian simulation time $\sim \beta$. This is precisely

the cost for implementing both the operator Fourier Transform (3.3) and the coherent term (3.6) via Linear Combination of Unitaries. To be more careful, the $\tilde{O}(\cdot)$ notation also includes polylogarithmic factors due to discretization, truncation, and Hamiltonian simulation error, which is typical in quantum algorithms. To clarify, the idealized map $\mathcal{L}_\beta$ remains exactly detailed-balanced (the nice object to analyze), and the algorithmic implementation error can be made arbitrarily small given the desired runtime.

We expect the *total Hamiltonian simulation time* to be the figure of merit for the algorithmic cost, among others. The jump operators $\boldsymbol{A}^a$ can be as simple as Pauli operators, but we consider a black-box query model in case more complex jumps are needed for faster mixing. In the last line, the use of other two-qubit gates comes from easier-to-implement unitaries, including the Quantum Fourier Transform, state preparation unitary for the Gaussian filter $|f\rangle$, controlled transition weight; see section 3.3 and [40, Section III.B].

Combining Theorem 3.1.1 and Theorem 3.1.2, we can prepare the Gibbs state by simulating the Lindblaidan until convergence, resulting in the cost

$$\text{(total Hamiltonian simulation time per Gibbs sample)} = \widetilde{O}\big(t_{mix}(\mathcal{L}_\beta) \cdot \beta\big).$$

Formally, the *mixing time* $t_{mix}(\mathcal{L}_\beta)$ quantifies the shortest time scale that any two input states become indistinguishable (see Proposition 3.9.4). To obtain end-to-end gate complexities, we should also instantiate the Hamiltonian simulation cost, a subroutine whose complexity for various systems has been thoroughly studied. For example, on D-dimensional lattices with local jumps $\boldsymbol{A}^a$, we expect the actual cost to be (up to logarithmic error dependence)

$$\text{(gate complexity per unit evolution time)} \sim \underbrace{\beta}_{\text{Ham. sim. time}} \times \underbrace{(v_{LR}\beta)^D}_{\text{volume}}, \qquad \text{(spatially local Hamiltonians)}$$

where v_{LR} is the Lieb-Robinson velocity and $v_{LR}\beta$ is roughly the radius of the Heisenberg evolution $\boldsymbol{A}^a(\cdot)$ at time $\sim \beta$. Indeed, the Lindbladian is a sum over quasi-local Lindbladian operators (Figure 3.2)

$$\mathcal{L}_\beta = \sum_{a\in A} \mathcal{L}_\beta^a \quad \text{each centered at} \quad \boldsymbol{A}^a \quad \text{with radius} \quad \tilde{O}(v_{LR}\beta).$$

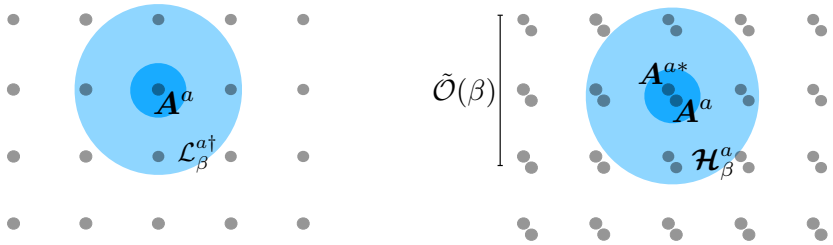


Figure 3.2: (Left) For lattice Hamiltonians, our Lindbladian is a sum of quasi-local terms $\mathcal{L}_\beta^a$ localized around each jump $\boldsymbol{A}^a$ with radius $\tilde{O}(\beta)$. Indeed, detailed balance is really about the *energy difference*, which can be diagnosed by Fourier Transforming the Heisenberg evolution $\boldsymbol{A}^a(t) = \mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}$. Due to the Lieb-Robinson bounds, the localized Lindbladian terms effectively only depend on the local Hamiltonian patch nearby (up to exponentially decaying tail). (Right) This locality persists after purification, where two copies of the system are glued together.

In particular, the cost per unit Lindbladian evolution time is essentially *independent* of the system size (up to logarithmic dependencies), as we only need to simulate the Hamiltonian patch surrounding each jump $\boldsymbol{A}^a$.[6]

Purifying the Lindbladians

We may purify the Lindbladian to prepare the *purified Gibbs state* [180]

$$\left|\sqrt{\boldsymbol{\rho}_\beta}\right\rangle := \frac{1}{\sqrt{\mathrm{Tr}[\mathrm{e}^{-\beta\boldsymbol{H}}]}}\sum_i \mathrm{e}^{-\beta E_i/2}|\psi_i\rangle\otimes\left|\psi_i^*\right\rangle; \tag{3.9}$$

this particular purification is reminiscent of the quantum walk formalism for detailed-balanced classical Markov chains. The relevant "parent Hamiltonian" (playing the role of a "quantum walk" operator) is the *discriminant* associated with the Lindbladian; the expression may appear intimidating but resembles how classical detailed-balanced Markov chains are quantized

$$\mathcal{H}_\beta := \boldsymbol{\rho}^{-1/4}\mathcal{L}[\boldsymbol{\rho}^{1/4}\cdot\boldsymbol{\rho}^{1/4}]\boldsymbol{\rho}^{-1/4}.$$

The above should be regarded as the Lindbladian under similarity transformation. Further, to implement the quantum walk, the superoperator needs to be *vectorized* into an operator on duplicated Hilbert spaces

$$\text{(vectorization)}\quad \mathcal{H}_\beta : \mathcal{B}(\mathbb{C}^{2^n}) \to \mathcal{B}(\mathbb{C}^{2^n}) \simeq \boldsymbol{\mathcal{H}_\beta} \in \mathcal{B}(\mathbb{C}^{2^n}\otimes\mathbb{C}^{2^n}).$$

[6] In fact, we can further parallelize the Lindbladian evolution to improve the circuit depth; see section 3.7.

Indeed, the Gibbs state $\boldsymbol{\rho}_\beta$, as the Lindbladian stationary state, corresponds to the purified Gibbs state $\left|\sqrt{\boldsymbol{\rho}_\beta}\right\rangle$ (3.9), as a zero-eigenvector of the discriminant. As a sanity check, the quantum detailed balance condition 3.7 naturally ensures that the operator is Hermitian $\boldsymbol{\mathcal{H}}_\beta = \boldsymbol{\mathcal{H}}_\beta^\dagger$.

Proposition 3.1.1 (Purifying Lindbladians)**.** *Instantiate the Lindbladian parameters of Theorem 3.1.1 for the Gaussian or Metropolis transition weight. Then, the corresponding discriminant $\boldsymbol{\mathcal{H}}_\beta$ is Hermitian, frustration-free, and annihilates the purified Gibbs state.*

$$\boldsymbol{\mathcal{H}}_\beta = \sum_{a\in A} \boldsymbol{\mathcal{H}}_\beta^a \quad \textit{such that} \quad \boldsymbol{\mathcal{H}}_\beta^a \left|\sqrt{\boldsymbol{\rho}_\beta}\right\rangle = 0 \quad \textit{for each} \quad a \in A.$$

In other words, preparing the purified Gibbs state boils down to the *ground state* problem (up to a negative sign) for a *frustration-free* parent Hamiltonian parameterized by β[7]; for lattice Hamiltonian with local jumps $\boldsymbol{A}^a$, the parent Hamiltonian inherits the *quasi-locality* of our Lindbladian, with individual terms of radius $\sim v_{LR}\beta$. The algorithmic cost to implement the parent Hamiltonian is analogous to the Lindbladian case, where $\boldsymbol{H}_\beta$ roughly corresponds to a constant-time Lindbladian evolution.

Theorem 3.1.3 (Block encodings for the discriminants)**.** *Instantiate the Lindbladian parameters of Theorem 3.1.1 for either the Gaussian or Metropolis transition weight. Then, the corresponding discriminant $\frac{1}{\tilde{O}(1)}\boldsymbol{\mathcal{H}}_\beta$ can be block-encoded approximately in ϵ-spectral norm using*

$\tilde{O}(\beta)$ *(Hamiltonian simulation time)*

$\tilde{O}(1)$ *(resettable ancilla)*

$\tilde{O}(1)$ *(block-encodings for the jumps $\sum_{a\in A}|a\rangle \otimes \boldsymbol{A}^a$ and its transposes $\sum_{a\in A}|a\rangle \otimes (\boldsymbol{A}^a)^T$)*

$\tilde{O}(1)$ *(other two-qubit gates).*

The $\tilde{O}(\cdot)$ notation absorbs logarithmic dependencies on $t, \beta, \|\boldsymbol{H}\|, n, 1/\epsilon, |A|$.

Like in Theorem 3.1.2, we have expanded many other unitaries in the last line; see section 3.3. Note that we had to downscale the discriminant to implement the

[7] Strictly speaking, originating from a Lindbladian, here the parent Hamiltonian is negative semi-definite, and the purified Gibbs state is the top-eigenstate. Introducing a global negative sign will make it the ground state.

block encodings; see section 3.3 for the proof. For readers familiar with quantum walks, obtaining a block-encoding for $\boldsymbol{I} + \mathcal{H}_\beta$ (instead of $\mathcal{H}_\beta$) would be nicer as it leads to a quadratic speedup in terms of the spectral gap of $\mathcal{H}_\beta$. However, we have not found such a direct block-encoding for our particular construction, leaving this an open problem. This is in contrast with our earlier approximate construction [40, 180], where such a direct block-encoding was possible, which might be advantageous in some cases.

Roadmap

The remaining body of text is organized thematically by the analysis and the algorithm. We begin with the analysis (section 3.2) regarding how we design our Lindbladian to satisfy quantum detailed balance exactly. Next, we give efficient algorithms (section 3.3) in terms of modularized block encodings to implement the advertised Lindbladian and its purification.

In the appendix, we include independent expositions: a connection between the discriminant gap, mixing time, and the area law of entanglement (section 3.7) and an alternative heuristic derivation of our detailed balance Lindbladian in the time domain (section 3.8).

3.2 Analysis

In this section, we execute the calculations circling the exact detailed balance condition. First, we review the operator Fourier Transform in the frequency domain. Second, we review the notion of detailed balance, including the stationary state and spectral theory of convergence. Third, we plug in the advertised functional forms and derive the required coherent term $\boldsymbol{B}$ for achieving detailed balance.

Operator Fourier Transform

Our Lindbladian features the operator Fourier Transform $\hat{\boldsymbol{A}}(\omega)$ of a jump operator $\boldsymbol{A}$ according to the Hamiltonian $\boldsymbol{H}$ (dropping the jump label for this section). To analyze it, we need to consider the frequency domain representation instead of the time domain. Decompose the operator in the energy basis, and regroup in terms of the energy change (called the *Bohr frequencies* $\nu \in B(\boldsymbol{H}) := \{E_i - E_j \,|\, E_i, E_j \in \mathrm{Spec}(\boldsymbol{H})\}$)

$$\begin{aligned}\boldsymbol{A} &:= \sum_{E_1, E_2 \in \mathrm{spec}(\boldsymbol{H})} \boldsymbol{P}_{E_2} \boldsymbol{A} \boldsymbol{P}_{E_1} = \sum_{\nu \in B(\boldsymbol{H})} \sum_{E_2 - E_1 = \nu} \boldsymbol{P}_{E_2} \boldsymbol{A} \boldsymbol{P}_{E_1} \\ &=: \sum_{\nu \in B(\boldsymbol{H})} \boldsymbol{A}_\nu \quad \text{such that} \quad (\boldsymbol{A}_\nu)^\dagger = (\boldsymbol{A}^\dagger)_{-\nu},\end{aligned}$$

where $\boldsymbol{P}_E$ denotes the orthogonal projector onto the eigensubspace of $\boldsymbol{H}$ with *exact* energy E. This decomposition naturally solves the Heisenberg evolution

$$\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} = \sum_{\nu\in B}\boldsymbol{A}_\nu \mathrm{e}^{\mathrm{i}\nu t} \quad \text{since} \quad [\boldsymbol{H}, \boldsymbol{A}_\nu] = \nu\boldsymbol{A}_\nu.$$

Indeed, the energy differences $\nu \in B$ (shorthand of $B(\boldsymbol{H})$) naturally arise from the commutator (as opposed to the absolute energies $E \in \mathrm{spec}(\boldsymbol{H})$).

At first glance, the reference to the exact energies seems unphysical as each of them individually requires a long (likely exponential) Hamiltonian simulation time to access algorithmically. Fortunately, all that we are manipulating are the *smooth* weights on these Bohr frequencies; indeed, the operator Fourier Transform can be conveniently expressed by

$$\hat{\boldsymbol{A}}_f(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}\mathrm{e}^{-\mathrm{i}\omega t}f(t)\mathrm{d}t = \sum_{\nu\in B}\boldsymbol{A}_\nu \hat{f}(\omega-\nu),$$

where $\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t$ is the Fourier Transform of the filter function $f(t)$. Our choice of $f(t)$ is

$$f(t) := \mathrm{e}^{-\sigma_E^2 t^2}\sqrt{\sigma_E\sqrt{2/\pi}} \quad \text{such that}$$
$$\hat{f}(\omega) = \frac{1}{\sqrt{2\sigma_E\sqrt{2\pi}}}\exp\left(-\frac{\omega^2}{4\sigma_E^2}\right) \quad \text{and} \quad \int_{-\infty}^{\infty}|f(t)|^2\mathrm{d}t = 1,$$

therefore $\hat{\boldsymbol{A}}(\omega)$ becomes simply

$$\hat{\boldsymbol{A}}(\omega) = \frac{1}{\sqrt{2\sigma_E\sqrt{2\pi}}}\sum_{\nu\in B}\exp\left(-\frac{(\omega-\nu)^2}{4\sigma_E^2}\right)\boldsymbol{A}_\nu.$$

The above uses the Gaussian integrals, which will also be constantly recalled.

Fact 3.2.1 (Gaussian integrals)**.** *For any $b \in \mathbb{C}$ and $\sigma > 0$, we have that* $\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{(\omega-b)^2}{2\sigma^2}}\mathrm{d}\omega = \sqrt{2\pi}\sigma$.

We can think of the width σ_E as the uncertainty in energy, which scales inversely proportionally to the time width $\sim \sigma_E^{-1}$.

Exact detailed balance from that of the transition part

Our notion of detailed balance for Lindbladians is analogous to its classical cousin, ensuring a stationary state $\boldsymbol{\rho}$. For the mathematical audience, we should mention that other forms of quantum detailed balance have also been studied (see section 3.9), but we will dominantly focus on the following as it appears to be especially nice.

Definition 3.2.1 (Kubo–Martin–Schwinger detailed balance condition)**.** *For a normalized, full-rank state $\rho \succ 0$, we say that an super-operator $\mathcal{L}$ satisfies ρ-detailed balance (or ρ-DB in short) if*

$$\mathcal{L}^\dagger[\cdot] = \rho^{-1/2}\mathcal{L}[\rho^{1/2} \cdot \rho^{1/2}]\rho^{-1/2},$$

or equivalently, whenever the associated discriminant is self-adjoint with respect to ρ, i.e.,

$$\begin{aligned}\mathcal{D}(\rho, \mathcal{L}) &:= \rho^{-1/4}\mathcal{L}[\rho^{1/4} \cdot \rho^{1/4}]\rho^{-1/4} \\ &= \rho^{1/4}\mathcal{L}^\dagger[\rho^{-1/4} \cdot \rho^{-1/4}]\rho^{1/4} = \mathcal{D}(\rho, \mathcal{L})^\dagger.\end{aligned}$$

In the above, $(\mathcal{L})^\dagger$ denotes the adjoint for superoperators with respect to trace (i.e., the Hilbert Schmidt inner product).

One may interpret the conjugation with the state as a similarity transformation under which the Lindbladian becomes Hermitian (w.r.t. to the KMS inner product). The above generalizes classical detailed balance by considering super-operators and permitting the stationary distribution to be an operator.

Proposition 3.2.1 (Fixed point)**.** *If a Lindbladian $\mathcal{L}$ is ρ-detailed-balanced, then*

$$\mathcal{L}[\rho] = 0.$$

Recently, *quantum approximate detailed balance* has also been studied in the precursor of this work [40, Section II.A], discussing nonasymptotic error bounds relating mixing times to fixed point error. Exact detailed balance gives a much simpler conceptual picture. Still, we may again need to recall approximate detailed balance for non-fine-tuned Lindbladians (such as those from Nature) or amid intermediate steps of analysis (such as when truncating the radius of the quasi-local jumps).

At first glance, the detailed balance condition is merely a linear equation that can be solved abstractly. However, the difficulty arises due to two additional constraints.

- (**Complelete Positivity**.) Lindbladians have a particular *quadratic* dependence on the Lindblad operators to ensure complete positivity and trace preservation of $e^{t\mathcal{L}}$ for any t.
- (**Efficiency**.) The Lindbladian (i.e., the block-encoding for the jumps and the Hamiltonian) must be efficiently implemented using a limited Hamiltonian simulation time.

The main challenge is to satisfy both constraints simultaneously. Indeed, Davies' generator [56, 58] satisfies the first but not the second because it uses an infinite-time operator Fourier Transform; using QSVT, one might be able directly to implement the Boltzmann weight smoothly at moderate costs, but it may break the Lindbladian structure. Our approach begins by isolating the "transition" part of the Lindbladian (3.1) with abstract Lindblad operators $\boldsymbol{L}_j$

$$\mathcal{L}[\cdot] := -\mathrm{i}[\boldsymbol{B}, \cdot] + \underbrace{\sum_j \boldsymbol{L}_j \cdot \boldsymbol{L}_j^\dagger}_{\mathcal{T}:=} - \underbrace{\frac{1}{2}\{\sum_j \boldsymbol{L}_j^\dagger \boldsymbol{L}_j, \cdot\}}_{\boldsymbol{R}:=},$$

where $\boldsymbol{B}$ and $\boldsymbol{R}$ are both Hermitian. This decomposition is helpful because conjugating with the stationary state preserves the form of the transition part

$$\sqrt{\boldsymbol{\rho}}^{-1}\left(\sum_j \boldsymbol{L}_j(\sqrt{\boldsymbol{\rho}} \cdot \sqrt{\boldsymbol{\rho}})\boldsymbol{L}_j^\dagger\right)\sqrt{\boldsymbol{\rho}}^{-1} = \sum_j \boldsymbol{L}_j'(\cdot)\boldsymbol{L}_j'^\dagger.$$

However, the commutator $\boldsymbol{B}$ and anti-commutator terms $\boldsymbol{R}$ *mix* with each other under conjugation with Gibbs state. Based on the above observation, our recipe for constructing a detailed-balanced Lindbladian consists of three steps:

1. *Guess* a set of Lindblad operators $\boldsymbol{L}_j$ such that the transition part (which is $\mathcal{T} = \sum_a \int_{-\infty}^{\infty} \gamma(\omega)\hat{\boldsymbol{A}}^a(\omega) \cdot \hat{\boldsymbol{A}}^a(\omega)^\dagger \mathrm{d}\omega$ in our case) obeys detailed balance.

2. According to the transition part $\mathcal{T}$, determine the decay part parameterized by $\boldsymbol{R}$. This gives a purely dissipative Lindbladian.

3. According to the decay part $\boldsymbol{R}$, tailor the commutator term $\boldsymbol{B}$ to ensure detailed balance. Remarkably, such a $\boldsymbol{B}$ always exists, can be found explicitly, and is essentially *unique*. Of course, whether the map is efficiently implementable is a separate question.[8]

To simplify the presentation, we introduce the following notation for conjugating any full-rank state $\boldsymbol{\rho}$:

$$\Gamma_{\boldsymbol{\rho}}[\cdot] := \boldsymbol{\rho}^{1/2}(\cdot)\boldsymbol{\rho}^{1/2} \quad \text{and} \quad \Lambda_{\boldsymbol{\rho}}[\cdot] := \boldsymbol{\rho}^{-1/2}(\cdot)\boldsymbol{\rho}^{1/2}.$$

Observe that for Hermitian operator $\boldsymbol{X}$, we have the identities $\Gamma_{\boldsymbol{\rho}}[\boldsymbol{X}]^\dagger = \Gamma_{\boldsymbol{\rho}}[\boldsymbol{X}]$ and $\Lambda_{\boldsymbol{\rho}}[\boldsymbol{X}]^\dagger = \Lambda_{\boldsymbol{\rho}}^{-1}[\boldsymbol{X}]$. When the context is clear, we will omit subscript $\boldsymbol{\rho}$.

[8] This is inspired by a related ongoing work [78] at an early stage.

The main calculation of this section is summarized as follows. For any $\boldsymbol{R}$, we can give a general solution for the coherent term $\boldsymbol{B}$; this calculation is possible because the coherent term is not constrained by the complete-positivity structure of Lindbladians and only needs to be Hermitian.

Lemma 3.2.1 (Prescribing the coherent term)**.** *For any full-rank state $\boldsymbol{\rho}$ and Hermitian operator $\boldsymbol{R}$, there exists a unique Hermitian operator $\boldsymbol{B}$ (up to adding any scalar multiples of the identity $\boldsymbol{I}$) such that the super-operator*

$$\mathcal{S}[\cdot] := -\mathrm{i}[\boldsymbol{B}, \cdot] - \frac{1}{2}\{\boldsymbol{R}, \cdot\}$$

satisfies $\boldsymbol{\rho}$-DB.[9] *For a Gibbs state $\boldsymbol{\rho} \propto \exp(-\beta \boldsymbol{H})$, we can express the solution decomposed according to the Bohr frequencies $\nu \in B$ as*

$$\boldsymbol{B} = \frac{\mathrm{i}}{2} \sum_{\nu \in B} \tanh\left(\frac{\beta\nu}{4}\right) \boldsymbol{R}_\nu.$$

The above can be applied to a purely dissipative Lindbladian, where the transition part already satisfies $\boldsymbol{\rho}$-DB.

Corollary 3.2.1 ($\boldsymbol{\rho}$-DB Lindbladians)**.** *Suppose we have a purely-dissipative $\mathcal{L}_{diss}$ Lindbladian such that the transition part satisfies $\boldsymbol{\rho}$-DB for a full-rank state $\boldsymbol{\rho}$*

$$\Gamma_{\boldsymbol{\rho}}^{-1} \circ \mathcal{T} \circ \Gamma_{\boldsymbol{\rho}}^{1} = \mathcal{T}^\dagger,$$

then we can accordingly prescribe $\boldsymbol{B}$ such that $-\mathrm{i}[\boldsymbol{B}, \cdot] + \mathcal{L}_{diss}$ satisfies $\boldsymbol{\rho}$-DB.

Proof of Lemma 3.2.1. Let $\boldsymbol{K} := \boldsymbol{B} - \frac{\mathrm{i}}{2}\boldsymbol{R}$ and observe that

$$\mathcal{S}[\cdot] := -\mathrm{i}[\boldsymbol{B}, \cdot] - \frac{1}{2}\{\boldsymbol{R}, \cdot\} = -\mathrm{i}\boldsymbol{K}(\cdot) + \mathrm{i}(\cdot)\boldsymbol{K}^\dagger$$
$$\mathcal{S}^\dagger[\cdot] := \mathrm{i}[\boldsymbol{B}, \cdot] - \frac{1}{2}\{\boldsymbol{R}, \cdot\} = \mathrm{i}\boldsymbol{K}^\dagger(\cdot) - \mathrm{i}(\cdot)\boldsymbol{K}$$

using that $\boldsymbol{B}$ and $\boldsymbol{R}$ are Hermitian. Then,

$$\begin{aligned}
\mathcal{S}^\dagger[\cdot] - \Gamma^{-1} \circ \mathcal{S} \circ \Gamma[\cdot] &= \mathrm{i}\boldsymbol{K}^\dagger(\cdot) - \mathrm{i}(\cdot)\boldsymbol{K} + \mathrm{i}\Gamma^{-1}\left(\boldsymbol{K}\Gamma[\cdot] - \Gamma[\cdot]\boldsymbol{K}^\dagger\right) \\
&= \mathrm{i}\left(\boldsymbol{K}^\dagger + \Lambda[\boldsymbol{K}]\right)(\cdot) - \mathrm{i}(\cdot)\left(\boldsymbol{K} + (\Lambda[\boldsymbol{K}])^\dagger\right) \\
&\qquad\qquad \text{(using that } \Gamma^{-1}\boldsymbol{K}\Gamma[\cdot] = \Lambda[\boldsymbol{K}](\cdot)\text{)} \\
&=: \mathrm{i}(\boldsymbol{Q}(\cdot) - (\cdot)\boldsymbol{Q}^\dagger),
\end{aligned}$$

[9] Actually, our proof shows an even stronger statement: for any Hermitian $\boldsymbol{F}$ commuting with $\boldsymbol{\rho}$ there is a unique $\boldsymbol{B}$ (up to an additive term proportional to $\boldsymbol{I}$) such that $\mathcal{S}^\dagger[\cdot] - \Gamma^{-1} \circ \mathcal{S} \circ \Gamma[\cdot] = -\mathrm{i}[\boldsymbol{F}, \cdot]$. The only change is that one should set $\boldsymbol{B}_0 := -\frac{1}{2}\boldsymbol{F}$. This relates to a more general notion of detailed balance that allows for a unitary drift Definition 3.9.1, see also [69, Section 5]. This might be useful for breaking degeneracies of the state $\boldsymbol{\rho}$ (or the Hamiltonian $\boldsymbol{H}$).

where we define $\boldsymbol{Q} := \boldsymbol{K}^\dagger + \Lambda[\boldsymbol{K}]$. To ensure quantum detailed balance, we need the RHS to vanish

$$\boldsymbol{Q}(\cdot) - (\cdot)\boldsymbol{Q}^\dagger = 0 \qquad \Longleftrightarrow \qquad \boldsymbol{Q} = 0 + \lambda \boldsymbol{I} \quad \text{for} \quad \lambda \in \mathbb{R},$$

that is, $\boldsymbol{Q}$ vanishes up to a real multiple of the identity $\lambda \boldsymbol{I}$; for simplicity, the identity part can be dropped for now and added back. Since $\boldsymbol{\rho}$ has full rank, we can assume without loss of generality that $\boldsymbol{\rho} \propto \exp(-\beta \boldsymbol{H})$ for some $\boldsymbol{H}$. Now we compute

$$\begin{aligned}
\boldsymbol{Q} &= \boldsymbol{K}^\dagger + \Lambda[\boldsymbol{K}] \\
&= (1 + \Lambda)\boldsymbol{B} + \frac{\mathrm{i}}{2}(1 - \Lambda)\boldsymbol{R} \\
&= \sum_{\nu \in B} (1 + \mathrm{e}^{\beta\nu/2})\boldsymbol{B}_\nu + \frac{\mathrm{i}}{2}(1 - \mathrm{e}^{\beta\nu/2})\boldsymbol{R}_\nu., \qquad \text{(using that } \Lambda(\boldsymbol{B}_\nu) = \mathrm{e}^{\beta\nu/2}\boldsymbol{B}_\nu)
\end{aligned}$$

where we denote $B := B(\boldsymbol{H})$ the set of Bohr frequencies of $\boldsymbol{H}$. Since the operators $\boldsymbol{B}_\nu$ (and $\boldsymbol{R}_\nu$) are linearly independent for different Bohr frequencies,

$$\boldsymbol{Q} = 0 \qquad \Longleftrightarrow \qquad \boldsymbol{B}_\nu = \frac{\mathrm{i}}{2}\tanh\left(\frac{\beta\nu}{4}\right)\boldsymbol{R}_\nu \quad \text{for each} \quad \nu \in B.$$

$$\text{(using that } \tanh(x) = \tfrac{\mathrm{e}^{2x}-1}{\mathrm{e}^{2x}+1})$$

In particular, $\boldsymbol{B}_0 = 0$ since $\tanh(0) = 0$.

Finally, since $\boldsymbol{R}$ is Hermitian and $\tanh(\omega)$ is an odd function, we have that $\boldsymbol{B}$ is Hermitian as well by

$$\begin{aligned}
\sum_{\nu \in B} (\boldsymbol{B}_\nu)^\dagger &= \frac{\mathrm{i}}{2} \sum_{\nu \in B} \tanh\left(-\frac{\beta\nu}{4}\right)(\boldsymbol{R}_\nu)^\dagger. && \text{(Using } -\tanh(x) = \tanh(-x)) \\
&= \frac{\mathrm{i}}{2} \sum_{\nu \in B} \tanh\left(-\frac{\beta\nu}{4}\right)\boldsymbol{R}_{-\nu} && \text{(Using that } \boldsymbol{R} = \boldsymbol{R}^\dagger \text{ implies } (\boldsymbol{R}_\nu)^\dagger = \boldsymbol{R}_{-\nu}) \\
&= \frac{\mathrm{i}}{2} \sum_{\nu \in B} \tanh\left(\frac{\beta\nu}{4}\right)\boldsymbol{R}_\nu = \sum_{\nu \in B} \boldsymbol{B}_\nu. && \text{(Change of variables } \nu \to -\nu) \quad \blacksquare
\end{aligned}$$

Adding the identity part $\lambda \boldsymbol{I}$ to conclude the proof.

Exact detailed balance for Gaussian weights and their linear combinations

In this section, we carry out the abstract recipe for our advertised Lindbladian to prove the exact detailed balance condition (Definition 3.2.1). We begin with the transition part and then solve for the coherent term.

Exact detailed balance of the transition part

First, we show that the Gaussian ansatz indeed leads to detailed balance for the "transition" part

$$\Gamma_{\boldsymbol{\rho}}^{-1} \circ \mathcal{T} \circ \Gamma_{\boldsymbol{\rho}} = \mathcal{T}^{\dagger}.$$

It is instructive to rewrite the abstract equation above in terms of Bohr-frequencies. Let $(\cdot)$ be any input matrix, then the transition part of our Lindbaldian reads

$$\begin{aligned}
\mathcal{T} &= \sum_{a\in A} \int_{-\infty}^{\infty} \gamma(\omega) \hat{\boldsymbol{A}}^a(\omega)(\cdot)\hat{\boldsymbol{A}}^a(\omega)^{\dagger} \mathrm{d}\omega \\
&= \frac{1}{2\sigma_E\sqrt{2\pi}} \sum_{a\in A} \sum_{\nu_1,\nu_2\in B} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(\omega+\omega_\gamma)^2}{2\sigma_\gamma^2}} \mathrm{e}^{-\frac{(\omega-\nu_1)^2}{4\sigma_E^2}} \mathrm{e}^{-\frac{(\omega-\nu_2)^2}{4\sigma_E^2}} \boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^{\dagger} \mathrm{d}\omega \\
&=: \sum_{a\in A} \sum_{\nu_1,\nu_2\in B} \alpha^{(\omega_\gamma,\sigma_\gamma)}_{\nu_1,\nu_2} \boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^{\dagger}. \qquad (3.11)
\end{aligned}$$

Since the bilinear expression holds for any input $(\cdot)$, taking traces on both sides yields that

$$\sum_{a\in A} \int_{-\infty}^{\infty} \gamma(\omega) \hat{\boldsymbol{A}}^a(\omega)^{\dagger} \hat{\boldsymbol{A}}^a(\omega) \mathrm{d}\omega = \sum_{a\in A} \sum_{\nu_1,\nu_2\in B} \alpha^{(\omega_\gamma,\sigma_\gamma)}_{\nu_1,\nu_2} (\boldsymbol{A}^a_{\nu_2})^{\dagger} \boldsymbol{A}^a_{\nu_1}.$$

In terms of Bohr frequencies, the exact detailed balance condition is a certain symmetry of the coefficient matrix $\alpha^{(\omega_\gamma,\sigma_\gamma)}$.

Proposition 3.2.2 (Detailed balance in the Energy domain)**.** *Consider a super-operator parameterized by a Hamiltonian $\boldsymbol{H}$, β, and a set of operators including its adjoints $\{\boldsymbol{A}^a : a \in A\} = \{\boldsymbol{A}^{a\dagger} : a \in A\}$:*

$$\mathcal{T} = \sum_{a\in A} \sum_{\nu_1,\nu_2\in B} \alpha_{\nu_1,\nu_2} \boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^{\dagger} \quad \textit{such that} \quad \alpha_{\nu_1,\nu_2} = \alpha_{-\nu_2,-\nu_1} \mathrm{e}^{-\beta(\nu_1+\nu_2)/2}$$

$$\textit{for each} \quad \nu_1, \nu_2 \in B$$

Then,

$$\Gamma_{\boldsymbol{\rho}}^{-1} \circ \mathcal{T} \circ \Gamma_{\boldsymbol{\rho}}^{1} = \mathcal{T}^{\dagger}.$$

Indeed, one recovers the classical detailed balance condition for inputs diagonal in the energy basis. However, the quantum detailed balance condition also constrains the amplitudes between off-diagonal matrix elements (Figure 3.1).

Proof. We can directly calculate

$$\begin{aligned}
\Gamma_{\boldsymbol{\rho}}^{-1} \circ \mathcal{T} \circ \Gamma_{\boldsymbol{\rho}}^{1} &= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \alpha_{\nu_1,\nu_2}\sqrt{\boldsymbol{\rho}_\beta^{-1}}\boldsymbol{A}^a_{\nu_1}\sqrt{\boldsymbol{\rho}_\beta}(\cdot)\sqrt{\boldsymbol{\rho}_\beta}(\boldsymbol{A}^a_{\nu_2})^\dagger\sqrt{\boldsymbol{\rho}_\beta^{-1}} \\
&= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \alpha_{\nu_1,\nu_2}\mathrm{e}^{\frac{\beta}{2}\nu_1}\boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^\dagger\mathrm{e}^{\frac{\beta}{2}\nu_2} \qquad (\text{since } \boldsymbol{\rho}_\beta \propto \mathrm{e}^{-\beta \boldsymbol{H}}) \\
&= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \alpha_{-\nu_2,-\nu_1}\boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^\dagger \quad (\text{since } \alpha_{\nu_1,\nu_2}\mathrm{e}^{\frac{\beta(\nu_1+\nu_2)}{2}} = \alpha_{-\nu_2,-\nu_1}) \\
&= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \alpha_{-\nu_2,-\nu_1}((\boldsymbol{A}^{a\dagger})_{-\nu_1})^\dagger(\cdot)(\boldsymbol{A}^{a\dagger})_{-\nu_2} \\
&\qquad\qquad (\text{since } (\boldsymbol{A}_\nu)^\dagger = (\boldsymbol{A}^\dagger)_{-\nu}) \\
&= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \alpha_{-\nu_2,-\nu_1}(\boldsymbol{A}^a_{-\nu_1})^\dagger(\cdot)\boldsymbol{A}^a_{-\nu_2} \\
&\qquad\qquad (\text{since } \{\boldsymbol{A}^a : a\in A\} = \{\boldsymbol{A}^{a\dagger} : a\in A\}) \\
&= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \alpha_{\nu_2,\nu_1}(\boldsymbol{A}^a_{\nu_1})^\dagger(\cdot)\boldsymbol{A}^a_{\nu_2} \\
&\qquad\qquad (\text{since } B = \mathrm{spec}(\boldsymbol{H}) - \mathrm{spec}(\boldsymbol{H}) = -B) \\
&= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \alpha_{\nu_1,\nu_2}(\boldsymbol{A}^a_{\nu_2})^\dagger(\cdot)\boldsymbol{A}^a_{\nu_1} = \mathcal{T}^\dagger. \quad (\text{relabelling } \nu_1 \leftrightarrow \nu_2) \quad \blacksquare
\end{aligned}$$

While the above representation explicitly addresses the energy basis, we note that positivity becomes obscured as the left and right energy labels ν_1, ν_2 can differ. The positivity now becomes implicit in the coefficient matrix.

Proposition 3.2.3 (Positive semi-definite)**.** *If the coefficients α_{ν_1,ν_2} as a matrix $\boldsymbol{\alpha}$ is positive-semi-definite*

$$\boldsymbol{\alpha} \geq 0,$$

(and thus Hermitian $\alpha_{\nu_1,\nu_2} = \left(\alpha_{\nu_2,\nu_1}\right)^*$), then

$$\sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \alpha_{\nu_1,\nu_2}\left(\boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^\dagger - \frac{1}{2}\{(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1}, \cdot\}\right) \quad \textit{gives a Lindbladian.}$$

Positivity indeed holds for (3.11) and can be seen by the integral form of the coefficients. To conclude this section, it remains to verify that the coefficients arising from Gaussian indeed satisfy the symmetry.

Lemma 3.2.2 (Exact "skew-symmetry" of coefficients)**.** *For each $\omega_\gamma, \sigma_\gamma$, the coefficients $\alpha^{(\omega_\gamma,\sigma_\gamma)}_{\nu_1,\nu_2}$ defined by (3.11) factorize*

$$\alpha^{(\omega_\gamma,\sigma_\gamma)}_{\nu_1,\nu_2} = \frac{\sigma_\gamma}{2\sqrt{\sigma_E^2+\sigma_\gamma^2}} \cdot \exp\left(-\frac{(\nu_1+\nu_2+2\omega_\gamma)^2}{8(\sigma_E^2+\sigma_\gamma^2)}\right) \cdot \exp\left(-\frac{(\nu_1-\nu_2)^2}{8\sigma_E^2}\right), \quad (3.12)$$

and have a certain "skew-symmetry" under negation and transpose

$$\alpha^{(\omega_\gamma,\sigma_\gamma)}_{\nu_1,\nu_2} = \alpha^{(\omega_\gamma,\sigma_\gamma)}_{-\nu_2,-\nu_1} e^{-\beta(\nu_1+\nu_2)/2} \quad \textit{for} \quad \beta := \frac{2\omega_\gamma}{\sigma_E^2+\sigma_\gamma^2}. \tag{3.13}$$

Proof. We directly calculate the Gaussian integrals in (3.11), preserving the quadratic nature of the exponents.

$$\alpha^{(\omega_\gamma,\sigma_\gamma)}_{\nu_1,\nu_2} = \frac{1}{\sigma_E\sqrt{8\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{\omega^2}{2}\cdot\left(\frac{1}{\sigma_\gamma^2}+\frac{1}{\sigma_E^2}\right) - \omega\cdot\left(\frac{\omega_\gamma}{\sigma_\gamma^2}-\frac{\nu_1+\nu_2}{2\sigma_E^2}\right) - \frac{\nu_1^2+\nu_2^2}{4\sigma_E^2} - \frac{\omega_\gamma^2}{2\sigma_\gamma^2}\right) d\omega$$

$$= \frac{1}{\sigma_E\sqrt{8\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{\left(\omega + \left(\frac{1}{\sigma_\gamma^2}+\frac{1}{\sigma_E^2}\right)^{-1}\left(\frac{\omega_\gamma}{\sigma_\gamma^2}-\frac{\nu_1+\nu_2}{2\sigma_E^2}\right)\right)^2}{2\left(\frac{1}{\sigma_\gamma^2}+\frac{1}{\sigma_E^2}\right)^{-1}} + \frac{\left(\frac{1}{\sigma_\gamma^2}+\frac{1}{\sigma_E^2}\right)^{-1}}{2}\left(\frac{\omega_\gamma}{\sigma_\gamma^2}-\frac{\nu_1+\nu_2}{2\sigma_E^2}\right)^2 - \frac{\nu_1^2+\nu_2^2}{4\sigma_E^2} - \frac{\omega_\gamma^2}{2\sigma_\gamma^2}\right) d\omega$$

$$= \frac{1}{2\sigma_E}\cdot\frac{1}{\sqrt{\frac{1}{\sigma_\gamma^2}+\frac{1}{\sigma_E^2}}} \exp\left(\frac{1}{2\left(\frac{1}{\sigma_\gamma^2}+\frac{1}{\sigma_E^2}\right)}\left(\frac{\omega_\gamma}{\sigma_\gamma^2}-\frac{\nu_1+\nu_2}{2\sigma_E^2}\right)^2 - \frac{\nu_1^2+\nu_2^2}{4\sigma_E^2} - \frac{\omega_\gamma^2}{2\sigma_\gamma^2}\right)$$

(by Fact 3.2.1)

$$= \frac{1}{2\sqrt{\frac{\sigma_E^2}{\sigma_\gamma^2}+1}}\cdot\exp\left(-\frac{(\nu_1+\nu_2+2\omega_\gamma)^2}{8(\sigma_E^2+\sigma_\gamma^2)}\right)\cdot\exp\left(-\frac{(\nu_1-\nu_2)^2}{8\sigma_E^2}\right)$$

(since $2(\nu_1^2+\nu_2^2) = (\nu_1+\nu_2)^2 + (\nu_1-\nu_2)^2$)

$$= \alpha^{(\omega_\gamma,\sigma_\gamma)}_{-\nu_2,-\nu_1} e^{-\beta(\nu_1+\nu_2)/2}.$$

(due to our choice of β) ■

We see that we may tune the parameters to match a desirable exponent β. Remarkably, the width σ_E can be finite (i.e., do not scale linearly the precision $\frac{1}{\epsilon}$ like in metrology) while retaining exact detailed balance; a reasonable choice is, e.g.,

$$\omega_\gamma = \sigma_E = \sigma_\gamma = \frac{1}{\beta}.$$

This will imply that the algorithmic cost for implementing the Gaussian weighted operator Fourier Transform will only need to scale with β (and polylogarithmically with the precision due to discretization and truncation error!) However, the Gaussian transition weight comes with the price of a narrower band of transitions peaked at $\omega_\gamma \pm O(\sigma_\gamma)$, which might result in a substantially increased mixing time compared to, e.g., Metropolis weight $\gamma(\omega) = \min(1, e^{-\beta\omega})$; we will revisit this issue by taking

linear combinations of Gaussians at section 3.2 since all our calculation are linear. For clarity, we first focus on the Gaussian weights.

Why does Gaussian interplay so perfectly with quantum detailed balance? In section 3.8, we attempted to derive Gaussian from the first principle. Indeed, Gaussians are very natural if we impose several conditions on the function.

Adding the unitary term

Now that the "transition" part satisfies the detailed balance condition exactly, we proceed to complete the Lindbaldian by adding the "decay" part and the "coherent" part. The decay part is fixed by trace-preserving; Lemma 3.2.1 then uniquely prescribes the required coherent term, which we display as follows in the frequency domain. The explicit form will be useful for implementation.

Corollary 3.2.2 (An exactly detailed-balanced Lindbladian with Gaussian filtering)**.** *The Lindbladian*

$$\mathcal{L}[\cdot] := -\mathrm{i}\sum_{\nu\in B}[\boldsymbol{B}_\nu,\cdot] + \sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\alpha_{\nu_1,\nu_2}\left(\boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^\dagger - \frac{1}{2}\{(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1},\cdot\}\right)$$

corresponding to a self-adjoint set of jump operators $\{\boldsymbol{A}^a : a\in A\} = \{\boldsymbol{A}^{a\dagger} : a\in A\}$*, parametrized by coefficients* $\alpha_{\nu_1,\nu_2}\in\mathbb{C}$ *satisfying* $\alpha^*_{\nu_1,\nu_2} = \alpha_{\nu_2,\nu_1}$ *and*

$$\alpha_{\nu_1,\nu_2}\mathrm{e}^{\frac{\beta(\nu_1+\nu_2)}{2}} = \alpha_{-\nu_2,-\nu_1},$$

and amended by the coherent terms

$$\boldsymbol{B}_\nu := \sum_{a\in A}\sum_{\substack{\nu_1-\nu_2=\nu\\ \nu_1,\nu_2\in B}}\underbrace{\frac{\tanh(-\beta(\nu_1-\nu_2)/4)}{2\mathrm{i}}\alpha_{\nu_1,\nu_2}}_{\hat{f}(\nu_1,\nu_2):=}(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1} \tag{3.14}$$

satisfies $\boldsymbol{\rho}_\beta$*-detailed balance.*

Proof. Apply Lemma 3.2.1. Note that the operator

$$(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1} = \sum_{E_i-E_j=\nu_1,E_i-E_k=\nu_2}\boldsymbol{P}_{E_k}\boldsymbol{A}^{a\dagger}\boldsymbol{P}_{E_i}\boldsymbol{A}^a\boldsymbol{P}_{E_j}$$

must have the energy difference contained in the set of Bohr frequencies B

$$\begin{aligned}\nu_1-\nu_2 &= E_i-E_j-(E_i-E_k) \quad \text{for some} \quad E_i,E_j,E_k\in Spec(\boldsymbol{H})\\ &= E_k-E_i\in B.\end{aligned}$$

■

The above corollary essentially leads to Theorem 3.1.1 but is written in the Bohr-frequency decomposition.

Proof of Theorem 3.1.1. Combine Proposition 3.2.3, Lemma 3.2.2, Proposition 3.2.1, and Corollary 3.2.2 to conclude the proof. ∎

Linear combination of Gaussians

Can we go beyond Gaussians? As we discussed, the Gaussians have a narrower band of transitions; it would be desirable to lift this restriction to accelerate the mixing time. In this section, we give a family of filters by exploiting the freedom to tune the Gaussian parameters $(\omega_\gamma, \sigma_E, \sigma_\gamma)$ and taking a linear combination of Gaussians.

Corollary 3.2.3 (Linear combination of Gaussians)**.** *Fix σ_E and $g \in \ell_1(\mathbb{R})$ and set $\gamma^{(g)}(\omega) := \int_{\frac{\beta\sigma_E^2}{2}}^{\infty} g(x) e^{-\frac{(\omega+x)^2}{4x/\beta - 2\sigma_E^2}} dx$.* [10] *Then, analogous to (3.11), the coefficients as a linear combination over integration variable x*

$$\alpha^{(g)}_{\nu_1,\nu_2} := \int_{\frac{\beta\sigma_E^2}{2}}^{\infty} g(x) \alpha^{(\omega_\gamma,\sigma_\gamma)}_{\nu_1,\nu_2} dx \quad \text{for} \quad (\omega_\gamma(x), \sigma_\gamma(x)) = (x, \sqrt{2x/\beta - \sigma_E^2}) \tag{3.15}$$

satisfy the symmetries $\alpha^{(g)*}_{\nu_1,\nu_2} = \alpha^{(g)}_{\nu_2,\nu_1}$ and $\alpha^{(g)}_{\nu_1,\nu_2} = \alpha^{(g)}_{-\nu_1,-\nu_2} e^{-\beta(\nu_1+\nu_2)/2}$ for each $\nu_1, \nu_2 \in \mathbb{R}$. If $g(x) \geq 0$ for each x, then we also have that $\boldsymbol{\alpha}^{(g)} \geq 0$, however, this is not a necessary condition.

Proof. Recall the meaning of the superscripts $\alpha^{(\omega_\gamma,\sigma_\gamma)}_{\nu_1,\nu_2}$. The proof is merely the linearity of symmetries and that convex combination preserves the cone of positive semidefinite matrices. ∎

To widen the band of transitions, a natural choice is to weigh each Gaussian $e^{-\frac{(\omega+\omega_\gamma)^2}{2\sigma_\gamma^2}}$ with its inverse ℓ_1-weight. Surprisingly, this leads to filters that resemble the Metropolis and Glauber weights; while other choices are plausible, we spell out the calculation for this as a natural representative.

[10] Note that in principle we could also vary σ_E, but that would complicate both the analysis and the implementation due to the required adjustments to the parameters of the performed operator Fourier Transform.

Proposition 3.2.4 (Metropolis and Glauber-like filters)**.** *Setting* $g(\omega_\gamma) = \frac{1}{\sqrt{2\pi}\sigma_\gamma} = \frac{1}{\sqrt{2\pi(\frac{2\omega_\gamma}{\beta}-\sigma_E^2)}}$ *yields*

$$\gamma^{(s)}_{\sigma_E}(\omega) := \int_{\frac{\beta\sigma_E^2}{2}}^{\frac{\beta\sigma_E^2}{2}+\frac{s^2}{\beta}} \frac{\mathrm{e}^{-\frac{(\omega+x)^2}{4x/\beta-2\sigma_E^2}}}{\sqrt{2\pi(2x/\beta-\sigma_E^2)}}\mathrm{d}x$$
$$= \mathrm{e}^{-\beta\max\left(\omega+\frac{\beta\sigma_E^2}{2},0\right)}\cdot \underbrace{\frac{1}{2}\left[\left(\mathrm{erf}\left(\frac{s}{2}-\frac{\beta}{s}\left|\omega+\frac{\beta\sigma_E^2}{2}\right|\right)+1\right)+\mathrm{e}^{\beta\left|\omega+\frac{\beta\sigma_E^2}{2}\right|}\left(\mathrm{erf}\left(\frac{s}{2}+\frac{\beta}{s}\left|\omega+\frac{\beta\sigma_E^2}{2}\right|\right)-1\right)\right]}_{\leq 1},$$

which, in the $s\to\infty$ *limit, coincides with the Metropolis weight shifted by* $\frac{\beta\sigma_E^2}{2}$*:*

$$\gamma^{(\infty)}_{\sigma_E}(\omega) = \mathrm{e}^{-\beta\max\left(\omega+\frac{\beta\sigma_E^2}{2},0\right)}. \tag{3.16}$$

Restricting the above $g(\omega_\gamma)$ *to the interval* $\omega_\gamma \in \left(\frac{3\beta\sigma_E^2}{2},\infty\right)$ *results in the following smooth variant of* (3.16)

$$\tilde{\gamma}^{(\infty)}_{\sigma_E}(\omega) = \mathrm{e}^{-\beta\max\left(\omega+\frac{\beta\sigma_E^2}{2},0\right)} \tag{3.17}$$
$$\underbrace{\frac{1}{2}\left[\mathrm{erfc}\left(\frac{1}{\sigma_E}\left(\frac{\beta\sigma_E^2}{2}-\left|\omega+\frac{\beta\sigma_E^2}{2}\right|\right)\right)+\mathrm{e}^{\beta\left|\omega+\frac{\beta\sigma_E^2}{2}\right|}\mathrm{erfc}\left(\frac{1}{\sigma_E}\left(\frac{\beta\sigma_E^2}{2}+\left|\omega+\frac{\beta\sigma_E^2}{2}\right|\right)\right)\right]}_{\leq 1},$$

which resembles the Glauber filter also shifted by $\frac{\beta\sigma_E^2}{2}$ *(Figure 3.3).*

Proof. We directly found the above by Mathematica. ∎

3.3 Algorithms

This section presents efficient quantum algorithms for simulating the advertised Lindbladian and the associated parent Hamiltonian. The former mainly builds on black-box Lindbladian simulation algorithms [43, 49, 112], more precisely their improved variant described in [40, Theorem III.2] whose complexity boils down to constructing block encoding for the Lindblad operators (Definition 3.3.1); the latter merely requires block-encoding the parent Hamiltonian, which feeds into quantum

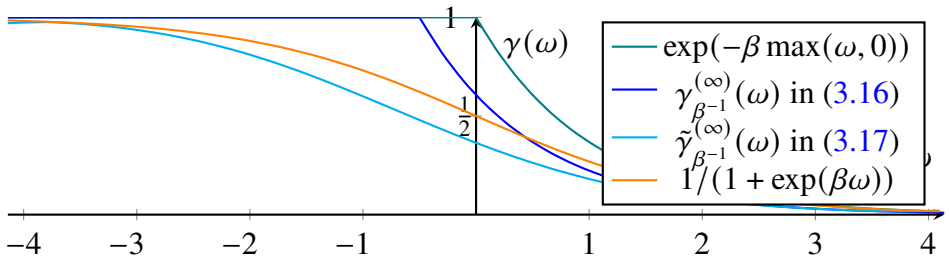


Figure 3.3: A plot of the filter functions $\gamma(\omega)$ for Metropolis, Glauber and our filters arising from Gaussian linear combination(3.16)-(3.17) (with $\sigma_E = \frac{1}{\beta}$).

simulated annealing (see [40, Appendix G] for a modern discussion) to prepare the purified Gibbs state.

Thus, the main algorithmic contribution is to assemble the block encodings associated with our synthetic Lindbladian (3.1) and our parent Hamiltonians (Theorem 3.1.3). The frequency domain representation (section 3.2), which is natural in the context of analyzing quantum detailed balance, is less instructive for algorithmic implementation. Indeed, addressing the exact energy eigenstates (or the exact Bohr frequencies) is generally inefficient. Nevertheless, the algorithmic task becomes straightforward in the time domain representation. Indeed, our Lindbladian can be expressed in terms of weighted time integrals $\int(\cdot)\mathrm{d}t$ of some rapidly decaying functions; a standard Linear-Combination-of-Unitary argument (under suitable discretization) leads to the algorithmic complexity in terms of controlled Hamiltonian simulation time.

For the Lindbladian (3.1), block-encodings for the dissipative part are already constructed in [40, Section III.B]. Thus, it remains to construct the coherent term; for the parent Hamiltonian, we will need to construct block encodings from scratch, but the manipulations are analogous. In the following sections, we first present the time-domain expressions (section 3.3-3.3), which immediately yield the corresponding block encodings (section 3.3) and the overall complexities (section 3.3-3.3).

Time-domain representation of our Lindbladians

Applying a two-dimensional Fourier Transform for the coherent term (3.14) leads to the following time-domain representation where LCU techniques are naturally applicable. See section 3.5 for the calculations.

Corollary 3.3.1 (Coherent term for the Gaussian case)**.** *For each $\beta > 0$ and parameters $\sigma_E = \sigma_\gamma = \omega_\gamma = \frac{1}{\beta}$, the coherent term $\boldsymbol{B}$ (3.14) corresponding to the*

Gaussian weight $\gamma(\omega) = \exp\left(-\frac{(\beta\omega+1)^2}{2}\right)$ *can be written as*

$$\boldsymbol{B} := \sum_{a\in A}\int_{-\infty}^{\infty} b_1(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\left(\int_{-\infty}^{\infty} b_2(t')\boldsymbol{A}^{a\dagger}(\beta t')\boldsymbol{A}^{a}(-\beta t')\mathrm{d}t'\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t, \quad (3.18)$$

where

$$b_1(t) := 2\sqrt{\pi}\mathrm{e}^{\frac{1}{8}}\left(\frac{1}{\cosh(2\pi t)} *_t \sin(-t)\exp\left(-2t^2\right)\right)$$

$$\textit{such that} \quad \|b_1\|_1 < 1 \quad (3.19)$$

$$b_2(t) := \frac{1}{2\pi}\sqrt{\frac{1}{\pi}}\exp\left(-4t^2 - 2\mathrm{i}t\right)$$

$$\textit{such that} \quad \|b_2\|_1 < \frac{1}{16}. \quad (3.20)$$

Indeed, we can verify that $\boldsymbol{B}$ is Hermitian by

$$\begin{aligned}
\boldsymbol{B}^\dagger &= \sum_{a\in A}\int_{-\infty}^{\infty} b_1^*(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\left(\int_{-\infty}^{\infty} b_2^*(t')\boldsymbol{A}^{a\dagger}(-\beta t')\boldsymbol{A}^{a}(\beta t')\mathrm{d}t'\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t \\
&= \sum_{a\in A}\int_{-\infty}^{\infty} b_1(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\left(\int_{-\infty}^{\infty} b_2(-t')\boldsymbol{A}^{a\dagger}(-\beta t')\boldsymbol{A}^{a}(\beta t')\mathrm{d}t'\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t \\
&\qquad\qquad \text{(By } b_1^*(t) = b_1(t) \text{ and } b_2^*(t) = b_2(-t)\text{)} \\
&= \sum_{a\in A}\int_{-\infty}^{\infty} b_1(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\left(\int_{-\infty}^{\infty} b_2(t')\boldsymbol{A}^{a\dagger}(\beta t')\boldsymbol{A}^{a}(-\beta t')\mathrm{d}t'\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t = \boldsymbol{B}. \\
&\qquad\qquad \text{(Change of variable } t' \to -t'\text{)}
\end{aligned}$$

The explicit weights corresponding to Metropolis weights are slightly more cumbersome due to taming a mild (logarithmic) singularity. We can also verify that $\boldsymbol{B}^{M,\eta}$ is Hermitian by $b_2^{M,\eta*}(t) = b_2^{M,\eta}(-t)$.

Corollary 3.3.2 (Approximate coherent term for the Metropolis-like weight)**.** *If* $\sigma_E = \frac{1}{\beta}$*, then the coherent term* $\boldsymbol{B}^M$ *corresponding to the Metropolis-like weight* $\gamma^M(\omega) = \exp\left(-\beta\max\left(\omega + \frac{1}{2\beta}, 0\right)\right)$ *satisfies*

$$\left\|\boldsymbol{B}^M - \boldsymbol{B}^{M,\eta}\right\| \le \left\|\sum_{a\in A}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^{a}\right\|\min\left(\frac{\eta\beta\|\boldsymbol{H}\|}{\sqrt{2}\pi}, O\left((\eta\beta\|\boldsymbol{H}\|)^3\right)\right), \quad (3.21)$$

where

$$\boldsymbol{B}^{M,\eta} := \sum_{a\in A}\int_{-\infty}^{\infty} b_1(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\left(\int_{-\infty}^{\infty} b_2^{M,\eta}(t')\boldsymbol{A}^{a\dagger}(\beta t')\boldsymbol{A}^{a}(-\beta t')\mathrm{d}t' + \frac{1}{16\sqrt{2}\pi}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^{a}\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t, \quad (3.22)$$

with $b_1(t)$ as in (3.19), and

$$b_2^{M,\eta}(t) := \frac{1}{4\sqrt{2}\pi} \frac{\exp(-2t^2 - \mathrm{i}t) + \mathbb{1}(|t| \le \eta)\mathrm{i}(2t+\mathrm{i})}{t(2t+\mathrm{i})}$$
$$\text{such that} \quad \|b_2^{M,\eta}\|_1 < \frac{1}{5} + \frac{1}{2\sqrt{2}\pi} \ln(1/\eta). \tag{3.23}$$

Further, if $\left[\sum_{a\in A} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a, \boldsymbol{H}\right] = 0$, we can drop the second term in (3.22) after the integral in t' since $\int_{-\infty}^{\infty} b_1(t) = 0$.

See section 3.6 for the proof. After suitable truncation, this merely incurs a *logarithmic* overhead ($\sim \log(\beta\|\boldsymbol{H}\|/\epsilon)$) to the algorithmic cost for an ϵ-approximation due to subnormalization $\frac{1}{\|b_2^{M,\eta}\|_1}$.

Time-domain representation of our parent Hamiltonians

Recall that the central mathematical object for coherent Gibbs sampling is the *discriminant* (i.e., the Lindbladian under a similarity transformation)

$$\mathcal{H}(\boldsymbol{\rho}, \mathcal{L}) := \boldsymbol{\rho}^{-1/4} \mathcal{L}[\boldsymbol{\rho}^{1/4} \cdot \boldsymbol{\rho}^{1/4}] \boldsymbol{\rho}^{-1/4}.$$

This amounts to a mild calculation that transfers the heavy lifting done already in the Lindbladidan context. One adaption for the coherent algorithm is that the fundamental object is not a superoperator but an operator on a doubled Hilbert space. Formally, we define the *vectorization* of a super-operator by

$$C[\cdot] = \sum_j \alpha_j \boldsymbol{A}_j [\cdot] \boldsymbol{B}_j \to \boldsymbol{C} = \sum_j \alpha_j \boldsymbol{A}_j \otimes \boldsymbol{B}_j^T, \quad \text{(vectorization)}$$

where $\boldsymbol{B}_j^T$ denotes the transpose of the matrix $\boldsymbol{B}_j$ in the computational basis $|i\rangle$. We use curly fonts $\mathbb{C}$ for super-operators and bold fonts $\mathbb{C}$ for the vectorized super-operators (which is, a matrix). For a matrix $\boldsymbol{\rho}$, let us denote its vectorized version by

$$|\boldsymbol{\rho}\rangle := (I \otimes T^{-1})\boldsymbol{\rho} \quad \text{where} \quad T|i\rangle = \langle i| \quad \text{for each} \quad |i\rangle.$$

In the time domain, our parent Hamiltonian takes the following form (see section 3.5 for the calculations):

$$\boldsymbol{\mathcal{H}}_\beta = \underbrace{\sum_{a\in A} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_-(t_-) h_+(t_+) \cdot \boldsymbol{A}^a(t_+ - t_-) \otimes \boldsymbol{A}^a(-t_- - t_+)^T \mathrm{d}t_+ \mathrm{d}t_-}_{\text{transition part}}$$
$$+ \underbrace{\frac{1}{2}(\boldsymbol{N} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{N}^*)}_{\text{decay and coherent part}}.$$

Compared with the Lindbladian case, the first term is essentially the transition part under a similarity transformation, and the second term combines the decay and coherent part. Here, we do not care about complete positivity or trace-preserving but merely Hermiticity.

The transition part

The transition parts for both Gaussian and Metropolis weights are as follows.

Corollary 3.3.3 (The transition part for Gaussian weights)**.** *For the Gaussian weight* $\gamma(\omega) = \exp\left(-\frac{(\beta\omega+1)^2}{2}\right)$ *with* $\omega_\gamma = \sigma_E = \sigma_\gamma = 1/\beta$*, the discriminant* $\boldsymbol{\mathcal{H}}_\beta$ *is described in the time domain by*

$$h_+(t) = \frac{1}{\beta}\mathrm{e}^{-1/4}\exp(-\frac{4t^2}{\beta^2}) \quad \textit{and} \quad h_-(t) = \frac{1}{\pi\beta}\exp(-\frac{2t^2}{\beta^2}) \quad \textit{such that} \quad \|h_-\|_1, \|h_+\|_1 \le 1. \tag{3.24}$$

Corollary 3.3.4 (The transition part for Metropolis weights)**.** *For* $\sigma_E = \frac{1}{\beta}$*, the Metropolis-like weight* $\gamma^M(\omega) = \exp\left(-\beta\max\left(\omega + \frac{1}{2\beta}, 0\right)\right)$ *yields* $\boldsymbol{\mathcal{H}}_\beta$ *described in the time domain by the same* $h_-(t)$ *as in Eq.* (3.24) *and by*

$$h_+(t) = \frac{\mathrm{e}^{-1/8}}{\beta}\frac{\mathrm{e}^{-2t^2/\beta^2}}{4\sqrt{2\pi}(\frac{t^2}{\beta^2} + \frac{1}{16})} \quad \textit{such that} \quad \|h_+(t)\|_1 \le 1.$$

The $\boldsymbol{N}$-term

The time-domain presentation of the $\boldsymbol{N}$ term is as follows.

Corollary 3.3.5 ($\boldsymbol{N}$ term for Gaussian weights)**.** *For each* β*, the Gaussian weight* $\gamma(\omega) = \exp\left(-\frac{(\beta\omega+1)^2}{2}\right)$ *with* $\sigma_E = \sigma_\gamma = \omega_\gamma = \frac{1}{\beta}$ *corresponds to the discriminant where*

$$\boldsymbol{N} = \sum_{a\in A}\int_{-\infty}^{\infty} n_1(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\left(\int_{-\infty}^{\infty} n_2(t')\boldsymbol{A}^{a\dagger}(\beta t')\boldsymbol{A}^a(-\beta t')\mathrm{d}t'\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t,$$

where

$$n_1(t) := \frac{1}{4}\cdot 2\sqrt{\pi}\left(\frac{1}{\cosh(2\pi t)} *_t \exp\left(-2t^2\right)\right) \qquad \textit{such that} \quad \|n_1\|_1 = \frac{\pi}{4\sqrt{2}} < 1 \tag{3.25}$$

$$n_2(t) := 4\cdot 2\cdot\frac{1}{2\pi}\sqrt{\frac{1}{\pi}}\exp\left(-4t^2 - 2\mathrm{i}t\right) = 8\cdot b_2(t) \qquad \textit{such that} \quad \|n_2\|_1 < \frac{1}{2}$$

with b_2 *as in* (3.20)*.*

Corollary 3.3.6 (***N*** term for Metropolis weights)**.** *If* $\sigma_E = \frac{1}{\beta}$*, then the Metropolis-like weight* $\gamma^M(\omega) = \exp\left(-\beta \max\left(\omega + \frac{1}{2\beta}, 0\right)\right)$ *corresponds to the discriminant where* $\boldsymbol{N}^M$ *satisfies*

$$\left\|\boldsymbol{N}^M - \boldsymbol{N}^{M,\eta}\right\| \leq \sum_{a\in A}\left\|\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right\| \min\left(\frac{\eta\beta\|\boldsymbol{H}\|}{\sqrt{2}\pi}, O\left((\eta\beta\|\boldsymbol{H}\|)^3\right)\right),$$

where

$$\boldsymbol{N}^{M,\eta} := \sum_{a\in A}\int_{-\infty}^{\infty} n_1(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\left(\int_{-\infty}^{\infty} n_2^{M,\eta}(t')\boldsymbol{A}^{a\dagger}(\beta t')\boldsymbol{A}^a(-\beta t')\mathrm{d}t' + \frac{1}{16\sqrt{2}\pi}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t,$$

with $n_1(t)$ *as in* (3.25)*, and* $n_2^{M,\eta} = b_2^{M,\eta}$ *as in* (3.23)*.*

In the above cases, we can verify that $\boldsymbol{N}$ is Hermitian by $n_1^*(t) = n_1(t)$ and $n_2^*(t) = n_2(-t)$. For $\boldsymbol{N}^M$, we also use that $(\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a)^\dagger = \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a$.

Block-encodings

Our simulation algorithm extensively uses block encodings that are largely borrowed from [40]. This section aims to instantiate them to state the theorems appropriately, and the curious reader may refer to [40, Section III.B].

Definition 3.3.1 (Block-encoding for Lindbladian)**.** *Given a purely irreversible Lindbladian*

$$\mathcal{L}[\boldsymbol{\rho}] := \sum_{j\in J}\left(\boldsymbol{L}_j\boldsymbol{\rho}\boldsymbol{L}_j^\dagger - \frac{1}{2}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\boldsymbol{\rho} - \frac{1}{2}\boldsymbol{\rho}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\right),$$

we say that a unitary $\boldsymbol{U}$ *is a block-encoding for Lindblad operators* $\{\boldsymbol{L}_j\}_{j\in J}$ *if* [11]

$$(\langle 0^b| \otimes \boldsymbol{I}) \cdot \boldsymbol{U} \cdot (|0^c\rangle \otimes \boldsymbol{I}) = \sum_{j\in J}|j\rangle \otimes \boldsymbol{L}_j \quad \textit{for} \quad b \leq c \in \mathbb{Z}^+.$$

- Block-encoding $\boldsymbol{V}_{jump}$ of the jump operators $\boldsymbol{A}^a$ in the form of Definition 3.3.1:

$$(\langle 0^b| \otimes \boldsymbol{I}_a \otimes \boldsymbol{I}_{sys})] \cdot \boldsymbol{V}_{jump} \cdot (|0^c\rangle \otimes \boldsymbol{I}_{sys}) = \sum_{a\in A}|a\rangle \otimes \boldsymbol{A}^a.$$

  To implement the discriminant, we also assume access to a block-encoding $\boldsymbol{V}_{jumpT}$ for the partial transpose $\sum_{a\in A}|a\rangle \otimes (\boldsymbol{A}^a)^T$.

[11] In the first register, we could use any orthonormal basis. Sticking to computational basis elements $|j\rangle$ is just for ease of presentation. Intuitively, one can think about b as the number of ancilla qubits used for implementing the Lindblad operators $\boldsymbol{L}_j$, while typically $a - b \approx \log|J|$.

- Quantum Fourier Transform

$$\boldsymbol{QFT}_N : |\bar{t}\rangle \to \frac{1}{\sqrt{N}} \sum_{\bar{\omega} \in S_{\omega_0}} \mathrm{e}^{-\mathrm{i}\bar{\omega}\bar{t}} |\bar{\omega}\rangle.$$

  We use "bar" to denote variables taking discrete values. In particular, the Fourier frequencies $\bar{\omega}$ and times $\bar{t}$ are integer multiples of ω_0 and t_0 respectively such that

$$\omega_0 t_0 = \frac{2\pi}{N}, \quad \text{and} \quad S^{\lceil N \rfloor} := \left\{ -\lceil (N-1)/2 \rceil, \ldots, -1, 0, 1, \ldots, \lfloor (N-1)/2 \rfloor \right\},$$

$$\text{and} \quad S_{\omega_0}^{\lceil N \rfloor} := \omega_0 \cdot S^{\lceil N \rfloor}, \quad S_{t_0}^{\lceil N \rfloor} := t_0 \cdot S^{\lceil N \rfloor}.$$

- Controlled Hamiltonian simulation

$$\sum_{\bar{t} \in S_{t_0}} |\bar{t}\rangle\langle\bar{t}| \otimes \mathrm{e}^{\pm \mathrm{i}\bar{t}\boldsymbol{H}}.$$

- State preparation oracles for the Fourier Transform weights, acting on the frequency register

$$\boldsymbol{Prep}_f : \left|\bar{0}\right\rangle \to |f\rangle.$$

- Controlled filter for the Boltzmann factors acting on the frequency register and the Boltzmann weight register

$$\boldsymbol{W} := \sum_{\bar{\omega} \in S_{\omega_0}} \boldsymbol{Y}_{1-\gamma(\bar{\omega})} \otimes |\bar{\omega}\rangle\langle\bar{\omega}| \quad \text{where} \quad 0 \le \gamma(\bar{\omega}) \le 1 \quad \text{and} \quad \gamma(\bar{\omega}) = \gamma(-\bar{\omega})\mathrm{e}^{-\beta\bar{\omega}}.$$

- Single qubit Pauli-Y rotations

$$\boldsymbol{Y}_\theta := \begin{pmatrix} \sqrt{1-\theta} & -\sqrt{\theta} \\ \sqrt{\theta} & \sqrt{1-\theta} \end{pmatrix}.$$

- Reflection on b-qubits

$$\boldsymbol{R}_b := 2\left|0^b\right\rangle\left\langle 0^b\right| - \boldsymbol{I}_b.$$

To feed into the black-box Lindbladian simulation algorithm [40, Theorem III.2], we need block-encodings for the dissipative part and the coherent term; for the coherent Gibbs sampler, we need block-encoding for the discriminant $\mathcal{H}_\beta$, which we obtain by adding the transition part and the $\boldsymbol{N}$ part.

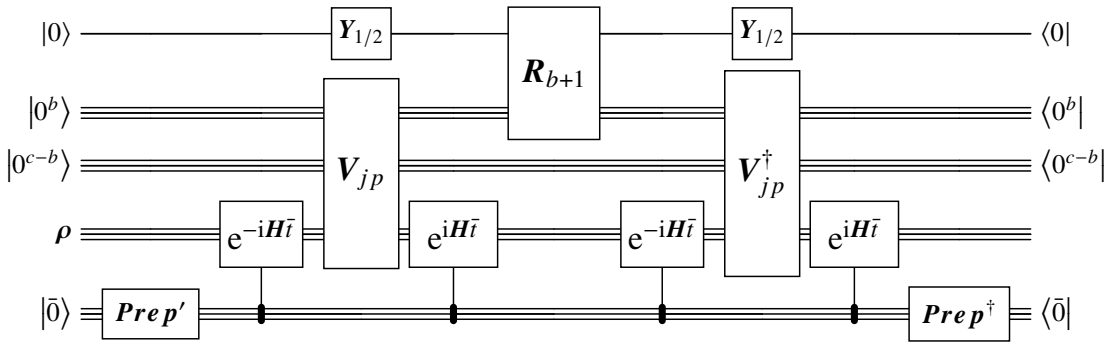


Figure 3.4: Circuit for block-encoding (3.3). The gate $\boldsymbol{prep}$ is a shorthand for $\boldsymbol{Prep}_{\sqrt{|f_+|}}$ and $\boldsymbol{prep}'$ for $\boldsymbol{Prep}_{f_+/\sqrt{|f_+|}}$.

Proposition 3.3.1 (Block-encoding for the coherent term)**.** *Suppose* $\|f_-\|_1, \|f_+\|_1 \leq 1$. *Then, there is a block-encoding for*

$$\sum_{\bar{t}_- \in S_{t_0}} f_-(\bar{t}_-) e^{-i\boldsymbol{H}\bar{t}_-} \left(\sum_{\bar{t}_+ \in S_{t_0}} f_+(\bar{t}_+) \sum_{a \in A} \boldsymbol{A}^{a\dagger}(\bar{t}_+) \boldsymbol{A}^a(-\bar{t}_+) \right) e^{i\boldsymbol{H}t_-}$$

using constant calls to controlled Hamiltonian simulation, $\boldsymbol{V}_{jump}$, $\boldsymbol{prep}_{\sqrt{|f_+|}}$, $\boldsymbol{prep}_{f_+/\sqrt{|f_+|}}$, $\boldsymbol{prep}_{\sqrt{|f_+|}}$, $\boldsymbol{prep}_{f_+/\sqrt{|f_+|}}$ *and their adjoints.*

The identical statement applies to the $\boldsymbol{N}$ term by replacing $f_\pm \to n_\pm$.

Proof. It suffices to construct

$$\sum_{\bar{t}_+ \in S_{t_0}} f_+(\bar{t}_+) \sum_{a \in A} \boldsymbol{A}^{a\dagger}(\bar{t}_+) \boldsymbol{A}^a(-\bar{t}_+)$$

and then apply the operator Fourier Transform. When there is only one jump ($|A| = 1$), this merely uses iterations of LCU and controlled Hamiltonian simulation. When a block-encoding gives the set of jumps, see Figure 3.4. To see that this yields the desired expression, observe that

$$\boldsymbol{Y}_{1/2}|0\rangle = |+\rangle \quad \text{and} \quad \boldsymbol{Y}_{1/2}^\dagger|0\rangle = |-\rangle.$$

Thus, the expression remains the same if we drop the $-\boldsymbol{I}_{b+1}$ term in $\boldsymbol{R}_{b+1}$. ∎

Proposition 3.3.2 (Bilinear)**.** *Suppose* $\|h_-\|_1 = \|h_+\|_1 = 1$. *Then, there is a block-encoding for*

$$\sum_{\bar{t}_- \in S_{t_0}} \sum_{\bar{t}_+ \in S_{t_0}} \sum_{a \in A} h_-(\bar{t}_-) h_+(\bar{t}_+) \cdot \boldsymbol{A}^a(\bar{t}_+ - \bar{t}_-) \otimes \boldsymbol{A}^a(-\bar{t}_- - \bar{t}_+)^T$$

using constant calls to controlled Hamiltonian simulation, $\boldsymbol{V}_{jump}$, $\boldsymbol{V}_{jumpT}$, $\boldsymbol{prep}_{\sqrt{|h_+|}}$, $\boldsymbol{prep}_{h_+/\sqrt{|h_+|}}$, $\boldsymbol{prep}_{\sqrt{|h_+|}}$, $\boldsymbol{prep}_{h_+/\sqrt{|h_+|}}$ *and their adjoints.*

Proof. The proof is a modification of Proposition 3.3.1 by replacing $\boldsymbol{V}_{jump}^{\dagger}$ with $\boldsymbol{V}_{jump}^{T}$ and duplicating the system register. ■

Proving Theorem 3.1.2: complexity for Lindbladian simulation

We now combine the block-encoding to give the overall cost of Gibbs sampling. We can implement the controlled time-evolution up to a truncation time as long as the profiles f_+, f_- are well-concentrated and smooth in the time domain (and that the Fourier Transforms are well-defined), which is the case since the frequency profiles $\hat{f}_+, \hat{f}_-$ are smooth and concentrated Corollary 3.3.1-Proposition 3.6.1. Also, the following results all require discretization of the time integrals, which is fortunately handled by [40, Appendix C]; this sets the required size of the Fourier Transform register, which uses polylogarithmically many qubits.

Proof of Theorem 3.1.2. Use the black-box Lindbladian simulation algorithm [40, Theorem III.2] for block-encoding for the coherent term (Proposition 3.3.1) and the dissipative part [40, Section III.B.1]. For the Metropolis weight, a logarithmic overhead is incurred for taming the mild singularity in the b_2^M; $\boldsymbol{B}$ is especially subnormalized by $\|b_2^M\|_1^{-1} = 1/O(\log(\beta\|H\|/\epsilon))$ to fit an approximation of $\boldsymbol{B}$ into a unitary block encoding. ■

Proving Theorem 3.1.3: complexity for the discriminant

We may now construct the advertised block encoding for the discriminant. The $\boldsymbol{N}$ term is analogous to the coherent term (Proposition 3.3.1).

Proof of Theorem 3.1.3. Add the block encodings for the transition part (from Proposition 3.3.2), $\boldsymbol{N} \otimes \boldsymbol{I}$, and $\boldsymbol{I} \otimes \boldsymbol{N}$ (from Proposition 3.3.1), up to mild subnormalization. Again, the Metropolis case incurs an additional logarithmic factor. ■

Note that we do not implement $\boldsymbol{I} + \mathcal{H}_\beta$ (as in [40, Proposition III.5]) but rather $\mathcal{H}_\beta$ itself; implementing the former would allow us to obtain a quadratic speedup on the discriminant gap, which we currently do not have.

3.4 Discussion

We have constructed the quantum analog of the classical Monte Carlo algorithms with desirable features. We highlight potential future directions as listed.

- **Quantum simulation applications.** A key factor in industrial quantum simulation applications [30, 37, 107] is effective quantum algorithms for low-energy states. Our algorithm can be employed for any Hamiltonian *without* substantial variational parameters or a case-by-case trial state or adiabatic path. While the mixing time can vary widely, the fact that physically relevant states (molecules or materials) *exist* in Nature suggests a reasonable mixing time in practice. Regarding practical gate complexities, the locality of our algorithm for lattice Hamiltonian may be favorable as we merely need to simulate a $\tilde{O}(\beta)$-radius Hamiltonian patch localized around each jump $\boldsymbol{A}^a$.

- **Locality and complexity of quantum Gibbs state.** Our algorithm opens new angles on the locality and complexity of Gibbs states (such as the decay of correlation, quantum conditional mutual information, recovery channels, and quantum belief propagation). In particular, the combination of localized jumps and exact detailed balance enables the rigorous study of convergence [31, 96] for noncommuting lattice Hamiltonians. Rapid mixing also directly implies the circuit complexity of the purification (Appendix 3.7) through the Lindbladian gap, giving a dynamic perspective on the area law of entanglement [84].

- **New open-system physics.** Just as quantum computing lacks a go-to Monte Carlo algorithm, open system physics lacks a simple, universal Lindbladian that succinctly captures open system thermodynamics. Our algorithm qualifies due to its elegant properties. For example, our Lindbladian enables a precise definition of *dynamical thermal phase transitions* in terms of mixing time that may contrast with static thermal phase transitions. Related concepts include metastable states, the energy landscape, *quantum spin glass*, and self-correcting quantum memories, whose precise formulation for noncommuting Hamiltonian has also been lacking.

- **A new algorithmic subroutine.** Classical MCMC algorithms have been widely employed to solve other problems beyond physical simulation, and we may expect the same for our algorithm. A natural example is optimization problems (e.g., constraint satisfaction problems and modern optimization problems), whether applying to classical Hamiltonian (in a setting similar

to Quantum Approximate Optimization Algorithms (QAOA) [66]) or quantum Hamiltonians. Another application is Quantum Semidefinite Program Solvers [11, 27], where Quantum Gibbs state preparation is routinely invoked.

- **Comparision with existing algorithms.** With a new algorithm at hand, we expect fruitful comparison with existing (quantum or classical) algorithms such as the adiabatic algorithm [64], phase estimation with trial states, tensor network, Quantum Monte Carlo, etc. In particular, understanding the distinction from classical algorithms could either inspire better classical algorithms or expose potential sources of quantum advantage in quantum simulation (e.g., the sign problem or difficulty in contracting PEPS).

- **Numerical studies.** As the complement to theory, the explicit form of our Lindbladian also enables direct numerical studies regarding the above notions, e.g., the scaling of mixing time for thermal state or ground states, dynamic phase transitions, and noncommuting quantum memories, and the interplay with tensor networks.

To conclude, given the celebrated theoretical and empirical triumph of Markov chain Monte Carlo methods and its successors over the past 70 years, we argue that this work should serve similar roles in quantum computing. Especially given the current skepticism on the practical applicability of quantum computers, our new algorithms bring hope to the community by initiating a new wave of directions covering theory, experiment, numerics, and application.

Acknowledgments

We thank Jonathan Moussa for raising the question of whether exact detailed balance is possible and pointing us to his related work [133] after the precursor of this work [40] became public; at that time, we thought that exact detailed balance is incompatible with the energy-time uncertainty principle. We also thank Aaron J. Friedman, Jinkang Guo, Oliver Hart, and Andrew Lucas for collaborations on related topics and Alvaro Alhambra, Anurag Anshu, Simon Apers, Mario Berta, Fernando Brandao, Angela Capel, Garnet Chan, ChatGPT-4, Alex Dalzell, Zhiyan Ding, Steve Flammia, Hsin-Yuan (Robert) Huang, Lin Lin, Yunchao Liu, Sam McArdle, Mehdi Soleimanifar, Frederik Nathan, Umesh Vazirani, and Tong Yu for helpful discussions. CFC was supported through an internship of the AWS Center for Quantum Computing. AG acknowledges funding from the AWS Center for Quantum Computing.

3.5 Appendix:Deriving time-domain represantations

Our calculation for detailed balance has focused on the frequency domain. This appendix applies Fourier transforms to obtain the time-domain representation. The arguments are conceptually straightforward but require some bookkeeping.

For both the Lindbladians and the parent Hamiltonians, we will often encounter a two-dimensional sum over Bohr frequencies. Since there are two energy labels, we employ a two-dimensional Fourier Transform. For any function of frequencies $\hat{f}(\nu_1, \nu_2)$, the time-domain representation of the bilinear expression gives

$$\sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\hat{f}(\nu_1,\nu_2)\boldsymbol{A}^a_{\nu_1}\otimes(\boldsymbol{A}^a_{\nu_2})^\dagger=\sum_{a\in A}\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(t_1,t_2)\boldsymbol{A}^a(t_1)\otimes\boldsymbol{A}^{a\dagger}(-t_2)\mathrm{d}t_1\mathrm{d}t_2,$$

where we introduced the two-dimensional Fourier Transform

$$f(t_1,t_2):=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{f}(\nu_1,\nu_2)\mathrm{e}^{\mathrm{i}\nu_1 t_1}\mathrm{e}^{\mathrm{i}\nu_2 t_2}\mathrm{d}\nu_1\mathrm{d}\nu_2.$$

Fortunately, for our usage, the Fourier Transform *decouples* into two iterations of one-dimensional Fourier Transforms, significantly simplifying the presentation and implementation.

Corollary 3.5.1 (Factorized time-domain functions)**.** *If the function factorizes in the energy domain such that*

$$\frac{1}{2\pi}\hat{f}(\nu_1,\nu_2)=\hat{f}_+(\nu_1+\nu_2)\cdot\hat{f}_-(\nu_1-\nu_2),$$

then

$$\begin{aligned}&\sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\hat{f}(\nu_1,\nu_2)\boldsymbol{A}^a_{\nu_1}\otimes(\boldsymbol{A}^a_{\nu_2})^\dagger\\&=\sum_{a\in A}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_-(t_-)f_+(t_+)\boldsymbol{A}^a(-t_--t_+)\otimes\boldsymbol{A}^{a\dagger}(t_+-t_-)\mathrm{d}t_+\mathrm{d}t_-\end{aligned}\tag{3.26}$$

and

$$\begin{aligned}&\sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\hat{f}(\nu_1,\nu_2)(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1}\\&=\sum_{a\in A}\int_{-\infty}^{\infty}f_-(t_-)\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t_-}\left(\int_{-\infty}^{\infty}f_+(t_+)\boldsymbol{A}^{a\dagger}(t_+)\boldsymbol{A}^a(-t_+)\mathrm{d}t_+\right)\mathrm{e}^{\mathrm{i}\boldsymbol{H}t_-}\mathrm{d}t_-,\end{aligned}$$

where the function $f_\pm$ are inverse Fourier Transforms of $\hat{f}_\pm$.

Crucially, the RHS can be implemented via Linear Combination of Unitaries by discretizing the integral.

Proof. Since the expression is linear in the sum over jumps $a \in A$, it suffices to prove for any operator $\boldsymbol{A}$, dropping the jump labels a.

$$\begin{aligned}
&\sum_{\nu_1,\nu_2\in B} 2\pi \hat{f}_+(\nu_1+\nu_2)\cdot \hat{f}_-(\nu_1-\nu_2)\boldsymbol{A}_{\nu_1}\otimes(\boldsymbol{A}_{\nu_2})^\dagger\\
&= \sum_{\nu_1,\nu_2\in B} 2\pi \hat{f}_+(\nu_+)\cdot \hat{f}_-(\nu_-)\left(\boldsymbol{A}_{\frac{\nu_++\nu_-}{2}}\otimes \boldsymbol{A}_{\frac{\nu_+-\nu_-}{2}}\right)^\dagger\\
&\hfill \text{(Let } \nu_+ := \nu_1+\nu_2 \text{ and} \nu_- := \nu_1-\nu_2)\\
&= \sum_{\nu_1,\nu_2\in B} \int_{-\infty}^{\infty} f_+(t_+)e^{-\mathrm{i}\nu_+t_+}\mathrm{d}t_+ \int_{-\infty}^{\infty} f_-(t_-)e^{-\mathrm{i}\nu_-t_-}\mathrm{d}t_- \boldsymbol{A}_{\frac{\nu_++\nu_-}{2}}\otimes\left(\boldsymbol{A}_{\frac{\nu_+-\nu_-}{2}}\right)^\dagger\\
&\hfill \text{(Fourier Transform)}\\
&= \sum_{\nu_1,\nu_2\in B} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_+(t_+)f_-(t_-)\boldsymbol{A}_{\frac{\nu_++\nu_-}{2}}(-t_+-t_-)\otimes\left(\boldsymbol{A}_{\frac{\nu_+-\nu_-}{2}}\right)^\dagger(t_+-t_-)\mathrm{d}t_+\mathrm{d}t_-\\
&\hfill \text{(Since } (\boldsymbol{A}_\nu)^\dagger = (\boldsymbol{A}^\dagger)_{-\nu})\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_+(t_+)f_-(t_-)\boldsymbol{A}(-t_+-t_-)\otimes \boldsymbol{A}^\dagger(t_+-t_-)\mathrm{d}t_+\mathrm{d}t_-.
\end{aligned}$$

The fourth equality uses that

$$\mathrm{e}^{-\mathrm{i}\nu_+t_+}\mathrm{e}^{-\mathrm{i}\nu_-t_-} = \exp\Big(\frac{\mathrm{i}(\nu_+-\nu_-)(t_+-t_-)}{2}\Big)\cdot\exp\Big(\frac{\mathrm{i}(\nu_++\nu_-)(-t_+-t_-)}{2}\Big)$$

to conclude the proof. ■

Now, we plug in the appropriate functions to arrive at the time-domain functions.

Our Lindbladians

We evaluate the Fourier transform for the coherent term $\boldsymbol{B}$. The expression looks intimidating, but all that matters for the algorithmic complexity is that they decay rapidly (in the time domain).

Corollary 3.5.2 (Explicit time-domain functions)**.** *In the time domain, the coherent term* $\boldsymbol{B}$ *in* (3.14) *corresponding to coefficients constructed in* (3.15) *reads*

$$f_+(t) = \int_{\frac{\beta\sigma_E^2}{2}}^{\infty} g(x)\sigma_\gamma(x)\exp\left(-\frac{4t^2x}{\beta}-2\mathrm{i}tx\right)\mathrm{d}x, \tag{3.27}$$

$$f_-(t) = \frac{\sigma_E}{\pi\beta}\mathrm{e}^{\frac{\beta^2\sigma_E^2}{8}}\left(\frac{1}{\cosh\left(\frac{2\pi t}{\beta}\right)} *_t \sin\left(-\beta\sigma_E^2 t\right)\mathrm{e}^{-2\sigma_E^2t^2}\right), \tag{3.28}$$

depending on parameters $\sigma_E, g(x), \beta$.[12]

Recall that the convolution of two functions over variable t is defined by

$$(f *_t g)(t) := \int_{-\infty}^{\infty} f(s)g(t-s)\mathrm{d}s.$$

Nicely, the product structure persists under a convex combination of Gaussians as only $f_+(t)$ depends on $g(x)$. Otherwise, we could have had to consider a linear combination of function products, which is messier to implement.

Proof. First, we confirm that the energy domain function indeed has a product structure $\frac{1}{2\pi}\hat{f}(\nu_1,\nu_2) = \hat{f}_+(\nu_1+\nu_2)\cdot\hat{f}_-(\nu_1-\nu_2)$ due to (3.12), (3.15) and (3.14) for

$$\hat{f}_+(\nu) = \int_{\frac{\beta\sigma_E^2}{2}}^{\infty} \frac{g(x)\sigma_\gamma(x)}{2\sqrt{\sigma_E^2+\sigma_\gamma^2(x)}} \exp\left(-\frac{(\nu+2x)^2}{16x/\beta}\right)\mathrm{d}x, \text{ and} \tag{3.29}$$

$$\hat{f}_-(\nu) = \frac{1}{2\pi}\frac{\tanh(-\beta\nu/4)}{2\mathrm{i}}\exp\left(-\frac{\nu^2}{8\sigma_E^2}\right) = \frac{1}{2\pi}\frac{1}{\cosh(-\beta\nu/4)}\cdot\frac{\sinh(-\beta\nu/4)}{2\mathrm{i}}\exp\left(-\frac{\nu^2}{8\sigma_E^2}\right).$$

Since we work with well-concentrated integrable functions, the Fourier Transforms exist, and we can compute them as follows. We begin with the Gaussian integral associated with f_+

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{(\nu+2x)^2}{16x/\beta}\right)\mathrm{e}^{\mathrm{i}\nu t}\mathrm{d}\nu = 2\sqrt{\frac{2x}{\beta}}\exp\left(-\frac{4t^2x}{\beta}-2\mathrm{i}tx\right).$$

Thus, using (3.29) and the definition of $\sigma_\gamma(x) = \sqrt{2x/\beta-\sigma_E^2}$, we get (3.27). In order to compute $f_-(t)$, we use the convolution theorem $\mathcal{F}^{-1}(\mathcal{F}(f)\cdot\mathcal{F}(g)) = f*g/\sqrt{2\pi}$. Individually, we have

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\sinh(-\beta\nu/4)}{2\mathrm{i}}\exp\left(-\frac{\nu^2}{8\sigma_E^2}\right)\mathrm{e}^{\mathrm{i}\nu t}\mathrm{d}\nu = \sigma_E\mathrm{e}^{\frac{\beta^2\sigma_E^2}{8}}\sin\left(-\beta\sigma_E^2 t\right)\mathrm{e}^{-2\sigma_E^2t^2} \tag{3.30}$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{2\pi\cosh(-\beta\nu/4)}\mathrm{e}^{\mathrm{i}\nu t}\mathrm{d}\nu = \frac{2}{\sqrt{2\pi}\beta\cosh\left(\frac{2\pi t}{\beta}\right)}.$$

[12] Note that the function $b_1(t)$ seems to have width $\sim\beta\sigma_E$ due to the convolution by $\frac{1}{\cosh(4\pi t/(\beta\sigma_E))}$, meaning that the integral in t seems to require Hamiltonian evolution times up to $\sim\beta$. In fact, the numerics show a $1/\mathrm{poly}$ decay until about $\sim\beta\sigma_E$ (after which the exponential decay starts), suggesting that about $\min\left(\sigma_E^{-1}\mathrm{Poly}(1/\epsilon), \beta\log(1/\epsilon)\right)$ Hamiltonian simulation time is required in order to achieve ϵ precision for the block-encoding of the coherent term. Thus, it might be difficult to obtain exact detailed balance below $\Omega(\beta)$ Hamiltonian evolution times. On the contrary, the function $b_2(t')$ has a width only about $1/\sqrt{\beta\omega_\gamma}$, implying that the corresponding other integral in t' can be well-approximated by only using Hamiltonian evolution time $\sim\sqrt{\beta/\omega_\gamma}$.

Take the convolution to conclude the calculation. ∎

The above explicit form allows us to compute the explicit form of the function $f_+(t)$ corresponding to our two main settings. We begin with the Gaussian case.

Corollary 3.3.1 (Coherent term for the Gaussian case)**.** *For each $\beta > 0$ and parameters $\sigma_E = \sigma_\gamma = \omega_\gamma = \frac{1}{\beta}$, the coherent term $\boldsymbol{B}$ (3.14) corresponding to the Gaussian weight $\gamma(\omega) = \exp\left(-\frac{(\beta\omega+1)^2}{2}\right)$ can be written as*

$$\boldsymbol{B} := \sum_{a\in A} \int_{-\infty}^{\infty} b_1(t) \mathrm{e}^{-\mathrm{i}\beta \boldsymbol{H} t} \left(\int_{-\infty}^{\infty} b_2(t') \boldsymbol{A}^{a^\dagger}(\beta t') \boldsymbol{A}^{a}(-\beta t') \mathrm{d}t' \right) \mathrm{e}^{\mathrm{i}\beta \boldsymbol{H} t} \mathrm{d}t, \quad (3.18)$$

where

$$b_1(t) := 2\sqrt{\pi} \mathrm{e}^{\frac{1}{8}} \left(\frac{1}{\cosh(2\pi t)} *_t \sin(-t) \exp\left(-2t^2\right) \right)$$

$$\textit{such that} \quad \|b_1\|_1 < 1 \quad (3.19)$$

$$b_2(t) := \frac{1}{2\pi} \sqrt{\frac{1}{\pi}} \exp\left(-4t^2 - 2\mathrm{i}t\right)$$

$$\textit{such that} \quad \|b_2\|_1 < \frac{1}{16}. \quad (3.20)$$

Indeed, both functions of time are rapidly decaying and have bounded ℓ_1-norm (as required for LCU implementation).

Proof. Setting $g(x) = \delta(x - \omega_\gamma)$ in Corollary 3.5.1 yields the desired Gaussian weight and $f_+(t)$ in (3.27) becomes

$$f_+(t) = \int_{-\infty}^{\infty} \delta(x - \omega_\gamma) \sigma_\gamma \exp\left(-\frac{4t^2 x}{\beta} - 2\mathrm{i}tx \right) dx = \sqrt{\frac{2\omega_\gamma}{\beta} - \sigma_E^2} \exp\left(-\frac{4t^2 \omega_\gamma}{\beta} - 2\mathrm{i}t\omega_\gamma \right).$$

Setting $b_1(t) := 2\pi\sqrt{\pi} f_-(t/\sigma_E)$ and $b_2(t') := f_+(\beta t')/(2\pi\sqrt{\pi})$ and applying a change of variables in the integral (3.26) yields the desired result (3.18). Note that the convolution $*_t$ implicitly is an integral over t, so we should not forget to rescale $\mathrm{d}t$ there.

Lastly, we bound the ℓ_1-norm of the functions by Hölder's inequality

$$\|b_1\|_1 \leq \left\|(1+t^2)^{-1}\right\|_2 \left\|(1+t^2) b_1\right\|_2 = \sqrt{\frac{\pi}{2}} \left\|(1+t^2) b_1\right\|_2 \leq 1,$$

using individual bounds $\int_{-\infty}^{\infty} \frac{1}{(1+2t^2)^2} \mathrm{d}t = \frac{\pi}{2}$ and $\int_{-\infty}^{\infty} \left|(1+t^2) b_1(t)\right|^2 \mathrm{d}t < 0.625$. The norm $\|b_2\|_1 = \frac{e^{-1/4}}{4\pi} < \frac{1}{16}$ is a Gaussian integral (Fact 3.2.1). ∎

The explicit weights corresponding to Metropolis weights are slightly more cumbersome due to taming a logarithmic singularity; see section 3.6.

Our parent Hamiltonians

Based on the Lindbladian, we explicitly evaluate the discriminant in the frequency domain.

Proposition 3.5.1 (Symmerized discriminant)**.** *In the setting of Corollary 3.2.2, the discriminant corresponding to the $\boldsymbol{\rho}_\beta$-DB Lindbladian reads*

$$\mathcal{H}_\beta = \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} h_{\nu_1,\nu_2}\boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^\dagger + \frac{1}{2}(\boldsymbol{N}(\cdot)+(\cdot)\boldsymbol{N})$$

$$\textit{or}\quad \boldsymbol{\mathcal{H}}_\beta = \sum_{a\in A}\sum_{\nu_1,\nu_2\in B} \underbrace{h_{\nu_1,\nu_2}\boldsymbol{A}^a_{\nu_1}\otimes(\boldsymbol{A}^a_{\nu_2})^*}_{\textit{transition part}} + \underbrace{\frac{1}{2}(\boldsymbol{N}\otimes\boldsymbol{I}+\boldsymbol{I}\otimes\boldsymbol{N}^*)}_{\textit{coherent and decay part}}$$

where $h_{\nu_1,\nu_2} := \mathrm{e}^{\beta(\nu_1+\nu_2)/4}\alpha_{\nu_1,\nu_2} = h_{-\nu_2,-\nu_1}$ *and*

$$\boldsymbol{N} := -\sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\frac{\alpha_{\nu_1,\nu_2}}{\cosh(\beta(\nu_1-\nu_2)/4)}(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1} = \boldsymbol{N}^\dagger.$$

Proof. We calculate

$$\begin{aligned}
\mathcal{H}_\beta &= \boldsymbol{\rho}^{-1/4}\mathcal{L}[\boldsymbol{\rho}^{1/4}\cdot\boldsymbol{\rho}^{1/4}]\boldsymbol{\rho}^{-1/4}\\
&= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\mathrm{e}^{\beta(\nu_1+\nu_2)/4}\alpha_{\nu_1,\nu_2}\boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^\dagger - \mathrm{i}\sum_{\nu\in B}(\mathrm{e}^{\frac{\beta\nu}{4}}\boldsymbol{B}_\nu(\cdot) - \mathrm{e}^{-\frac{\beta\nu}{4}}(\cdot)\boldsymbol{B}_\nu)\\
&\quad -\frac{1}{2}\sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\left(\alpha_{\nu_1,\nu_2}\mathrm{e}^{\frac{\beta(\nu_1-\nu_2)}{4}}(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1}(\cdot) + \alpha_{\nu_1,\nu_2}\mathrm{e}^{-\frac{\beta(\nu_1-\nu_2)}{4}}(\cdot)(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1}\right)\\
&= \sum_{a\in A}\sum_{\nu_1,\nu_2\in B}h_{\nu_1,\nu_2}\boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^\dagger + \frac{1}{2}\boldsymbol{N}(\cdot) + \frac{1}{2}(\cdot)\boldsymbol{N}^\dagger,
\end{aligned}$$

where

$$\boldsymbol{N} = -2\mathrm{i}\sum_{\nu\in B}\mathrm{e}^{\beta\nu/4}\boldsymbol{B}_\nu - \sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\alpha_{\nu_1,\nu_2}\mathrm{e}^{\frac{\beta(\nu_1-\nu_2)}{4}}(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1}.$$

We further simplify $\boldsymbol{N}$ by expressing $\boldsymbol{B}$ as a linear combination of $(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1}$ as in (3.14).

$$\boldsymbol{N} = \sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\Big(\exp(\beta(\nu_1-\nu_2)/4)\tanh(\beta(\nu_1-\nu_2)/4) - \exp(\beta(\nu_1-\nu_2)/4)\Big)\alpha_{\nu_1,\nu_2}(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1}$$

$$\begin{aligned}
&= -\sum_{a\in A}\sum_{\nu_1,\nu_2\in B}\frac{\alpha_{\nu_1,\nu_2}}{\cosh(\beta(\nu_1-\nu_2)/4)}(\boldsymbol{A}^a_{\nu_2})^\dagger\boldsymbol{A}^a_{\nu_1} && (\text{By } \mathrm{e}^x(\tanh(x)-1) = \tfrac{-1}{\cosh(x)})\\
&= \boldsymbol{N}^\dagger, && (\text{Invariance under } \nu_1\leftrightarrow\nu_2)
\end{aligned}$$

as advertised.

■

The transition part

For the transition part, we quickly obtain the time-domain representation using a two-dimensional Fourier Transform as a corollary of Corollary 3.5.1.

Corollary 3.5.3 (Time integrals)**.** *Suppose* $h_{\nu_1,\nu_2} = 2\pi \cdot \hat{h}_+(\nu_+) \cdot \hat{h}_-(\nu_-)$*, then*

$$\begin{aligned}&\sum_{a\in A}\sum_{\nu_1,\nu_2\in B} h_{\nu_1,\nu_2} \boldsymbol{A}^a_{\nu_1} \otimes (\boldsymbol{A}^a_{\nu_2})^* \\ &= \sum_{a\in A}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_-(t_-)h_+(t_+)\cdot \boldsymbol{A}^a(t_+ - t_-)\otimes \boldsymbol{A}^a(-t_- - t_+)^T \mathrm{d}t_+\mathrm{d}t_-,\end{aligned}$$

where $h_\pm(t)$ *are Fourier Transforms of* $\hat{h}_\pm(\nu)$*.*

Now, we can evaluate the Fourier transforms explicitly.

Proposition 3.5.2 (Linear combination for h)**.** *For* α_{ν_1,ν_2} *defined in* (3.15) *and each* $\sigma_E, g(x)$*, we have that* $h_{\nu_1,\nu_2} = 2\pi\hat{h}_+(\nu_+)\cdot\hat{h}_-(\nu_-)$ *for the discriminant (Proposition 3.5.1) where*

$$\hat{h}_+(\nu) = \int_{\frac{\beta\sigma_E^2}{2}}^{\infty} \frac{g(x)\sigma_\gamma(x)}{2\sqrt{\sigma_E^2+\sigma_\gamma^2(x)}} \exp\left(-\frac{\beta\nu^2}{16x} - \frac{\beta x}{4}\right)\mathrm{d}x$$

$$\textit{and}\quad \hat{h}_-(\nu) := \frac{1}{2\pi}\exp\left(-\frac{\nu^2}{8\sigma_E^2}\right),$$

with the corresponding time-domain functions

$$\begin{aligned}h_+(t) &= \int_{\frac{\beta\sigma_E^2}{2}}^{\infty} g(x)\sigma_\gamma(x)\exp\left(-\frac{4t^2x}{\beta} - \frac{\beta x}{4}\right)\mathrm{d}x \\ \textit{and}\quad h_-(t) &= \frac{\sigma_E}{\pi}\exp(-2\sigma_E^2 t^2).\end{aligned} \tag{3.31}$$

Proof. For $\hat{h}_+(\nu)$ term, the cross term in the exponential $\mathrm{e}^{-\beta\nu/4}$ is precisely cancelled by $h_{\nu_1,\nu_2} := \mathrm{e}^{\beta(\nu_1+\nu_2)/4}\alpha_{\nu_1,\nu_2}$; the $\hat{h}_-(\nu)$ term remains the same. To obtain the time-domain functions, we simply carry out the Gaussian integral

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{\beta\nu^2}{16x} - \frac{\beta x}{4}\right)\mathrm{e}^{\mathrm{i}\nu t}\mathrm{d}\nu = 2\sqrt{\frac{2x}{\beta}}\exp\left(-\frac{4t^2x}{\beta} - \frac{\beta x}{4}\right)$$

and use that $\sigma_E^2 + \sigma_\gamma^2(x) = \beta/2x$. ■

Proposition 3.5.3 (The Gaussian case)**.** *In the setting of Proposition 3.5.2 and for* α_{ν_1,ν_2} *defined in* (3.13)*, we have*

$$\hat{h}_+(\nu) := \frac{\sigma_\gamma}{2\sqrt{\sigma_E^2+\sigma_\gamma^2}} \cdot \exp\left(-\frac{\nu^2+(2\omega_\gamma)^2}{8(\sigma_E^2+\sigma_\gamma^2)}\right) \quad \textit{and} \quad \hat{h}_-(\nu) := \frac{1}{2\pi}\exp\left(-\frac{\nu^2}{8\sigma_E^2}\right).$$

Corollary 3.3.3 (The transition part for Gaussian weights)**.** *For the Gaussian weight* $\gamma(\omega) = \exp\left(-\frac{(\beta\omega+1)^2}{2}\right)$ *with* $\omega_\gamma = \sigma_E = \sigma_\gamma = 1/\beta$*, the discriminant* $\boldsymbol{\mathcal{H}}_\beta$ *is described in the time domain by*

$$h_+(t) = \frac{1}{\beta}\mathrm{e}^{-1/4}\exp(-\frac{4t^2}{\beta^2}) \quad \textit{and} \quad h_-(t) = \frac{1}{\pi\beta}\exp(-\frac{2t^2}{\beta^2}) \quad \textit{such that} \quad \|h_-\|_1, \|h_+\|_1 \le 1. \tag{3.24}$$

Corollary 3.3.4 (The transition part for Metropolis weights)**.** *For* $\sigma_E = \frac{1}{\beta}$*, the Metropolis-like weight* $\gamma^M(\omega) = \exp\left(-\beta\max\left(\omega+\frac{1}{2\beta},0\right)\right)$ *yields* $\boldsymbol{\mathcal{H}}_\beta$ *described in the time domain by the same* $h_-(t)$ *as in Eq.* (3.24) *and by*

$$h_+(t) = \frac{\mathrm{e}^{-1/8}}{\beta}\frac{\mathrm{e}^{-2t^2/\beta^2}}{4\sqrt{2\pi}(\frac{t^2}{\beta^2}+\frac{1}{16})} \quad \textit{such that} \quad \|h_+(t)\|_1 \le 1.$$

Proof. Setting $g(x) = \frac{1}{\sqrt{2\pi}\sigma_\gamma(x)}$ on the interval $\left(\frac{\beta\sigma_E^2}{2},\infty\right)$, yields in (3.31)

$$h_+(t) = \int_{\frac{\beta\sigma_E^2}{2}}^\infty g(x)\sigma_\gamma(x)\exp\left(-\frac{4t^2x}{\beta}-\frac{\beta x}{4}\right)\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{\frac{\beta\sigma_E^2}{2}}^\infty \exp\left(-\frac{4t^2x}{\beta}-\frac{\beta x}{4}\right)\mathrm{d}x$$
$$= \frac{1}{\beta}\frac{\mathrm{e}^{-2\sigma_E^2t^2-\beta^2\sigma_E^2/8}}{4\sqrt{2\pi}(\frac{t^2}{\beta^2}+\frac{1}{16})}.$$

Set $\sigma_E = 1/\beta$ to conclude the calculation. To obtain the ℓ_1-norm bound, integrate $\frac{\mathrm{e}^{-1/8}}{4\sqrt{2\pi}}\int_{-\infty}^\infty \frac{\mathrm{e}^{-2x^2}}{x^2+1/16}\mathrm{d}x = \sqrt{\frac{\pi}{2}}\,\mathrm{erfc}(1/\sqrt{8}) < 0.78.$ ∎

The N-term

We instantiate the time-domain presentation of the $\boldsymbol{N}$ term (adapted from the calculation for the coherent term $\boldsymbol{B}$ (Corollary 3.5.2)).

Corollary 3.5.4 (Explicit time-domain functions)**.** *In the time domain, the* $\boldsymbol{N}$ *term corresponding to coefficients constructed in* (3.15) *can be written as*

$$\boldsymbol{N} = \sum_{a\in A}\int_{-\infty}^\infty f_-(t_-)\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t_-}\left(\int_{-\infty}^\infty f_+(t_+)\boldsymbol{A}^{a^\dagger}(t_+)\boldsymbol{A}^a(-t_+)\mathrm{d}t_+\right)\mathrm{e}^{\mathrm{i}\boldsymbol{H}t_-}\mathrm{d}t_-$$

for $f_+(t)$ as in (3.27) and

$$f_-(t) = \frac{2\sigma_E}{\pi\beta}\left(\frac{1}{\cosh\left(\frac{2\pi t}{\beta}\right)} *_t \mathrm{e}^{-2\sigma_E^2 t^2}\right),$$

depending on parameters $\sigma_\gamma, \sigma_E, g(x), \beta$.

Proof. Follow the proof of Corollary 3.5.2, but drop the $\sinh(-\beta\nu/4)/(2\mathrm{i})$ term in (3.30). The convolution then gives

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{\nu^2}{8\sigma_E^2}\right)\mathrm{e}^{\mathrm{i}\nu t}\mathrm{d}\nu = 2\sigma_E\mathrm{e}^{-2\sigma_E^2 t^2}.$$

■

We instantiate similar results with minor modifications without replicating the proofs.

Corollary 3.3.5 (***N*** term for Gaussian weights)**.** *For each β, the Gaussian weight $\gamma(\omega) = \exp\left(-\frac{(\beta\omega+1)^2}{2}\right)$ with $\sigma_E = \sigma_\gamma = \omega_\gamma = \frac{1}{\beta}$ corresponds to the discriminant where*

$$\boldsymbol{N} = \sum_{a\in A}\int_{-\infty}^{\infty} n_1(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\left(\int_{-\infty}^{\infty} n_2(t')\boldsymbol{A}^{a\dagger}(\beta t')\boldsymbol{A}^{a}(-\beta t')\mathrm{d}t'\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t,$$

where

$$n_1(t) := \frac{1}{4}\cdot 2\sqrt{\pi}\left(\frac{1}{\cosh(2\pi t)} *_t \exp\left(-2t^2\right)\right) \qquad \text{such that} \quad \|n_1\|_1 = \frac{\pi}{4\sqrt{2}} < 1 \tag{3.25}$$

$$n_2(t) := 4\cdot 2\cdot\frac{1}{2\pi}\sqrt{\frac{1}{\pi}}\exp\left(-4t^2 - 2\mathrm{i}t\right) = 8\cdot b_2(t) \qquad \text{such that} \quad \|n_2\|_1 < \frac{1}{2}$$

with b_2 as in (3.20).

Proof. As both functions are positive $\frac{1}{\cosh(2\pi t)}, \exp(-2t^2) > 0$, we have that $\|n_1(t)\|_1 = \int_{-\infty}^{\infty} n_1(t)\mathrm{d}t = \frac{\sqrt{\pi}}{2}\left(\int_{-\infty}^{\infty}\frac{1}{\cosh(2\pi t)}\mathrm{d}t\right)\cdot\left(\int_{-\infty}^{\infty}\exp(-2t^2)\mathrm{d}t\right)$, where the second equality follows from the fact that the integral of a convolution is the product of the integrals of the convolved functions due to Fubini's theorem. The first integral evaluates to $1/2$ and the second evaluates to $\sqrt{\pi/2}$ due to Fact 3.2.1, therefore

$\|n_1(t)\|_1 = \pi/\sqrt{32}$. The second bound follows from the analogous bound on b_2 in Corollary 3.3.1. The factor of 4 and $\frac{1}{4}$ is redistributed to ensure both n_1 and n_2 are normalized. Note the overall extra factor of 2 in Corollary 3.5.4 compared with Corollary 3.5.2. ■

Corollary 3.3.6 ($\boldsymbol{N}$ term for Metropolis weights)**.** *If $\sigma_E = \frac{1}{\beta}$, then the Metropolis-like weight $\gamma^M(\omega) = \exp\Big(-\beta \max\Big(\omega + \frac{1}{2\beta}, 0\Big)\Big)$ corresponds to the discriminant where $\boldsymbol{N}^M$ satisfies*

$$\big\|\boldsymbol{N}^M - \boldsymbol{N}^{M,\eta}\big\| \leq \sum_{a\in A} \big\|\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\big\| \min\left(\frac{\eta\beta\|\boldsymbol{H}\|}{\sqrt{2}\pi}, O\Big((\eta\beta\|\boldsymbol{H}\|)^3\Big)\right),$$

where

$$\boldsymbol{N}^{M,\eta} := \sum_{a\in A} \int_{-\infty}^{\infty} n_1(t) \mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t} \left(\int_{-\infty}^{\infty} n_2^{M,\eta}(t') \boldsymbol{A}^{a\dagger}(\beta t') \boldsymbol{A}^a(-\beta t') \mathrm{d}t' + \frac{1}{16\sqrt{2}\pi} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a \right) \mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t} \mathrm{d}t,$$

with $n_1(t)$ as in (3.25), and $n_2^{M,\eta} = b_2^{M,\eta}$ as in (3.23).

3.6 Appendix:Calculating the coherent term for the Metropolis-like weights

In this section, we dedicate to calculating the weights corresponding to the Metropolis-like weight in order to arrive at an expression that enables efficient implementation via LCU.

Proposition 3.6.1 (Metropolis-like weights)**.** *Setting $g(x) = \frac{1}{\sqrt{2\pi}\sigma_\gamma(x)}$ on the interval $\Big(\frac{\beta\sigma_E^2}{2}, \frac{s^2}{\beta}\Big)$, yields in (3.27)*

$$f_+^{(s)}(t) = \frac{1}{\beta} \frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t} - \mathrm{e}^{-4\frac{t^2}{\beta^2}s^2 - 2\mathrm{i}\frac{t}{\beta}s^2}}{2\sqrt{2\pi}\frac{t}{\beta}(2\frac{t}{\beta} + \mathrm{i})}. \tag{3.32}$$

In particular, in the $s \to \infty$ limit, the second term vanishes, but we can only interpret the result as a distribution

$$f_+^{(\infty)}(t) = \lim_{\eta\to 0+} \mathbb{1}(|t| \geq \eta) \frac{1}{\beta} \frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t}}{2\sqrt{2\pi}\frac{t}{\beta}(2\frac{t}{\beta} + \mathrm{i})} + \sqrt{\frac{\pi}{8}}\delta(t),$$

which again decays rapidly; we would need to pay attention to the $t \sim 0$ regime due to the delta function $\delta(t)$ and $1/t$ dependence.

Proof. We substitute $g(x) = \frac{1}{\sqrt{2\pi}\sigma_\gamma(x)}$ in (3.27), which due to $\sigma_E^2 + \sigma_\gamma^2 = \frac{2\omega_\gamma}{\beta}$ gives

$$f_+^{(s)}(t) := \int_{\frac{\beta\sigma_E^2}{2}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\gamma(x)}\sigma_\gamma(x)\exp\left(-\frac{4t^2x}{\beta} - 2itx\right)\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{\frac{\beta\sigma_E^2}{2}}^{\frac{s^2}{\beta}} \exp\left(-\frac{4t^2x}{\beta} - 2itx\right)\mathrm{d}x$$

$$= \frac{1}{\beta}\frac{\mathrm{e}^{-2\sigma_E^2t^2 - \mathrm{i}\beta\sigma_E^2t} - \mathrm{e}^{-4\frac{t^2}{\beta^2}s^2 - 2\mathrm{i}\frac{t}{\beta}s^2}}{2\sqrt{2\pi}\frac{t}{\beta}(2\frac{t}{\beta} + \mathrm{i})}.$$

In order to properly understand the $s \to \infty$ limit we also need to analyze (3.29):

$$\hat{f}_+^{(s)}(\nu) = \int_{\frac{\beta\sigma_E^2}{2}}^{\infty} \frac{g(x)\sigma_\gamma}{2\sqrt{\sigma_E^2 + \sigma_\gamma^2}}\exp\left(-\frac{(\nu + 2x)^2}{16x/\beta}\right)\mathrm{d}x = \int_{\frac{\beta\sigma_E^2}{2}}^{\frac{s^2}{\beta}} \sqrt{\frac{\beta}{16\pi x}}\exp\left(-\frac{(\nu + 2x)^2}{16x/\beta}\right)\mathrm{d}x \tag{3.33}$$

$$= \frac{1}{4}\left(\mathrm{e}^{-\frac{\beta\nu}{2}}\operatorname{erf}\left(\frac{\nu}{\sqrt{8}\sigma_E} - \frac{\beta\sigma_E}{\sqrt{8}}\right) - \operatorname{erf}\left(\frac{\nu}{\sqrt{8}\sigma_E} + \frac{\beta\sigma_E}{\sqrt{8}}\right) + \mathrm{e}^{-\frac{\beta\nu}{2}}\operatorname{erf}\left(\frac{s}{2} - \frac{\beta\nu}{4s}\right) + \operatorname{erf}\left(\frac{s}{2} + \frac{\beta\nu}{4s}\right)\right),$$

which in the $s \to \infty$ limit becomes

$$\hat{f}_+^{(\infty)}(\nu) = \frac{1}{4}\left(\operatorname{erfc}\left(\frac{\beta\sigma_E}{\sqrt{8}} + \frac{\nu}{\sqrt{8}\sigma_E}\right) + \mathrm{e}^{-\frac{\beta\nu}{2}}\operatorname{erfc}\left(\frac{\beta\sigma_E}{\sqrt{8}} - \frac{\nu}{\sqrt{8}\sigma_E}\right)\right). \tag{3.34}$$

In order to compute $f_+^{(\infty)}(\nu)$, let us consider the function $\hat{f}_{+,0}^{(\infty)}(\nu) := \hat{f}_+^{(\infty)}(\nu) - \frac{1-\operatorname{sgn}(\nu)}{4}$, where

$$\operatorname{sgn}(\nu) = \begin{cases} 0 & \text{if} \quad \nu = 0 \\ \operatorname{sgn}(\nu) = \nu/|\nu| & \text{if} \quad \nu \neq 0 \end{cases}.$$

Since $\hat{f}_{+,0}^{(\infty)}(\nu) \in \ell_1$, its inverse Fourier Transform exists, and it can be computed as $f_{+,0}^{(\infty)}(t) = \frac{1}{\beta}\frac{\mathrm{e}^{-2\sigma_E^2t^2 - \mathrm{i}\beta\sigma_E^2t + 2\mathrm{i}\frac{t}{\beta} - 1}}{2\sqrt{2\pi}\frac{t}{\beta}(2\frac{t}{\beta} + \mathrm{i})}$ after a change of variables by utilizing the (complex) Laplace transform of the error function erf [146]. Moreover, since $\hat{f}_{+,0}^{(\infty)}(\nu)$ is smooth, apart from $\nu = 0$ where its value is the mean of its left and right side limits, standard results in the theory of Fourier integrals (see, e.g., [189, Chapter 18.3.1.d]) imply

$$\hat{f}_{+,0}^{(\infty)}(\nu) = \lim_{\eta \to 0+}\frac{1}{\sqrt{2\pi}}\int_{-\frac{1}{\eta}}^{\frac{1}{\eta}} f_{+,0}^{(\infty)}(\nu)e^{-\mathrm{i}\nu t}\mathrm{d}t = \lim_{\eta \to 0+}\frac{1}{\sqrt{2\pi}}\int_{-\frac{1}{\eta}}^{\frac{1}{\eta}}\frac{1}{\beta}\frac{\mathrm{e}^{-2\sigma_E^2t^2 - \mathrm{i}\beta\sigma_E^2t} + 2\mathrm{i}\frac{t}{\beta} - 1}{2\sqrt{2\pi}\frac{t}{\beta}(2\frac{t}{\beta} + \mathrm{i})}\mathrm{d}t.$$

On the other hand, it is a standard calculation that

$$\frac{1}{4}\operatorname{sgn}(\nu) = \lim_{\eta\to 0+}\frac{1}{8\pi}\int_{\eta}^{\infty}\frac{\sin(\nu t)}{t}\mathrm{d}t = \lim_{\eta\to 0+}\frac{1}{\sqrt{2\pi}}\left(\int_{-\frac{1}{\eta}}^{-\eta}\frac{\mathrm{i}}{\sqrt{8\pi t}}\mathrm{e}^{-\mathrm{i}\nu t}\mathrm{d}t + \int_{\eta}^{\frac{1}{\eta}}\frac{\mathrm{i}}{\sqrt{8\pi t}}\mathrm{e}^{-\mathrm{i}\nu t}\mathrm{d}t\right).$$

Since $f_{+,0}^{(\infty)}(\nu)$ is bounded in the neighborhood of 0 we can combine the above two expression, yielding

$$\begin{aligned}
\hat{f}_{+}^{(\infty)}(\nu) - \frac{\operatorname{sgn}(\nu)}{4} &= \hat{f}_{+,0}^{(\infty)}(\nu) - \frac{1}{4} \\
&= \lim_{\eta\to 0+}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathbb{1}(|t|\in(\eta,1/\eta))\left(f_{+,0}^{(\infty)}(\nu) - \frac{\mathrm{i}}{\sqrt{8\pi t}}\right)e^{-\mathrm{i}\nu t}\mathrm{d}t \\
&= \lim_{\eta\to 0+}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathbb{1}(|t|\in(\eta,1/\eta))\frac{1}{\beta}\frac{\mathrm{e}^{-2\sigma_E^2t^2-\mathrm{i}\beta\sigma_E^2t}}{2\sqrt{2\pi}\frac{t}{\beta}(2\frac{t}{\beta}+\mathrm{i})}\mathrm{e}^{-\mathrm{i}\nu t}\mathrm{d}t \\
&= \lim_{\eta\to 0+}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathbb{1}(|t|\geq\eta)\frac{1}{\beta}\frac{\mathrm{e}^{-2\sigma_E^2t^2-\mathrm{i}\beta\sigma_E^2t}}{2\sqrt{2\pi}\frac{t}{\beta}(2\frac{t}{\beta}+\mathrm{i})}\mathrm{e}^{-\mathrm{i}\nu t}\mathrm{d}t,
\end{aligned}$$

where the last equality holds because of the rapid decay of the function for large t values. We can conclude the proof by observing that the Fourier Transform of $\sqrt{\frac{\pi}{8}}\delta(t)$ is $\frac{1}{4}$. ∎

In the $s=\infty$ case, the diverging ℓ_1 norm of $f_{+}^{(\infty)}(t)$ raises the question: how big could the norm $\left\|\boldsymbol{B}^M\right\|$ be? In the following we spell out the explicit form of $\boldsymbol{B}$ by removing the singularity at $t=0$, which shows that in fact $\left\|\boldsymbol{B}^M\right\| \leq O(\log(\beta\|\boldsymbol{H}\|))$, because $\|f_-\|_1 = O(1)$. Although the exact formula becomes cumbersome, it enables us to find a very precise approximation by a much simpler formula in Corollary 3.3.2.

Proposition 3.6.2 (Exact form of Metropolis coherent term)**.** *The coherent term $\boldsymbol{B}^M$ corresponding to the quasi-Metropolis weight $\gamma_{\sigma_E}^{(\infty)}(\omega)$ in (3.16) can be written as*

$$\boldsymbol{B}^M = \sum_{a\in A}\int_{-\infty}^{\infty} f_-(t_-)\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t_-}\boldsymbol{O}_a^M\mathrm{e}^{\mathrm{i}\boldsymbol{H}t_-}\mathrm{d}t_-,$$

where $f_-(t_-)$ is defined in (3.28), and for arbitrary $\theta > 0$ we have

$$\boldsymbol{O}_a^M = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^{a\dagger} \left(\frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t} + \mathbb{1}(|t| \le \theta)\mathrm{i}\left(2\frac{t}{\beta} + \mathrm{i}\right)}{\sqrt{8\pi}t(2\frac{t}{\beta} + \mathrm{i})} \mathrm{e}^{-2\mathrm{i}\boldsymbol{H}t} - \mathbb{1}(|t| \le \theta)\boldsymbol{H}\frac{\operatorname{sinc}(2\boldsymbol{H}t)}{\sqrt{2\pi}} \right) \boldsymbol{A}^a \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \mathrm{d}t$$

$$+ \frac{1}{\sqrt{8\pi}} \int_{-\theta}^{\theta} \cos(\boldsymbol{H}t)\boldsymbol{A}^{a\dagger} \cos(2\boldsymbol{H}t)\boldsymbol{A}^a \boldsymbol{H} \operatorname{sinc}(\boldsymbol{H}t) + \operatorname{sinc}(\boldsymbol{H}t)\boldsymbol{H}\boldsymbol{A}^{a\dagger} \cos(2\boldsymbol{H}t)\boldsymbol{A}^a \cos(\boldsymbol{H}t)\mathrm{d}t$$

$$+ \sqrt{\frac{\pi}{8}} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a.$$

Proof. By Corollary 3.5.1-Proposition 3.6.1 we know that

$$\boldsymbol{O}_a^M = \lim_{\eta \to 0+} \int_{-\infty}^{\infty} \left(\mathbb{1}(|t| \ge \eta) \frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t}}{\sqrt{8\pi}t(2\frac{t}{\beta} + \mathrm{i})} + \sqrt{\frac{\pi}{8}}\delta(t) \right) \boldsymbol{A}^{a\dagger}(t)\boldsymbol{A}^a(-t)dt.$$

We decompose the above integral in order to remove its singularity at $t = 0$. We start with the decomposition

$$\frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t}}{\sqrt{8\pi}t(2\frac{t}{\beta} + \mathrm{i})} = \frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t} + \mathbb{1}(|t| \le \theta)\mathrm{i}\left(2\frac{t}{\beta} + \mathrm{i}\right)}{\sqrt{8\pi}t(2\frac{t}{\beta} + \mathrm{i})} - \mathbb{1}(|t| \le \theta)\frac{\mathrm{i}}{\sqrt{8\pi}t},$$

implying that

$$\begin{aligned}
&\frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t}}{\sqrt{8\pi}t(2\frac{t}{\beta} + \mathrm{i})} \boldsymbol{A}^{a\dagger}(t)\boldsymbol{A}^a(-t) \\
&= \mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^{a\dagger} \cdot \left(\frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t}}{\sqrt{8\pi}t(2\frac{t}{\beta} + \mathrm{i})} \mathrm{e}^{-2\mathrm{i}\boldsymbol{H}t} \right) \boldsymbol{A}^a \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \\
&= \mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^{a\dagger} \cdot \left(\frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t} + \mathbb{1}(|t| \le \theta)\mathrm{i}\left(2\frac{t}{\beta} + \mathrm{i}\right)}{\sqrt{8\pi}t(2\frac{t}{\beta} + \mathrm{i})} - \mathbb{1}(|t| \le \theta)\frac{\mathrm{i}}{\sqrt{8\pi}t} \right) \mathrm{e}^{-2\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^a \mathrm{e}^{\mathrm{i}\boldsymbol{H}t},
\end{aligned}$$

where

$$\frac{\mathrm{i}}{\sqrt{8\pi}t}\mathrm{e}^{-2\mathrm{i}\boldsymbol{H}t} = \frac{\mathrm{i}}{\sqrt{8\pi}t}\cos(2\boldsymbol{H}t) + \frac{1}{\sqrt{8\pi}t}\underbrace{\sin(2\boldsymbol{H}t)}_{2\boldsymbol{H}t\operatorname{sinc}(2\boldsymbol{H}t)} = \mathrm{i}\frac{\cos(2\boldsymbol{H}t)}{\sqrt{8\pi}t} + \boldsymbol{H}\frac{\operatorname{sinc}(2\boldsymbol{H}t)}{\sqrt{2\pi}}.$$

Finally, observe that due to parity reasons, we have

$$\int_{-\infty}^{\infty} \mathbb{1}(|t| \geq \eta) \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^{a\dagger} \mathbb{1}(|t| \leq \theta) \frac{\mathrm{i}}{\sqrt{8\pi}} \frac{\cos(2\boldsymbol{H}t)}{t} \boldsymbol{A}^{a} \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \mathrm{d}t$$
$$= \frac{-1}{\sqrt{8\pi}} \int_{-\infty}^{\infty} \mathbb{1}(\eta \leq |t| \leq \theta) \left(\cos(\boldsymbol{H}t) \boldsymbol{A}^{a\dagger} \frac{\cos(2\boldsymbol{H}t)}{t} \boldsymbol{A}^{a} \sin(\boldsymbol{H}t) + \sin(\boldsymbol{H}t) \boldsymbol{A}^{a\dagger} \frac{\cos(2\boldsymbol{H}t)}{t} \boldsymbol{A}^{a} \cos(\boldsymbol{H}t) \right) \mathrm{d}t$$

$$= \frac{-1}{\sqrt{8\pi}} \int_{-\infty}^{\infty} \mathbb{1}(\eta \leq |t| \leq \theta)$$
$$\left(\cos(\boldsymbol{H}t) \boldsymbol{A}^{a\dagger} \cos(2\boldsymbol{H}t) \boldsymbol{A}^{a} \boldsymbol{H} \operatorname{sinc}(\boldsymbol{H}t) + \operatorname{sinc}(\boldsymbol{H}t) \boldsymbol{H} \boldsymbol{A}^{a\dagger} \cos(2\boldsymbol{H}t) \boldsymbol{A}^{a} \cos(\boldsymbol{H}t) \right) \mathrm{d}t.$$

Since in the above provided decomposition, every (matrix) function is bounded in the neighborhood of 0, we can obtain the $\eta \to 0+$ limit by simply removing the indicator $\mathbb{1}(|t| \geq \eta)$.

$$\lim_{\eta \to 0+} \int_{-\infty}^{\infty} \mathbb{1}(|t| \geq \eta) \frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t}}{\sqrt{8\pi} t (2\frac{t}{\beta} + \mathrm{i})} \boldsymbol{A}^{a\dagger}(t) \boldsymbol{A}^{a}(-t) dt$$
$$= \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A}^{a\dagger} \cdot \left(\frac{\mathrm{e}^{-2\sigma_E^2 t^2 - \mathrm{i}\beta\sigma_E^2 t} + \mathbb{1}(|t| \leq \theta) \mathrm{i} \left(2\frac{t}{\beta} + \mathrm{i} \right)}{\sqrt{8\pi} t (2\frac{t}{\beta} + \mathrm{i})} \mathrm{e}^{-2\mathrm{i}\boldsymbol{H}t} - \mathbb{1}(|t| \leq \theta) \boldsymbol{H} \frac{\operatorname{sinc}(2\boldsymbol{H}t)}{\sqrt{2\pi}} \right) \boldsymbol{A}^{a} \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \mathrm{d}t$$

$$+ \frac{1}{\sqrt{8\pi}} \int_{-\theta}^{\theta} \cos(\boldsymbol{H}t) \boldsymbol{A}^{a\dagger} \cos(2\boldsymbol{H}t) \boldsymbol{A}^{a} \boldsymbol{H} \operatorname{sinc}(\boldsymbol{H}t) + \operatorname{sinc}(\boldsymbol{H}t) \boldsymbol{H} \boldsymbol{A}^{a\dagger} \cos(2\boldsymbol{H}t) \boldsymbol{A}^{a} \cos(\boldsymbol{H}t) \mathrm{d}t.$$

We conclude by noting that $\int_{-\infty}^{\infty} \sqrt{\frac{\pi}{8}} \delta(t) \boldsymbol{A}^{a\dagger}(t) \boldsymbol{A}^{a}(-t) dt = \sqrt{\frac{\pi}{8}} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^{a}$. ∎

Note that this exact formula could be directly and efficiently implemented using QSVT. However, the gains are minimal, as it would only reduce subnormalization from $O(\log(\beta\|H\|/\epsilon))$ (this being the ℓ_1 norm required to achieve ϵ precision in Corollary 3.3.2) to $O(\log(\beta\|H\|))$ (the ℓ_1 norm of the weight function corresponding to the natural choice $\theta = 1/\beta\|\boldsymbol{H}\|$ in Proposition 3.6.2).

More interestingly, this exact formula seems also to hold in the infinite-dimensional case, giving rise to exact detailed balance in the infinite-dimensional version as well. We leave it for future work to verify that the construction and its analysis can indeed be generalized to infinite dimensional systems.

Moreover, the above result actually also implies that the Lindbladian corresponding to (3.32) also well-approximates its $s \to \infty$ limit, i.e., the above Metropolis-like Lindbladian. The argument goes as follows: first, "round" the Hamiltonian $\boldsymbol{H}$ to discretize its spectrum at some finite resolution $\ll 1/\beta$. Due to the form of (3.32), the resulting perturbation of the Lindbladian is bounded. Then, take the $s \to \infty$ limit. Since (3.33) is exponentially close to its limit (3.34) and there is a limited number of Bohr frequencies due to rounding, the resulting perturbation of the Lindbladian is once again bounded. Finally, undo the rounding, which again causes a bounded perturbation due to the exact form of Metropolis coherent term (Proposition 3.6.2). Carefully executing these bounds in the $\sigma_E = \beta$ case should show that for $s = \Theta((\beta\|H\| + \log(1/\epsilon) + 1)^2)$ the resulting Lindbladian is ϵ-close to its $s = \infty$ limit.

Now we show how to approximate the exact Metropolis Lindbladian in a different efficient way.

Corollary 3.3.2 (Approximate coherent term for the Metropolis-like weight)**.** *If $\sigma_E = \frac{1}{\beta}$, then the coherent term $\boldsymbol{B}^M$ corresponding to the Metropolis-like weight $\gamma^M(\omega) = \exp\left(-\beta \max\left(\omega + \frac{1}{2\beta}, 0\right)\right)$ satisfies*

$$\left\|\boldsymbol{B}^M - \boldsymbol{B}^{M,\eta}\right\| \leq \left\|\sum_{a\in A} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right\| \min\left(\frac{\eta\beta\|\boldsymbol{H}\|}{\sqrt{2}\pi}, O\left((\eta\beta\|\boldsymbol{H}\|)^3\right)\right), \tag{3.21}$$

where

$$\boldsymbol{B}^{M,\eta} := \sum_{a\in A} \int_{-\infty}^{\infty} b_1(t) \mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t} \left(\int_{-\infty}^{\infty} b_2^{M,\eta}(t') \boldsymbol{A}^{a\dagger}(\beta t')\boldsymbol{A}^a(-\beta t')\mathrm{d}t' + \frac{1}{16\sqrt{2}\pi}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right)\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t, \tag{3.22}$$

with $b_1(t)$ as in (3.19), and

$$b_2^{M,\eta}(t) := \frac{1}{4\sqrt{2}\pi} \frac{\exp(-2t^2 - \mathrm{i}t) + \mathbb{1}(|t| \leq \eta)\mathrm{i}(2t + \mathrm{i})}{t(2t + \mathrm{i})}$$
$$\text{such that} \quad \|b_2^{M,\eta}\|_1 < \frac{1}{5} + \frac{1}{2\sqrt{2}\pi}\ln(1/\eta). \tag{3.23}$$

Further, if $\left[\sum_{a\in A} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a, \boldsymbol{H}\right] = 0$, we can drop the second term in (3.22) after the integral in t' since $\int_{-\infty}^{\infty} b_1(t) = 0$.

Proof. First, let us establish the norm bound (3.23). Observe that

$$b_2^{M,\eta}(t) = b_2^{M,1}(t) - \frac{1}{4\sqrt{2}\pi}\mathbb{1}(\eta < |t| \leq 1)\frac{\mathrm{i}}{t}.$$

Using the above identity and applying a triangle inequality and then Hölder's inequality gives

$$\|b_2^{M,\eta}\|_1 \leq \left\|(1+4t^2)^{-1}\right\|_2 \left\|(1+4t^2)b_2^{M,1}\right\|_2 + \frac{1}{4\sqrt{2}\pi}\left\|\mathbb{1}(\eta < |t| \leq 1)\frac{\mathrm{i}}{t}\right\|_1.$$

To obtain (3.23), use that $\int_\eta^1 \frac{1}{t}\mathrm{d}t = \ln(1/\eta)$, $\int_{-\infty}^{\infty} \frac{1}{(1+4t^2)^2}\mathrm{d}t = \frac{\pi}{4}$, $\int_1^\infty \left|\frac{\exp(-2t^2)(2t-\mathrm{i})}{t}\right|^2 \mathrm{d}t = \mathrm{e}^{-4} - \sqrt{\pi}\,\mathrm{erfc}(2)$, and a direct computation of $\int_{-1}^1 \left|\frac{\exp\left(-2t^2-\mathrm{i}t\right)(2t-\mathrm{i})+\mathrm{i}\left(1+4t^2\right)}{t}\right|^2 < 16$.

Next, setting $\theta = \eta\beta$ in Proposition 3.6.2 and scaling the variables t, t' by a factor of β reveals that

$$\boldsymbol{B}^M - \boldsymbol{B}^{M,\eta} = \frac{1}{2\sqrt{2}\pi^2}\sum_{a\in A}\int_{-\infty}^{\infty} b_1(t)\mathrm{e}^{-\mathrm{i}\beta\boldsymbol{H}t}\boldsymbol{Q}_a\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t}\mathrm{d}t, \tag{3.35}$$

where

$$\begin{aligned}\boldsymbol{Q}_a = \frac{1}{2}\int_{-\eta}^{\eta} &\cos(\beta\boldsymbol{H}t')\boldsymbol{A}^{a\dagger}\cos(2\beta\boldsymbol{H}t')\boldsymbol{A}^a\beta\boldsymbol{H}\,\mathrm{sinc}(\beta\boldsymbol{H}t')\\ &+ \mathrm{sinc}(\beta\boldsymbol{H}t')\beta\boldsymbol{H}\boldsymbol{A}^{a\dagger}\cos(2\beta\boldsymbol{H}t')\boldsymbol{A}^a\cos(\beta\boldsymbol{H}t')\mathrm{d}t'\\ &- \int_{-\eta}^{\eta}\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t'}\boldsymbol{A}^{a\dagger}\big(\beta\boldsymbol{H}\,\mathrm{sinc}(2\beta\boldsymbol{H}t')\big)\boldsymbol{A}^a\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t'}\mathrm{d}t'.\end{aligned}$$

We can decompose in the second term $\mathrm{e}^{\mathrm{i}\beta\boldsymbol{H}t'} = \cos(\beta\boldsymbol{H}t') + \mathrm{i}\sin(\beta\boldsymbol{H}t')$. Due to parity reasons, we can see that the second term can be replaced by

$$-\int_{-\eta}^{\eta}\cos(\beta\boldsymbol{H}t')\boldsymbol{A}^{a\dagger}\big(\beta\boldsymbol{H}\,\mathrm{sinc}(2\beta\boldsymbol{H}t')\big)\boldsymbol{A}^a\cos(\beta\boldsymbol{H}t') - \sin(\beta\boldsymbol{H}t')\boldsymbol{A}^{a\dagger}\big(\beta\boldsymbol{H}\,\mathrm{sinc}(2\beta\boldsymbol{H}t')\big)\boldsymbol{A}^a\sin(\beta\boldsymbol{H}t')\mathrm{d}t'.$$

Now, let us define

$$\begin{aligned}\mu_a(\boldsymbol{X}) :=& \cos(\boldsymbol{X})\boldsymbol{A}^{a\dagger}\cos(2\boldsymbol{X})\boldsymbol{A}^a\boldsymbol{X}\,\mathrm{sinc}(\boldsymbol{X}) + \mathrm{sinc}(\boldsymbol{X})\boldsymbol{X}\boldsymbol{A}^{a\dagger}\cos(2\boldsymbol{X})\boldsymbol{A}^a\cos(\boldsymbol{X})\\ &- 2\cos(\boldsymbol{X})\boldsymbol{A}^{a\dagger}\big(\boldsymbol{X}\,\mathrm{sinc}(2\boldsymbol{X})\big)\boldsymbol{A}^a\cos(\boldsymbol{X}) + 2\sin(\boldsymbol{X})\boldsymbol{A}^{a\dagger}\big(\boldsymbol{X}\,\mathrm{sinc}(2\boldsymbol{X})\big)\boldsymbol{A}^a\sin(\boldsymbol{X}),\end{aligned}$$

so that

$$\boldsymbol{Q}_a = \int_{-\eta}^{\eta}\frac{\mu_a(\beta\boldsymbol{H}t')}{2t'}\mathrm{d}t'.$$

If $\boldsymbol{X}$ is Hermitian, by the triangle inequality, we can see that $\|\mu_a(\boldsymbol{X})\| \leq 6\|\boldsymbol{A}^a\|^2\|\boldsymbol{X}\|$. Moreover, from its definition, we can see that $\mu_a(\boldsymbol{X})$ is an odd analytic entire

function, and its derivative is 0 at $\boldsymbol{X} = 0$. From this it directly follows that $\mu_a(\boldsymbol{X}) = O\left(\|\boldsymbol{A}^a\|^2\|\boldsymbol{X}\|^3\right)$ for every Hermitian $\boldsymbol{X}$. Due to the triangle inequality,

$$\begin{aligned}\|\boldsymbol{Q}_a\| \leq \int_{-\eta}^{\eta} \frac{\|\mu_a(\beta\boldsymbol{H}t')\|}{2|t'|}\mathrm{d}t' &\leq \int_{-\eta}^{\eta} \|\boldsymbol{A}^a\|^2\|\beta\boldsymbol{H}\| \min\left(\frac{6}{2}, O\left(\|\beta\boldsymbol{H}t'\|^2\right)\right)\mathrm{d}t' \\ &= \|\boldsymbol{A}^a\|^2 \min\left(6\eta\beta\|\boldsymbol{H}\|, O\left((\eta\beta\|\boldsymbol{H}\|)^3\right)\right).\end{aligned}$$

Finally, we can prove (3.21) for individual jumps by combining the above bound and the inequality $\|b_1\|_1 \leq 1$ within (3.35). To obtain the bound in terms of the norm-of-sum $\|\sum_a \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|$, note that for any matrices $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$, we have

$$\begin{aligned}\|\sum_a \boldsymbol{X}\boldsymbol{A}^{a\dagger}\boldsymbol{Y}\boldsymbol{A}^a\boldsymbol{Z}\| &\leq \|\boldsymbol{X}\| \cdot \|\sum_a \boldsymbol{A}^{a\dagger} \otimes \langle a|\| \cdot \|\boldsymbol{Y} \otimes \boldsymbol{I}\| \cdot \|\sum_a \boldsymbol{A}^a \otimes |a\rangle\| \cdot \|\boldsymbol{Z}\| \\ &= \|\boldsymbol{X}\|\|\boldsymbol{Y}\|\|\boldsymbol{Z}\| \cdot \|\sum_a \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|.\end{aligned}$$

And rewrite $\mu_a(\boldsymbol{X})$ as a linear combination of terms each bounded by $O(\|\sum_a \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\|\boldsymbol{X}\|^3)$ by canceling out the linear $O(\boldsymbol{X})$ terms. ∎

3.7 Appendix:The discriminant gap and area law

The fact that the discriminant $\mathcal{H}_\beta$ can be regarded as frustration-free, spatially local Hamiltonian (Proposition 3.1.1) has direct implications on the "locality" of its zero-eigenstate, the purified Gibbs state. In particular, existing analytic techniques for studying gapped phases immediately lead to *explicit* low-depth circuits assuming a large discriminant gap for $\mathcal{H}_\beta$. Elegantly, lowering the temperature can now be interpreted as the Nature-given adiabatic paths parameterized by the inverse temperature β (known as quantum simulated annealing).

The particularly helpful tool is quasi-adiabatic evolution [85], a version of adiabatic evolution that exploits spatial locality.

Lemma 3.7.1 (Quasi-adiabatic evolution [15, Proposition 2.4])**.** *Consider a one-parameter family of Hamiltonians*

$$\boldsymbol{H}(s) \quad \textit{for each} \quad s \in [0,1] \quad \textit{with minimal gap} \quad \Delta.$$

Then, the family of ground-state projectors can be generated by a time-dependent Hamiltonian

$$\begin{aligned}&\boldsymbol{P}'(s) = \mathrm{i}[\boldsymbol{W}(s), \boldsymbol{P}(s)] \quad \textit{where} \\ &\boldsymbol{W}(s) := \int_{-\infty}^{\infty} \mathrm{d}t\; w(t) \int_0^t du\; \mathrm{e}^{\mathrm{i}u\boldsymbol{H}(s)}\boldsymbol{H}'(s)\mathrm{e}^{-\mathrm{i}u\boldsymbol{H}(s)} \quad \textit{for each} \quad s \in [0,1]\end{aligned}$$

for any weight function $w(t)$ satisfying

$$\int_{-\infty}^{\infty} w(t)\mathrm{d}t = 1 \quad \textit{and} \quad \tilde{w}(\omega) = 0 \quad \textit{if} \quad |\omega| \geq |\Delta|.$$

For geometrically local Hamiltonian on d-dimention lattice with single-site jumps $\|\boldsymbol{A}^a\| = 1$ (note that the natural normalization here is to make the strength extensive $\|\sum_{a\in A} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \propto n$), we can plug the discriminant into the above by

$$\boldsymbol{H}(s) \to \boldsymbol{\mathcal{H}}_{s\beta} \quad \text{and} \quad \Delta \to \min_s \lambda_{gap}(\boldsymbol{\mathcal{H}}_{s\beta}).$$

Strictly speaking, we are interested in the top eigenvector instead of the ground state due to its Lindbladian origin (Lindbladian spectrum is always nonpositive), but this is also handled by quasi-adiabatic continuation. We obtain a unitary circuit preparing the purified Gibbs state driven by the following time-dependent Hamiltonian:

$$\boldsymbol{W}(s) = \sum_a \underbrace{\int_{-\infty}^{\infty} \mathrm{d}t\; w(t) \int_0^t du\; \mathrm{e}^{\mathrm{i}u\boldsymbol{\mathcal{H}}_{s\beta}} \boldsymbol{\mathcal{H}}_{s\beta}^{a'} \mathrm{e}^{-\mathrm{i}u\boldsymbol{\mathcal{H}}_{s\beta}}}_{=:\boldsymbol{W}^a(s)}. \tag{3.36}$$

Algorithmically, the above should be considered an alternative to modernized adiabatic algorithms based on amplitude amplification [40, Appedix G]. The Heisenberg evolution is particularly powerful in the case of local Hamiltonians as it manifestly gives a quasi-local unitary circuit. Based on existing tools, we quickly instantiate several notations of locality. First, the unitary preserves spatially local operators.

Proposition 3.7.1 (Quasi-local unitary [15])**.** *Consider the one-parameter family of unitaries generated by the time-dependent Hermitian matrix $\boldsymbol{W}(s)$ as in* (3.36)

$$\boldsymbol{V}'(s) = \mathrm{i}\boldsymbol{W}(s)\boldsymbol{V}(s) \quad \textit{for each} \quad s \in [0,1].$$

Then, the unitary satisfies a Lieb-Robinson bound with almost exponential decay

$$\|[\boldsymbol{V}(s)\boldsymbol{O}_1\boldsymbol{V}^\dagger(s), \boldsymbol{O}_2]\| \leq O\left(\exp(-\frac{d}{\mathrm{Poly}_D(\beta)\ln(d^2)})\right) \quad \textit{where} \quad d = d(\boldsymbol{O}_1, \boldsymbol{O}_2).$$

Second, the unitary generates limited entanglement. Since the unitary is generated by (quasi)-local operators, integrating the entangling rate from $0 \to \beta$ [13] gives an area law of entanglement [171] for purified Gibbs state.

[13] The infinite temperature Gibbs state is the maximally entangled state. It is a tensor product of bell pairs across the two copies and has trivially entanglement across regions.

Corollary 3.7.1 (Thermal area-law from entangling rates (similar argument to [171]))**.** *The purified Gibbs state satisfies an entanglement area-law*

$$\textit{(the entanglement entropy for any region } R) \leq \frac{\mathrm{Poly}_D(\beta)}{\Delta} \cdot |\partial R|.$$

The above thermal area-law is weaker than existing bounds[14] and should be thought of as a sanity check. However, what's new is the explicitly parameterized low-depth unitary circuit (which is not a consequence of the area law). The unitary $\boldsymbol{V}(1)$ corresponding to (3.36) for a D-dimensional lattice Hamiltonian can be decomposed into a

$$\text{depth} \quad \frac{\mathrm{Poly}_D(\beta)}{\Delta} \cdot \mathrm{Poly}\log(n/\epsilon) \quad \text{circuit of 2-qubit gates}$$

up to ϵ-error in the operator norm using the HHKL algorithm [80]. The discriminant gap Δ is expected to be independent of the system size in the rapid mixing regime (under the normalization $\|\boldsymbol{A}^a\| = 1$). To obtain gapped ground states, we could scale the inverse temperature with the Hamiltonian gap $\Delta(\boldsymbol{H})$(not to confuse with the discriminant gap Δ)

$$\beta = O\big(\Delta(\boldsymbol{H}) \log(n/\epsilon)\big).$$

Of course, there might be phase transitions at low temperatures where the discriminant gap may close. The logarithmic dependence on the system size marks a technical difference from the area law setting. Nevertheless, the explicit circuit appears novel and could potentially lead to new numerical ansatz for low-energy states.

The Lindbladian case

Of course, we may apply the same line of thought to Lindbladians, assuming the quasi-local Lindbladian mixes in logarithmic time. However, there are subtle differences. The spectral gap of the parent Hamiltonian is qualitatively the same as the mixing time of our Lindbladian (which has a vanishing anti-Hermitian part)

$$\frac{\ln(2)}{\lambda_{gap}(\mathcal{H})} \leq t_{mix}(\mathcal{L}) \leq \frac{\ln(2\|\boldsymbol{\rho}^{-1/2}\|)}{\lambda_{gap}(\mathcal{H})}.$$

The conversion overhead $\ln(2\|\boldsymbol{\rho}^{-1/2}\|) = O(\beta\|\boldsymbol{H}\|)$ could scale with the system size in general. However, to drop this $O(\beta\|\boldsymbol{H}\|)$ factor (i.e., proving *rapid mixing*

[14] Ref. [104] states that the *mutual information* between regions satisfies $I(R:R') \leq O(\beta^{2/3})|\partial R|$ in any D-dimension lattice.

such that the mixing time scaling logarithmically with the system size) amounts to proving a more stringent quantum log-Sobelov inequality, which has been highly nontrivial [31]. Even if we assume rapid mixing, to get a low-depth circuit for the Gibbs state, we still need an efficient algorithm to implement the Lindbladian evolution that parallelizes the Lindblad operators. However, the HHKL algorithm [80] achieving this parallelization for local Hamiltonians made critical use of *reversing* time evolutions, which does not obviously apply to the dissipative setting.

As a remedy, one may consider trotterizing into quasi-local brickwork nonunitary circuits, giving a *discrete-time* Quantum Markov chain[15]

$$e^{\mathcal{L}\theta} \to \mathcal{N} := \prod_g e^{\mathcal{L}_g \theta} \quad \text{where} \quad \mathcal{L} = \sum_g \mathcal{L}_g$$

such that each $\mathcal{L}_g$ is a sum of quasi-local, nearly commuting block.

That is, we regroup the Lindbladian $\mathcal{L} = \sum_{a\in A} \mathcal{L}^a$ into quasi-local blocks to exploit the fact that nearly disjoint Lindbladians can be efficiently implemented in parallel (after spatial truncation). However, the Trotter error from first-order product formulas is extensive; thus, the discrete-time channel would not approximate the continuous-time evolution in general, and the mixing time analysis might require quantitatively different formalism. Nevertheless, the Gibbs state remains stationary

$$\mathcal{N}[\rho_\beta] = \rho_\beta,$$

and could potentially share similar mixing behavior as the continuous case in practice.

3.8 Appendix:Why Gaussians?

Our direct calculation confirms the correctness of our ansatz. But why does the filter $f(t)$ need to be Gaussian (3.3)? In this section, we try to derive the Gaussians from scratch, which can be viewed as an alternative view of detailed balance in the time domain. While we try to make our arguments precise, we stop at physicists' level of rigor and do not attempt to extract a mathematical theorem.

Assume that the filter function is real on the real axis

$$f(t) = f(t)^* \quad \text{for each} \quad t \in \mathbb{R}$$

and complex analytic, independent of β, and that the jump operator $A = A^\dagger$ is Hermitian (Let us focus on a single jump and drop the jump labels $a \in A$.). The goal is to solve for the viable choices of $f(t)$.

[15] The observation that discretization can be helpful despite deviating from the continuum was made in [61].

We calculate the associated "transition" part (3.1) in the time domain

$$\begin{aligned}\mathcal{T}^\dagger[\cdot] &= \int_{-\infty}^{\infty} \gamma(\omega)\hat{\boldsymbol{A}}(\omega)^\dagger(\cdot)\hat{\boldsymbol{A}}(\omega)\mathrm{d}\omega \\ &= \int \gamma(\omega)\mathrm{e}^{\mathrm{i}\omega(t-t')}f(t)f(t')\boldsymbol{A}(t)(\cdot)\boldsymbol{A}(t')\mathrm{d}\omega\mathrm{d}t\mathrm{d}t' \quad (\text{using} \quad f(t)=f(t)^*) \\ &= \frac{1}{\sqrt{2\pi}}\int c(t-t')f(t)f(t')\boldsymbol{A}(t)(\cdot)\boldsymbol{A}(t')\mathrm{d}t\mathrm{d}t'.\end{aligned}$$

The third line uses the inverse Fourier Transform

$$c(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\omega t}\gamma(\omega)\mathrm{d}\omega.$$

We may interpret $c(t)$ as a certain correlation function (hence the notation). Now, note that conjugating the Gibbs state yields

$$\begin{aligned}\Lambda(\hat{\boldsymbol{A}}(\omega)) &= \int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\omega t}f(t)\mathrm{e}^{\mathrm{i}\boldsymbol{H}(t-\mathrm{i}\beta/2)}\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}(t-\mathrm{i}\beta/2)}\mathrm{d}t \\ &= \int \mathrm{e}^{-\mathrm{i}s\omega}\mathrm{e}^{\beta\omega/2}\boldsymbol{A}(s)f(s+\mathrm{i}\beta/2)\mathrm{d}s \qquad (\text{setting} \quad s := t-\mathrm{i}\beta/2).\end{aligned}$$

Then, we get that

$$\begin{aligned}&\Gamma^{-1}\circ\mathcal{T}\circ\Gamma[\cdot] \\ &= \int_{-\infty}^{\infty}\gamma(\omega)\Lambda(\hat{\boldsymbol{A}}(\omega))(\cdot)\Lambda^{-1}(\hat{\boldsymbol{A}}(\omega)^\dagger)\mathrm{d}\omega \\ &= \int_{-\infty}^{\infty}\int\int\gamma(\omega)\mathrm{e}^{-\mathrm{i}\omega(s-s'+\mathrm{i}\beta)}f(s+\mathrm{i}\beta/2)f(s'-\mathrm{i}\beta/2)\boldsymbol{A}(s)(\cdot)\boldsymbol{A}(s')\mathrm{d}s\mathrm{d}s'\mathrm{d}\omega \\ &\qquad\qquad (\text{setting } s' := t'+\mathrm{i}\beta) \\ &= \frac{1}{\sqrt{2\pi}}\int\int c(s'-s-\mathrm{i}\beta)f(s+\mathrm{i}\beta/2)f(s'-\mathrm{i}\beta/2)\boldsymbol{A}(s)(\cdot)\boldsymbol{A}(s')\mathrm{d}s\mathrm{d}s' \\ &\qquad\qquad (\text{continuing } c(z),\ z\in\mathbb{C}) \\ &= \frac{1}{\sqrt{2\pi}}\int\int c(t'-t-\mathrm{i}\beta)f(t+\mathrm{i}\beta/2)f(t'-\mathrm{i}\beta/2)\boldsymbol{A}(t)(\cdot)\boldsymbol{A}(t')\mathrm{d}t\mathrm{d}t' \quad . \\ &\qquad\qquad (\text{shifting integration})\end{aligned}$$

The third line uses that $(\boldsymbol{A}(z))^\dagger = \boldsymbol{A}(z^*)$ for Heisenberg evolution at complex times. The last line shifts the integrals by $s\longrightarrow t$ and $s'\longrightarrow t'$ (assuming the absence of poles across the strip).

Comparing the coefficient of product integrals over $\boldsymbol{A}(t)\cdot\boldsymbol{A}(t')$ for each t,t', the condition

$$c(t-t')f(t)f(t') = c(t'-t-\mathrm{i}\beta)f(t+\mathrm{i}\beta/2)f(t'-\mathrm{i}\beta/2) \quad \text{for each} \quad t,t'\in\mathbb{R}$$

$$\text{ensures} \quad \mathcal{T}^\dagger = \Gamma^{-1}\circ\mathcal{T}\circ\Gamma.$$

This condition can be rearranged as (whenever the denominators are non-zero)

$$\frac{c(t-t')}{c(t'-t-\mathrm{i}\beta)} = \frac{f(t+\mathrm{i}\beta/2)f(t'-\mathrm{i}\beta/2)}{f(t)f(t')} \quad \text{for each} \quad t,t' \in \mathbb{R}. \tag{3.40}$$

Solving the functional equation

We proceed to solve (3.40) for a fixed β. Consider a change of variable $t \leftrightarrow t'$

$$\frac{c(t'-t)}{c(t-t'-\mathrm{i}\beta)} = \frac{f(t'+\mathrm{i}\beta/2)f(t-\mathrm{i}\beta/2)}{f(t')f(t)}. \tag{3.41}$$

Then, divide the two equations and observe that the LHS depends only on the time difference $t-t'$

$$(3.40)/(3.41) : f_1(t-t') = \frac{f(t+i\beta/2)}{f(t-\mathrm{i}\beta/2)} \cdot \frac{f(t'-\mathrm{i}\beta/2)}{f(t'+\mathrm{i}\beta/2)} := g(t)/g(t').$$

We must have that

$$g(t) = a\mathrm{e}^{bt} \quad \text{for constants} \quad a,b \quad \text{independent of} \quad t.$$

Thus,

$$\begin{aligned} & a\mathrm{e}^{bt} f(t-\mathrm{i}\beta/2) = f(t+\mathrm{i}\beta/2) \\ \Longrightarrow \quad & a\mathrm{e}^{bt}\mathrm{e}^{-i\beta/2\cdot\partial_t} f(t) = \mathrm{e}^{i\beta/2\cdot\partial_t} f(t) && \text{(analyticity)} \\ \Longrightarrow \quad & \mathrm{e}^{a'+b't-i\beta\cdot\partial_t} f(t) = f(t) && \text{(since} \quad [t,\partial_t] \propto 1). \end{aligned}$$

The last equality regards t and ∂_t as linear operators acting on functions and use the matrix-exponential fact that

$$[\boldsymbol{X},\boldsymbol{Y}] \quad \text{commutes with} \quad \boldsymbol{X},\boldsymbol{Y} \quad \text{implies} \quad \mathrm{e}^{\boldsymbol{X}}\mathrm{e}^{\boldsymbol{Y}} = \mathrm{e}^{\boldsymbol{X}+\boldsymbol{Y}+[\boldsymbol{X},\boldsymbol{Y}]/2}.$$

We can get rid of the exponential by regarding it as an eigenproblem

$$\mathrm{e}^{\boldsymbol{X}}|v\rangle = |v\rangle \quad \text{implies} \quad \boldsymbol{X}|v\rangle = \mathrm{i}2\pi\mathbb{Z}\cdot|v\rangle.$$

This amounts to solving the differential equation

$$(a_1\partial_t + a_2 t + a_3) f(t) = 0$$

which has the general solution being Gaussians (the constants may depend on the fixed β.). Plugging back into (3.40), we can solve for $c(t)$, which would not be unique as it allows for linear combinations.

As a sanity check, we re-derive the constraints between the Gaussian parameters and the temperature in the time picture. Let

$$c(x) = e^{-iax}\mathrm{e}^{-x^2/\delta^2}, \qquad and \qquad f(t) = e^{-t^2/\kappa^2},$$

where a, δ, κ are all real parameters. Then, letting $t - t' := x$ we get:

$$\begin{aligned} \log\left[\frac{c(t-t')}{c(t'-t-\mathrm{i}\beta)}\right] &= \log\left[\frac{f(t+\mathrm{i}\beta/2)f(t'-\mathrm{i}\beta/2)}{f(t)f(t')}\right] \\ \implies \log\left[\frac{e^{-iax}\mathrm{e}^{-x^2/\delta^2}}{e^{ia(x+\mathrm{i}\beta)}\mathrm{e}^{-(x+\mathrm{i}\beta)^2/\delta^2}}\right] &= \log\left[\frac{e^{-(t+\mathrm{i}\beta/2)^2/\kappa^2}\mathrm{e}^{-(t'-\mathrm{i}\beta/2)^2/\kappa^2}}{e^{-t^2/\kappa^2-t'^2/\kappa^2}}\right], \\ \text{it suffices if} \quad 2ix(\frac{\beta}{\delta^2} - a) + a\beta - \frac{\beta}{\delta^2} &= \frac{-ix\beta}{\kappa^2} + \frac{\beta^2}{2\kappa^2}. \end{aligned}$$

Equating the imaginary and real parts leads to two linearly dependent equations with the same solution:

$$a = \beta(\frac{1}{\delta^2} + \frac{1}{2\kappa^2}).$$

Then identifying $a \equiv \omega_\gamma$, $\gamma^{-2} = \sigma_\gamma^2/2$, and $\kappa^{-2} = \sigma_E^2$ leads to the same relationship between $\omega_\gamma, \sigma_\gamma^2, \sigma_E^2$ and β as in (3.4).

3.9 Appendix:Other notions of detailed balance

In addition to our KMS detailed balance condition (Definition 3.2.1), other quantum detailed balance has also been studied (see, e.g., [8, 32, 69]). In this section, we will discuss two variants of quantum detailed balance, and only the first one seems to work.

Detailed balance with unitary drift

We constructed a Lindbladian satisfying the KMS detailed balance condition. However, much of our results remain to hold even if we add a suitable Hamiltonian term, which might enable alternative constructions.

Definition 3.9.1 (KMS-detailed balance with unitary drift [69, Section 5])**.** *We say that the Lindbladian $\mathcal{L}$ satisfies $\boldsymbol{\rho}$-detailed balance with unitary drift (in short $\boldsymbol{\rho}$-DBU) with respect to a full-rank state $\boldsymbol{\rho}$ if there exists a Hermitian operator $\boldsymbol{Q}$ such that*

$$\mathcal{L}^\dagger[\cdot] - \sqrt{\boldsymbol{\rho}}^{-1}\mathcal{L}[\sqrt{\boldsymbol{\rho}}\cdot\sqrt{\boldsymbol{\rho}}]\sqrt{\boldsymbol{\rho}}^{-1} = \mathrm{i}[\boldsymbol{Q}, \cdot]. \tag{3.45}$$

Or, in terms of discriminants,

$$\mathcal{A}(\boldsymbol{\rho}, \mathcal{L}) := \frac{\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}) - \mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger}{2} = -\mathrm{i}\boldsymbol{\rho}^{1/4}[\boldsymbol{Q}, \boldsymbol{\rho}^{-1/4}(\cdot)\boldsymbol{\rho}^{-1/4}]\boldsymbol{\rho}^{1/4}.$$

That is, we relax the KMS detailed balance condition (Definition 3.2.1) by allowing the RHS to be any commutator term $\mathrm{i}[\boldsymbol{Q}, \cdot]$[16]; this enlarges the family of possible Lindbladian for the stationary state $\boldsymbol{\rho}$.

Proposition 3.9.1 (Fixed point)**.** *If a Lindbladian $\mathcal{L}$ is $\boldsymbol{\rho}$-detailed-balanced with unitary drift, then*

$$\mathcal{L}[\boldsymbol{\rho}] = 0.$$

Proof. Evaluate the superoperator (3.45) for the identity $\boldsymbol{I}$ and conclude $\sqrt{\boldsymbol{\rho}}^{-1}\mathcal{L}[\sqrt{\boldsymbol{\rho}}\boldsymbol{I}\sqrt{\boldsymbol{\rho}}]\sqrt{\boldsymbol{\rho}}^{-1} = \mathcal{L}^{\dagger}[\boldsymbol{I}] = 0$. ∎

Intuitively, we can certainly add any Hermitian $\boldsymbol{Q}$ that commutes with $\boldsymbol{\rho}$ without changing the stationary state. In fact, this is the only possibility[17].

Proposition 3.9.2 (Structure of $\boldsymbol{Q}$ [69, Lemma 28])**.** *In the setting of Definition 3.9.1, $\boldsymbol{Q}$ must commute with $\boldsymbol{\rho}$.*

Indeed, we can solve for $\boldsymbol{B}$ and $\boldsymbol{Q}$ by modifying the argument for Corollary 3.2.1

$$\begin{aligned}
\boldsymbol{Q} &= \boldsymbol{Q}^{\dagger} \\
&\Updownarrow \\
(\Lambda - \Lambda^{-1})(\boldsymbol{B}) &= \frac{\mathrm{i}}{2}(2\mathcal{I} - (\Lambda + \Lambda^{-1}))(\boldsymbol{R}) \\
&\Updownarrow \\
\sum_{\nu\in B}(\mathrm{e}^{\frac{\beta\nu}{2}} - \mathrm{e}^{\frac{-\beta\nu}{2}})\boldsymbol{B}_{\nu} &= \frac{\mathrm{i}}{2}\sum_{\nu\in B}(2 - \mathrm{e}^{-\frac{\beta\nu}{2}} - \mathrm{e}^{\frac{\beta\nu}{2}})\boldsymbol{R}_{\nu} \\
&\Updownarrow \\
\sum_{\nu\in B\setminus\{0\}}\boldsymbol{B}_{\nu} &= \frac{\mathrm{i}}{2}\sum_{\nu\in B\setminus\{0\}}\tanh\left(\frac{\beta\nu}{4}\right)\boldsymbol{R}_{\nu}. \qquad (\mathrm{e}^{\frac{\beta\nu}{2}} - \mathrm{e}^{\frac{-\beta\nu}{2}} \neq 0 \iff \nu \neq 0)
\end{aligned}$$

Thus, allowing the term $\boldsymbol{Q} \neq 0$ merely amounts to dropping constraints on the $\boldsymbol{B}_0$ component (which is exactly the set of operators that commute with the Hamiltonian).

This relaxed version of detailed balance comes with the conceptual price of making the Lindbladian "non-self-adjoint" under similarity transformation. Due to the anti-self-adjoint component $\mathcal{A}(\boldsymbol{\rho}, \mathcal{L})^{\dagger}$, the spectral theory of convergence (at first glance)

[16] In an email exchange with Jonathan Moussa, he mentioned analogous unitary effect for discrete quantum channels.

[17] We thank Jonathan Moussa for pointing us to [69, Lemma 28]. This further simplifies the derivation and the presentation, clarifying that our construction actually has $\boldsymbol{Q} = 0$ and a vanishing anti-Hermitian component.

seems to break since the right eigenvectors are not orthogonal in the Hilbert-Schmidt norm. Fortunately, the following observation comes to the rescue.

Proposition 3.9.3. *If a Lindbladian $\mathcal{L}$ is $\boldsymbol{\rho}$-detailed-balanced with unitary drift, then*

$$\mathcal{A}(\boldsymbol{\rho}, \mathcal{L})[\sqrt{\boldsymbol{\rho}}] = \mathcal{H}(\boldsymbol{\rho}, \mathcal{L})[\sqrt{\boldsymbol{\rho}}] = 0.$$

In other words, the specific form of detailed balance implies that the anti-self-adjoint component preserves the eigenvector $\sqrt{\boldsymbol{\rho}}$; the particular eigenvector $\sqrt{\boldsymbol{\rho}}$, which we care about, is orthogonal to other eigenvectors. Therefore, the second eigenvalue of $\mathcal{H}$ corresponds to the contraction of the Hilbert-Schmidt norm of the other eigenvector, controlling the mixing time.

Proposition 3.9.4 (Mixing time from spectral gap (adapted from [40, Proposition II.2]))**.** *If a Lindbladian $\mathcal{L}$ satisfies $\boldsymbol{\rho}$-DBU, then*

$$t_{mix}(\mathcal{L}) \leq \frac{\ln(2\|\boldsymbol{\rho}^{-1/2}\|)}{\lambda_{gap}(\mathcal{H})} \quad \textit{where} \quad \mathcal{H} := \frac{\mathcal{D}(\boldsymbol{\rho}, \mathcal{L})^\dagger + \mathcal{D}(\boldsymbol{\rho}, \mathcal{L})}{2},$$

and the mixing time t_{mix} is the smallest time for which

$$\left\|\mathrm{e}^{\mathcal{L}t_{mix}}[\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2]\right\|_1 \leq \frac{1}{2}\|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\|_1 \quad \textit{for any states} \quad \boldsymbol{\rho}_1, \boldsymbol{\rho}_2.$$

In other words, adding a coherent term only seems to help with the convergence.

Proof. Write $\boldsymbol{R} = \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2$, then

$$\begin{aligned}
\left\|\mathrm{e}^{\mathcal{L}t}[\boldsymbol{R}]\right\|_1 &= \left\|\boldsymbol{\rho}^{1/4}\mathrm{e}^{\mathcal{D}t}[\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}]\boldsymbol{\rho}^{1/4}\right\|_1 \\
&\leq \left\|\boldsymbol{\rho}^{1/4}\right\|_4 \cdot \left\|\mathrm{e}^{\mathcal{D}t}[\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}]\right\|_2 \cdot \left\|\boldsymbol{\rho}^{1/4}\right\|_4 \\
&\leq \mathrm{e}^{-\lambda_{gap}(\mathcal{H})t}\left\|\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}\right\|_2 \\
&\leq \mathrm{e}^{-\lambda_{gap}(\mathcal{H})t}\|\boldsymbol{\rho}^{-1/4}\|^2\|\boldsymbol{R}\|_2 \\
&\leq \mathrm{e}^{-\lambda_{gap}(\mathcal{H})t}\|\boldsymbol{\rho}^{-1/4}\|^2\|\boldsymbol{R}\|_1 \\
&= \mathrm{e}^{-\lambda_{gap}(\mathcal{L}^\dagger)t}\|\boldsymbol{\rho}^{-1/2}\|\|\boldsymbol{R}\|_1.
\end{aligned}$$

The first inequality uses Hölder's inequality. The second inequality uses the orthogonality to the leading eigenvector such that $\mathrm{Tr}[\sqrt{\boldsymbol{\rho}}\cdot\boldsymbol{\rho}^{-1/4}\boldsymbol{R}\boldsymbol{\rho}^{-1/4}] = \mathrm{Tr}[\boldsymbol{R}] = 0$. Take the logarithm to conclude the proof. ∎

However, if we wish to obtain a parent Hamiltonian corresponding to this Lindbladian, the anti-Hermitian component forces us to consider the *symmetrized discriminant*

$$\mathcal{H}_\beta = \frac{\mathcal{D} + \mathcal{D}^\dagger}{2}.$$

Unfortunately, the presence of the anti-Hermitian component breaks the qualitative analogy between mixing times and discriminant gaps; fast mixing can hold despite a small symmetrized discriminant gap, marking a limitation of the parent Hamiltonian approach. Incorporating a coherent term, the general conversion bound is as follows [40]; unfortunately, some bounds are only meaningful if the anti-Hermitian part $\mathcal{A}$ is small enough.

Proposition 3.9.5 (Spectral gap from mixing time [40, Proposition E.5])**.** *For any Lindbladian $\mathcal{L}$, let $-\lambda_{\mathrm{Re}(gap)}(\mathcal{L})$ be the second largest real part in its spectrum (counted by algebraic multiplicity), then*

$$\lambda_{gap}(\mathcal{H}) + 2\|\mathcal{A}\|_{2-2} \geq \|\mathcal{A}\|_{2-2} - \lambda_2(\mathcal{H}) \geq \lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \geq \frac{\ln(2)}{t_{mix}(\mathcal{L})}. \quad (3.46)$$

Moreover, if $\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \geq 2\|\mathcal{A}\|_{2-2}$, then there is unique eigenvalue $\lambda_1(\mathcal{H}) \geq -\|\mathcal{A}\|_{2-2}$ and

$$\lambda_{gap}(\mathcal{H}) + 2\|\mathcal{A}\|_{2-2} \geq \|\mathcal{A}\|_{2-2} - \lambda_2(\mathcal{H}) \geq \lambda_{\mathrm{Re}(gap)}(\mathcal{L}). \quad (3.47)$$

Corollary 3.9.1 (Spectral gap from mixing time)**.** *If a Lindbladian $\mathcal{L}$ satisfies ρ-DBU and $\frac{\ln(2)}{2t_{mix}(\mathcal{L})} \geq \|\mathcal{A}\|_{2-2}$, then*[18]

$$\lambda_{gap}(\mathcal{H}) \geq \frac{\ln(2)}{2t_{mix}(\mathcal{L})}.$$

Proof. Observe that $\frac{\ln(2)}{2t_{mix}(\mathcal{L})} \geq \|\mathcal{A}\|_{2-2}$ implies $\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \geq 2\|\mathcal{A}\|_{2-2}$ due to (3.46), which in turn by (2.79) also implies that $\lambda_1(\mathcal{H}) = 0$ (since 0 is an eigenvalue due to Proposition 3.9.3). Thus combining (3.46)-(3.47) yields

$$\|\mathcal{A}\|_{2-2} + \lambda_{\mathrm{gap}}(\mathcal{H}) \geq \lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \geq \frac{\ln(2)}{t_{mix}(\mathcal{L})}.$$ ∎

[18] The bound cannot be qualitatively strengthened without additional structural understanding or assumptions, e.g., consider $\boldsymbol{H} := \boldsymbol{Z} - (1+\epsilon)\boldsymbol{I}$, $\boldsymbol{A} := \mathrm{i}\boldsymbol{X}$; the eigenvalues of $\boldsymbol{H} + \boldsymbol{A}$ are $-1-\epsilon$, while $\lambda_1(\boldsymbol{H}) = -\epsilon$, $\|\boldsymbol{A}\| = 1$, and the relaxation (mixing) time is $O(1)$.

The proof of the above proceeds by showing that $\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \geq \frac{\ln(2)}{t_{mix}(\mathcal{L})}$, however without any prior knowledge on $\|\mathcal{A}\|_{2-2}$ it does not seem to be possible to lower bound $\lambda_{gap}(\mathcal{H})$ in general. Indeed, if the Lindbladian $\mathcal{L}$ satisfies $\boldsymbol{\rho}$-detailed balance with unitary drift, then $\mathcal{L}$ and $\mathcal{D}(\boldsymbol{\rho}, \mathcal{L}) = \mathcal{H} + \mathcal{A}$ are related by a similarity transform and are thus co-spectral. Due to Proposition 3.9.3 we know that $V := \mathrm{Span}(\sqrt{\boldsymbol{\rho}})^{\perp}$ is an invariant subspace of both $\mathcal{H}$ and $\mathcal{A}$, and thus also of $\mathcal{L}$. Since $\mathcal{H}_{|V} \leq -\lambda_{gap}(\mathcal{H})\mathcal{I}$ and $\mathcal{A}_{|V}$ is anti-Hermitian we get that every eigenvalue of $\mathcal{L}^{\dagger}_{|V}$ has real part at most $-\lambda_{gap}(\mathcal{H})$, thus $\lambda_{\mathrm{Re}(gap)}(\mathcal{L}) \geq \lambda_{gap}(\mathcal{H})$, which is an inequality in the "wrong" direction.

s-detailed balance

Our KMS detailed balance condition (Definition 3.9.2) is a special case ($s = 1/2$) of a larger family of s-detailed balance condition.

Definition 3.9.2 (s-detailed balance condition)**.** *For a normalized, full-rank state $\boldsymbol{\rho} \succ 0$ and an scalar $0 \leq s \leq 1$, we say that an super-operator $\mathcal{L}$ satisfies $(s, \boldsymbol{\rho})$-detailed balance (or s-DB in short) if*

$$\mathcal{L}^{\dagger}[\cdot] = \boldsymbol{\rho}^{s-1}\mathcal{L}[\boldsymbol{\rho}^{1-s} \cdot \boldsymbol{\rho}^{s}]\boldsymbol{\rho}^{-s}.$$

Since $\boldsymbol{\rho}$ is an operator, different choices of $0 \leq s \leq 1$ yield different detailed balance conditions. Nevertheless, they all prescribe the same fixed point.

Proposition 3.9.6 (Fixed point)**.** *If a Lindbladian $\mathcal{L}$ is $(s, \boldsymbol{\rho})$-detailed-balanced, then $\mathcal{L}[\boldsymbol{\rho}] = 0$.*

The case $s = 0$ corresponds to the so-called Gelfand-Naimark-Segal (GNS) inner product. One naturally wonders if our constructions also apply here. Unfortunately, the case $s = 1/2$ appears to be a special point in the interval $0 \leq s \leq 1$. Let us revisit the energy domain representation of detailed balance (Proposition 3.2.2) for the transition part. Consider a super-operator parameterized by a Hamiltonian $\boldsymbol{H}$, β, and a set of operators including its adjoints $\{\boldsymbol{A}^a : a \in A\} = \{\boldsymbol{A}^{a\dagger} : a \in A\}$:

$$\mathcal{T} = \sum_{a \in A} \sum_{\nu_1, \nu_2 \in B} \alpha_{\nu_1, \nu_2} \boldsymbol{A}^a_{\nu_1}(\cdot)(\boldsymbol{A}^a_{\nu_2})^{\dagger}.$$

Then, the s-detailed balance condition

$$\mathcal{T}^{\dagger}[\cdot] = \boldsymbol{\rho}^{s-1}\mathcal{T}[\boldsymbol{\rho}^{1-s} \cdot \boldsymbol{\rho}^{s}]\boldsymbol{\rho}^{-s} \quad \text{demands that}$$
$$\alpha_{\nu_1, \nu_2} = \alpha_{-\nu_2, -\nu_1} \exp(-\beta(1-s)\nu_1 - \beta s \nu_2) \quad \text{for each} \quad \nu_1, \nu_2 \in B.$$

However, this appears to be a strong condition; if we apply a change of variable $(\nu_1, \nu_2) \to (-\nu_2, -\nu_1)$,

$$\alpha_{-\nu_2,-\nu_1} = \alpha_{\nu_1,\nu_2} \exp(\beta(1-s)\nu_2 + \beta s \nu_1) \quad \text{for each} \quad \nu_1, \nu_2 \in B.$$

Multiply the two to see that (see [32, Lemma 2.5] for an abstract argument)

$$\alpha_{\nu_1,\nu_2}\alpha_{-\nu_2,-\nu_1} = \alpha_{\nu_1,\nu_2}\alpha_{-\nu_2,-\nu_1} \exp(\beta(1-2s)(\nu_2 - \nu_1)) \quad \text{for each} \quad \nu_1, \nu_2 \in B.$$

Therefore, if $s \neq \frac{1}{2}$ and $\beta \neq 0$, we must have that

$$\alpha_{\nu_1,\nu_2} = 0 \quad \text{if} \quad \nu_1 \neq \nu_2 \tag{3.48}$$

which contradicts our construction, and currently, we do not know how to algorithmically ensure (3.48). The only existing Lindbladian we knew that satifies (3.48) is the Davies' generator [56], which requires resolving the level spacing using a (exponentially) long Hamiltonian simulation time.

Chapter 4

COOLING IS UNIVERSAL FOR QUANTUM COMPUTATION

4.1 Introduction

Finding ground states and other low-energy states of quantum many-body systems is a central problem in physics, materials science, and chemistry. To address this problem, many powerful computational methods, such as density functional theory (DFT) [89, 103], quantum Monte Carlo (QMC) [17, 34, 155], variational optimization with tensor network ansatzes [82, 93, 145, 156, 173, 177, 178] or neural network ansatzes [33, 59, 88], and data-driven machine learning approaches [71, 91, 111, 147], have been developed. These methods, which can be implemented using classical computers, work well for many physically relevant problem instances but fail badly in other cases.

Quantum computers may be able to find ground states efficiently in cases that are too hard to solve classically, but conclusively identifying such cases is challenging. Even when known classical methods falter, one cannot easily rule out that alternative classical methods will have much better performance. Furthermore, finding ground states of local Hamiltonians is known to be QMA-hard [98, 100], and therefore is expected to be intractable even for quantum computers for some physical systems. To establish the efficacy of existing quantum algorithms for the ground-state problem, assumptions are required that might not be justified [108], such as the presence of a trial state with sufficient ground state overlap [70, 114] or a parameterized adiabatic path whose spectral gap remains open [65].

If finding ground states is hard for quantum computers, we expect that the task is hard for Nature as well. In searching for physics problems where quantum computers could have an advantage over classical computers, we should follow Nature's guidance. With that in mind, we observe that when a quantum system is placed in a low-temperature thermal bath [28, 55, 115, 135], it seeks a *local* minimum of the energy which may be far from the *global* minimum. For some physical systems, such as spin glasses [23, 63, 99, 136], finding a ground state is indeed known to be hard; such systems, when cooled, almost always find a local minimum instead of the ground state. In these cases, the ground state of the Hamiltonian is physically *irrelevant* because it is unlikely to be observed in experiments.

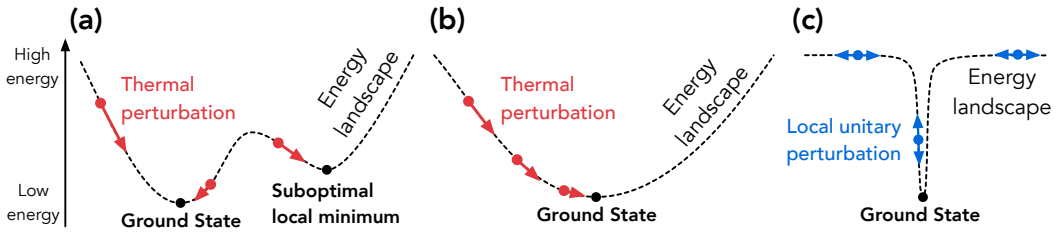


Figure 4.1: *(a) Energy landscape with multiple local minima.* Some Hamiltonians, such as magnets in a small external magnetic field, have multiple local minima. One is the ground state while the other local minima are suboptimal. *(b) Energy landscape with one local minimum.* For some local Hamiltonians, such as a family of BQP-hard Hamiltonians, the energy landscape over the entire n-qubit state space has a nice bowl shape, and the only local minimum is the global minimum. However, for QMA-hard Hamiltonians, the energy landscape necessarily contains many suboptimal local minima. *(c) Energy landscape under local unitary perturbations.* For any local Hamiltonian $\boldsymbol{H}$, there are always doubly exponentially many local minima within the n-qubit state space that stems from a large barren plateau. Local unitary perturbations are reversible, while thermal perturbations are irreversible.

This perspective encourages us to formulate and study a physically motivated alternative to the ground-state problem. We ask:

How tractable is the problem of finding local minima of the energy in quantum systems using classical and quantum computers?

We address this question by proving that a machine that cools physical systems to local minima is a universal quantum computer; that is, by finding local minima one can solve any problem that is efficiently solvable by a quantum computer, and furthermore a quantum computer can always find local minima efficiently for any physical system. Hence, under the widely accepted assumption that quantum computing is more powerful than classical computing, finding local minima is both classically hard and quantumly easy. This result raises the hope that exploring local minima in quantum systems arising from physics, chemistry, and materials science may offer new opportunities for solving classically intractable and physically relevant problems using quantum computers.

4.2 Results

We now present our main results concerning the tractability of finding local minima in quantum systems. A collection of notational conventions and some background

on quantum thermodynamics can be found in Appendix 4.5. Based on the standard definition in mathematical optimization [2, 5, 25, 94, 144], we consider a local minimum in a quantum system governed by Hamiltonian $\boldsymbol{H}$ to be a quantum state such that the expectation value of $\boldsymbol{H}$ does not decrease under any small perturbation applied to the state; see a brief review of mathematical optimization in Appendix 4.6.

Let $\mathcal{P}_\alpha$ be a perturbation parameterized by a small vector $\boldsymbol{\alpha}$ that maps quantum states to quantum states. An ϵ-approximate local minimum of an n-qubit Hamiltonian $\boldsymbol{H}$ under perturbation $\mathcal{P}$ is a state $\boldsymbol{\rho}$ with an energy $\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})$ that is an approximate minimum under perturbations, i.e.,

$$\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}) \leq \mathrm{tr}(\boldsymbol{H}\mathcal{P}_\alpha(\boldsymbol{\rho})) + \epsilon\|\boldsymbol{\alpha}\|$$

for all small enough $\boldsymbol{\alpha}$. The formal definition is given in Appendix 4.6. The local minima of $\boldsymbol{H}$ form a subset of the entire quantum state space, which contains the global minima, the ground states of $\boldsymbol{H}$. See Fig. 4.1 for energy landscapes with different numbers of local minima. We say an algorithm $\mathcal{A}$ has solved the problem of *finding* local minima under perturbation $\mathcal{P}$ if given any n-qubit Hamiltonian $\boldsymbol{H}$, written as a sum of few-qubit Hermitian operators, and any few-qubit observable $\boldsymbol{O}$, the algorithm $\mathcal{A}$ can output a real value $\mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho})$ corresponding to any approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}$ under perturbations $\mathcal{P}$ up to a small error.[1]

Inspired by the physical process that Nature uses to cool quantum systems, we define local minima based on *thermal perturbations* rooted in open system thermodynamics [28, 39, 55, 115, 135]. Thermal perturbations are fundamentally different from *local unitary perturbations*, which are short-time unitary evolution governed by a sum of few-body Hermitian operators, as may arise in an adaptive variational quantum eigensolver (VQE) [35, 77, 143]. These two types of perturbations yield distinct energy landscapes as illustrated in Fig. 4.1, leading to vastly different computational complexity for the task of finding a local minimum. Indeed, we argue in Section 4.2 that for local unitary perturbations, in contrast to thermal perturbations, finding a local minimum is classically easy.

Local minima under thermal perturbations

In this section, we consider local minima under thermal perturbations induced by a heat bath, formally defined in Appendix 4.6. When the coupling between an n-qubit system and a thermal bath is weak, and the bath is memoryless, the complicated joint

[1] Since there could be multiple local minima and we consider finding one instance to be sufficient, this problem is closer to a *relational* problem than to a *decision* problem.

system-bath Hamiltonian dynamics reduces to a Markovian Lindbladian evolution of the system alone, $\boldsymbol{\rho}(t) = \mathrm{e}^{\mathcal{L}t}[\boldsymbol{\rho}]$. Remarkably, this continuous time generator $\mathcal{L}$ can be defined by merely the system Hamiltonian $\boldsymbol{H}$, the *jump operators* $\boldsymbol{A}^a$ through which the bath interacts with the system, and thermodynamic quantities of the bath: inverse temperature β and a characteristic time-scale τ. See Appendix 4.5 for an introduction and Appendix 4.10 for an in-depth discussion. Under these assumptions, we may effectively consider a thermal perturbation of n-qubit state $\boldsymbol{\rho}$ to be

$$\text{(thermal perturbation):} \qquad \boldsymbol{\rho} \to \exp\left(\sum_{a=1}^{m} \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho}), \tag{4.1}$$

where $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ is the thermal Lindbladian associated with each jump operator $\boldsymbol{A}^a$ acting on a few qubits, $m = \mathrm{poly}(n)$ is the number of jump operators, and $\boldsymbol{\alpha} = \sum_a \alpha_a \hat{\boldsymbol{e}}_a \in \mathbb{R}^m_{\geq 0}$ is a *nonnegative* vector close to zero. Here, the vector is nonnegative because thermodynamic processes are generally irreversible. The irreversibility in thermal perturbations is crucial to ensure that there are fewer than doubly-exponentially many local minima in the energy landscape; see the discussion in Appendix 4.6.

We define a local minimum under thermal perturbations to be a state $\boldsymbol{\rho}$ with the minimum energy $\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})$ under thermal perturbations given in Eq. (4.1). Given the definition, we next study how tractable is the problem of finding a local minimum under thermal perturbations. Our complexity-theoretic results show that finding local minima under thermal perturbations is both quantumly easy (Section 4.2) and classically hard (Section 4.2), the latter under the well-accepted assumption that not all quantum circuits can be efficiently simulated on classical computers ($\mathsf{BPP} \neq \mathsf{BQP}$).

Finding local minima is easy for quantum computers

If our notion of local minima properly captures how a quantum system behaves in a cold environment, we expect finding local minima to be quantumly easy. Indeed, in the following theorem, we prove that a quantum computer can always efficiently find a local minimum of $\boldsymbol{H}$ under thermal perturbations starting from any initial state.

Theorem 4.2.1 (Quantumly easy to find local minima under thermal perturbations; informal)**.** *The problem of finding an ϵ-approximate local minimum of an n-qubit local Hamiltonian $\boldsymbol{H}$ under thermal perturbations with inverse temperature β and time scale τ can be solved in* $\mathrm{poly}(n, 1/\epsilon, \beta, \tau)$ *quantum computational time.*

The formal statement is given in Theorem 4.9.1 and is proven in Appendix 4.11. To establish the theorem, we propose a *quantum thermal gradient descent algorithm* that cools down a quantum system by performing gradient descent based on quantum thermodynamics. This result builds on the recent algorithmic developments for quantum Markov Chain Monte Carlo methods [41, 61, 149, 167].

Finding local minima is hard for classical computers

While local minima are easier to find than ground states for quantum computers, how hard is it to find them? As our second main result, we address this question by considering a class of geometrically local Hamiltonians $\{\boldsymbol{H}_C\}$ on a 2D lattice, such that the ground state of $\boldsymbol{H}_C$ encodes the outcome of a polynomial-size quantum circuit C using a modified Kitaev's circuit-to-Hamiltonian construction [3, 100, 142]. To understand the computational hardness, we provide the following characterization of the energy landscape.

Theorem 4.2.2 (No suboptimal local minimum in BQP-hard Hamiltonians; informal)**.** *For any quantum circuit C with size $|C|$, all approximate local minima of the geometrically local 2D Hamiltonian $\boldsymbol{H}_C$ under thermal perturbations with inverse temperature $\beta = \mathrm{Poly}(|C|)$ and time scale $\tau = \mathrm{Poly}(|C|)$ are close to the ground state.*

This theorem is the most technically involved contribution of this work. The formal statement is given in Theorem 4.9.2 and is proven in Appendix 4.14. Conceptually, the landscape of these 2D Hamiltonians has a nice *bowl shape*, like in convex optimization [25]. Therefore, performing thermal gradient descent (Theorem 4.2.1) allows us to prepare the ground state starting from an *arbitrary* initial state. To prove the theorem, we develop a set of techniques for establishing that a Hamiltonian $\boldsymbol{H}$ has *no suboptimal local minima*, i.e., all local minima of $\boldsymbol{H}$ are global minima.

Using this energy landscape characterization, we can show that finding a local minimum under thermal perturbations is classically intractable, assuming quantum computation is more powerful than classical computation. Suppose that a classical computer can efficiently find any local minima under thermal perturbations. Then by Theorem 4.2.2, the classical computer can efficiently find the ground state of $\boldsymbol{H}_C$ and thus simulate any efficient quantum circuit C, which is widely believed to be impossible. See Theorem 4.9.3 for a formal statement and the proof.

Theorem 4.2.3 (Classically hard to find local minima under thermal perturbations; informal)**.** *Assume the widely believed conjecture that* BPP $\neq$ BQP. *The problem of finding an approximate local minimum of an n-qubit local Hamiltonian* $\boldsymbol{H}$ *under thermal perturbations is universal for quantum computation and is thus classically hard.*

Local minima under local unitary perturbations

Instead of defining local minima in terms of thermal perturbations, one could consider *local unitary* perturbations instead. Local unitary perturbations are mathematically simple and commonly considered in adaptive variational quantum eigensolvers [35, 77, 143]. However, as we now explain, the task of finding a local minimum with respect to unitary perturbations is classically easy and, therefore, does not fulfill our goal of identifying a problem that separates quantum from classical computation.

Local unitary perturbations are short-time unitary evolutions under a sum of few-body Hermitian operators (e.g., a quantum circuit consisting of near-identity two-qubit gates). A local unitary perturbation of an n-qubit pure state $|\psi\rangle$ is given by

$$\text{(local unitary perturbation):} \qquad |\psi\rangle \to \exp\left(-\mathrm{i}\sum_{a=1}^{m}\alpha_a \boldsymbol{h}^a\right)|\psi\rangle,$$

where $\boldsymbol{h}^a$ is a Hermitian operator acting on a few qubits, $m = \mathrm{poly}(n)$ is the number of such Hermitian operators, and $\boldsymbol{\alpha} = \sum_a \alpha_a \hat{\boldsymbol{e}}_a \in \mathbb{R}^m$ is a vector close to zero. This definition is inspired by adaptive VQE [35, 77, 143], and is the state version of the Riemannian geometry of quantum computation defined in [140]. When one variationally minimizes the energy by applying unitary gates, one finds a local minimum under local unitary perturbations.

We prove that for any Hamiltonian $\boldsymbol{H}$, a random n-qubit pure state is almost always a local minimum of $\boldsymbol{H}$ under local unitary perturbations; see Lemma 4.4.1 in Methods. Hence, there are $\exp(\exp(\Omega(n)))$ many suboptimal local minima in the energy landscape, i.e., a large barren plateau. Because the number of local minima is enormous, the properties of these local minima concentrate on fixed values, making this problem *classically easy*.

Proposition 4.2.1 (Classically easy to find local minima under local unitary perturbations; informal)**.** *The problem of finding approximate local minima of n-qubit local Hamiltonian* $\boldsymbol{H}$ *under local unitary perturbations is classically easy.*

See Proposition 4.9.1 for the detailed statement and Appendix 4.9 for its proof. Aside from being classically easy, the landscape defined by local unitary perturbation is not physically well motivated because a physical system in contact with a thermal bath cools by open system dynamics rather than by unitary dynamics acting on the system.

4.3 Discussion

We have good reasons for believing that scalable fault-tolerant quantum computers will be more powerful than classical computers, but for what problems of practical interest should we expect a superpolynomial quantum advantage? Quantum computers might substantially speed up the task of characterizing properties of ground states for some local Hamiltonians that arise in physics, chemistry, and materials science, but it is not clear how to identify particular problems for which such speedups occur [109]. In some cases, classical methods provide good solutions, while in other cases, the problem is hard even for quantum computers.

Here, we have focused on an easier problem, namely finding local minima rather than global minima of a Hamiltonian. This problem is well motivated physically because the task of finding a local minimum under thermal perturbations is routinely carried out by actual physical systems when in contact with a cold thermal bath. We showed that this problem is solved efficiently by a proposed quantum optimization algorithm, the *quantum thermal gradient descent algorithm*. Furthermore, we showed that finding a local minimum is classically hard in general (assuming that $\mathsf{BPP} \neq \mathsf{BQP}$). Hence, the local minimum problem is a quantumly tractable alternative to the ground state problem for which superpolynomial quantum advantage can be achieved for some problem instances.

There have been other proposals for solving classically hard problems by finding suitable quantum states, such as designing a gapped adiabatic path for Hamiltonians to find ground states [3], engineering Lindbladians to have rapid dissipative evolution towards steady states [174] and performing quantum phase estimation on an initial state with high ground-state overlap [70]. These approaches draw inspiration from physics to motivate algorithms for solving problems on analog and digital quantum devices but do not emulate naturally occurring physical processes. In contrast, the problem of finding a local minimum is motivated by ubiquitous physical processes in Nature that produce the low-energy states studied in physics, chemistry, and materials science. Furthermore, the local minima problem enjoys the robustness of

thermodynamics: one merely needs to specify macroscopic bath quantities β and τ without worrying about microscopic details, and the choice of jump operators can be flexible since adding more jumps (even unwanted ones) only *removes* suboptimal local minima.

There are a plethora of classical algorithms for minimizing energies of quantum systems based on classical variational ansatzes for quantum states, such as tensor networks [1, 13, 52, 82, 93, 95, 106, 145, 156, 162, 163, 172, 173, 177, 178, 183, 187] and neural quantum states [33, 46, 47, 59, 60, 67, 74, 88, 120, 131, 141, 175]. These classical algorithms find a local minimum within a family of states defined by the classical variational ansatz. However, such a local minimum for the classical ansatz might not be a local minimum under thermal perturbations. If not, we can load the state found by the classical algorithm into a quantum computer and find a lower energy state by running the quantum thermal gradient descent algorithm. A corollary of our main results states the following.

Corollary 4.3.1 (Quantum advantage in finding lower-energy state; informal)**.** *Assume that not all polynomial-size quantum circuits can be efficiently simulated classically. Then there are 2D geometrically local Hamiltonians such that given any classical ansatz that allows efficient estimation of single-qubit observables and an output state $\boldsymbol{\rho}^{\#}$ of any efficient classical algorithm that optimizes the classical ansatz, running quantum thermal gradient descent starting at $\boldsymbol{\rho}^{\#}$ strictly lowers the energy.*

A formal statement is in Corollary 4.9.1, and its proof is in Appendix 4.9. In many cases, we can check if the classically optimized state $\boldsymbol{\rho}^{\#}$ is a local minimum under thermal perturbation by executing an efficient classical computation to compute the energy gradient. A negative energy gradient confirms that a quantum algorithm starting from $\boldsymbol{\rho}^{\#}$ could outperform the classical algorithm.

Furthermore, we may be able to do better than quantum thermodynamics by sharpening our quantum algorithms. Indeed, we may modify the thermal Lindbladians to have nicer analytic properties or lower algorithmic cost [39]. While these synthetic Lindbladians may not simulate Nature, they constitute a broader class of Quantum Markov Chain Monte Carlo algorithms [39, 41, 61, 149, 167] that may improve upon Nature. Apart from Lindbladians, other families of perturbations, such as unitary perturbations accompanied by mid-circuit measurements and/or qubit resets, may also yield nice bowl-shaped energy landscapes without suboptimal local minima. Progress

in this direction could lead to more efficient quantum optimization algorithms for finding low-energy states or for other applications.

Many other interesting and challenging questions remain open. Sometimes, when a system performs a random walk over a large plateau of suboptimal local minima for a sufficiently long time, the system escapes the plateau and reaches the true ground state (see, e.g., Case 1 in section 4.13). Could we characterize when ground states can be found efficiently despite having many suboptimal local minima? Our conclusion that finding local minima under thermal perturbations is classically hard relied on the complexity-theoretic conjecture that BPP $\neq$ BQP. Can we prove unconditionally that finding local minima is hard for classical algorithms, perhaps within a black-box oracle model? We have shown that there is a quantum advantage in finding local minima of quantum systems. Might there also be a quantum advantage in finding better local minima in classical optimization problems under some variant of quantum thermal gradient descent?

While ground state problems are hard to solve in general, many experimentally observed quantum systems efficiently relax to their ground states when cooled. This physical phenomenon suggests that perhaps many Hamiltonians of interest in physics, chemistry, and materials science have no suboptimal local minima. We have shown in Theorem 4.2.2 that a particular family of BQP-hard Hamiltonians has no suboptimal local minima under thermal perturbation. An important future goal is to characterize broader classes of Hamiltonians that have a similarly good energy landscape. Our proposed negative gradient condition suffices to rule out suboptimal local minima (Lemma 4.8.3), but checking this condition for a general Hamilton involves highly complex calculations. It would be helpful to develop more general-purpose and efficient methods to verify this property for specified physical Hamiltonians over spins, fermions, or bosons. We hope the ideas and techniques presented here will yield a deeper understanding of the energy landscapes of quantum systems and point toward promising opportunities for achieving quantum advantage for physically relevant problems.

Data availability

We do not analyse or generate any datasets, because our work proceeds within a theoretical and mathematical approach.

Code availability

We do not have any computer code, because our work proceeds within a theoretical and mathematical approach. All algorithms are analyzed mathematically in the Supplementary Information.

Acknowledgments:

The authors thank Anurag Anshu, Ryan Babbush, Fernando Brandao, Garnet Chan, Sitan Chen, Soonwon Choi, Jordan Cotler, Jarrod R. McClean, and Mehdi Soleimanifar for their valuable input. HH thanks Patrick Coles, Gavin Crooks, and Faris Sbahi for the inspiring discussions and for sharing their recent works on classical thermodynamics for AI applications [6, 50]. CFC is supported by the AWS Center for Quantum Computing internship. HH is supported by a Google PhD fellowship and a MediaTek Research Young Scholarship. HH acknowledges the visiting associate position at the Massachusetts Institute of Technology. LZ acknowledges funding from the Walter Burke Institute for Theoretical Physics at Caltech. JP acknowledges support from the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (DE-NA0003525, DE-SC0020290), the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator, and the National Science Foundation (PHY-1733907). The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center.

4.4 Methods

Local minima under thermal perturbations

A central concept that enables us to understand the energy landscape and establish the computational complexity of finding local minima under thermal perturbations in Theorem 4.2.1 and 4.2.3 is the *energy gradient operator*,

$$\text{(energy gradient operator):} \qquad \sum_{a=1}^{m} \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\hat{\boldsymbol{e}}_a,$$

where the adjoint $\mathcal{L}^\dagger$ is the Heisenberg-picture Lindbladian, i.e., $\mathrm{tr}(\mathcal{L}^\dagger[\boldsymbol{O}]\boldsymbol{\rho}) = \mathrm{tr}(\boldsymbol{O}\mathcal{L}[\boldsymbol{\rho}])$. The energy gradient operator is a vector of individual gradient operators[2] associated with each jump operator $\boldsymbol{A}^a$. Indeed, the energy gradient operator naturally

[2]This is similar to the spin operator $\boldsymbol{\sigma} = \sigma^x\hat{x} + \sigma^y\hat{y} + \sigma^z\hat{z}$, which is a vector of Hermitian observables.

emerges by taking an infinitesimal perturbation, i.e., the gradient of the energy $\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})$,

$$\mathrm{tr}\left(\boldsymbol{H}\exp\left(\sum_{a=1}^{m}\alpha_a\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho})\right)=\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})+\boldsymbol{\alpha}\cdot\sum_{a=1}^{m}\mathrm{tr}\left(\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\boldsymbol{\rho}\right)\hat{\boldsymbol{e}}_a+O(\|\boldsymbol{\alpha}\|^2).$$

In Appendix 4.8, we describe the formal definition and some properties of the energy gradient. We provide a concrete example by showing the sets of local minima for ferromagnetic Ising chains under different longitudinal field strengths in Appendix 4.13.

Quantum thermal gradient descent

To establish Theorem 4.2.1 showing that finding local minima under thermal perturbations is quantumly easy, we propose an algorithm called *quantum thermal gradient descent*. Gradient descent is necessary when the inverse temperature β and time scale τ are not infinite. When $\beta=\tau=\infty$, the energy gradient $\mathcal{L}_a^{\dagger\infty,\infty,\boldsymbol{H}}(\boldsymbol{H})\preceq 0$ is nonpositive. In this case, the algorithm can just perform a random walk along random directions because no perturbations increase energy, and the use of gradient descent is not necessary. But when β and τ are finite, the energy gradient can be positive. To find a local minimum that is a minimum under all thermal perturbations, the algorithm needs to carefully walk in directions with negative energy gradients.

To prove the convergence of quantum thermal gradient descent to a local minimum, we show that every small gradient step starting from a state that is not an approximate local minimum will decrease the energy. Because the energy of a quantum system cannot go unboundedly to negative infinity, the algorithm must terminate at some point, which means it will find an approximate local minimum. To establish this claim, we derive analytic properties of thermal Lindbladians based on a smoothness bound on the second derivatives in [39]. To implement a gradient step based on thermal perturbations on a quantum computer, we build on a recently developed efficient quantum algorithm that simulates thermal Lindbladian evolution using a quantum circuit augmented by mid-circuit measurements [39].

All local minima are global

We now highlight the proof idea for Theorem 4.2.2 showing that all approximate local minima are close to ground states in a family of Hamiltonians with classically-hard ground states. The family of geometrically local n-qubit Hamiltonians $\{\boldsymbol{H}_C\}$ in a

2D lattice studied in this work is a modification of Kitaev's circuit-to-Hamiltonian construction [100, 142] where the ground state encodes the computation of a quantum circuit $\boldsymbol{U}_C = \boldsymbol{U}_T \dots \boldsymbol{U}_1$. In particular, we design the ground state of $\boldsymbol{H}_C$ to be

$$\sum_{t=0}^{T} \sqrt{\xi_t}\left(\boldsymbol{U}_t \cdots \boldsymbol{U}_1 |0^n\rangle\right) \otimes \left|1^t 0^{T-t}\right\rangle, \qquad \text{where} \quad \xi_t := \frac{1}{2^T}\binom{T}{t}.$$

The binomial coefficient ξ_t is our modification of Kitaev's construction and is chosen to ensure that desired properties hold for the spectrum and the energy gradients.[3] Estimating local properties of the ground state of $\boldsymbol{H}_C$ is equivalent to simulating the quantum circuit C, which is BQP-hard. Hence, under the standard assumption that not all quantum circuits can be simulated by classical computations, estimating local properties of the ground state of $\boldsymbol{H}_C$ is classically hard.

Given the Hamiltonian $\boldsymbol{H}_C$, showing that all of its approximate local minima under thermal perturbations are also approximate global minima seems daunting due to the complex expression for the thermal Lindbladian $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ and the doubly exponentially large space of possible quantum states. Previous studies on circuit-to-Hamiltonian mappings mainly focused on the lowest energy states. Here, we need to worry about potential local minima in all excited states in any superposition. To make progress, we propose a sufficient condition in Appendix 4.8 that captures the nice landscape of $\boldsymbol{H}_C$ and rules out the presence of *any* suboptimal local minimum. Let $\boldsymbol{P}_G(\boldsymbol{H})$ be the projector onto the ground state space of $\boldsymbol{H}$. Assume there exists a unit vector $\hat{\boldsymbol{\alpha}} \in \mathbb{R}^m_{\geq 0}$ and $r > 0$ with

$$\text{(negative gradient condition):} \qquad -\sum_{a=1}^{m} \hat{\alpha}_a \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H}) \geq r(\boldsymbol{I} - \boldsymbol{P}_G(\boldsymbol{H})).$$

This negative gradient condition implies that any state with a small ground state overlap must experience a substantially negative energy gradient, i.e., it must not be a local minimum.

To prove that $\boldsymbol{H}_C$ satisfies the negative gradient condition, we propose a series of lemmas and mathematical techniques for characterizing energy gradients in few-qubit systems, in commuting Hamiltonians, and in subspaces of the Hamiltonian, which are stated in Appendix 4.12 and proven in Appendix 4.16. These new techniques build on the operator Fourier transform, and the secular approximation given in [39]

[3] The binomial distribution ensures the Bohr-frequency gap is sufficiently large, which is central to the robustness of energy gradients under errors due to finite temperature and small perturbations. We believe that the standard circuit-to-Hamiltonian construction also has a large Bohr-frequency gap, but the proof seems more difficult.

for systematically handling energy uncertainty in thermal Lindbladians, which we review and adapt for our purpose in Appendix 4.15. Using these techniques, we analyze the energy gradient of the entire system perturbatively by considering a sequence of Hamiltonians

$$\boldsymbol{H}_1 \to \boldsymbol{H}_2 \to \boldsymbol{H}_3 = \boldsymbol{H}_C \quad \text{with refining ground spaces} \quad \boldsymbol{P}_1 \supset \boldsymbol{P}_2 \supset \boldsymbol{P}_3,$$
$$\text{where} \quad \|\boldsymbol{H}_1\| \gg \|\boldsymbol{H}_2 - \boldsymbol{H}_1\| \gg \|\boldsymbol{H}_3 - \boldsymbol{H}_2\|.$$

Through these perturbations, we sequentially rule out local minima in excited states of the Hamiltonian $\boldsymbol{H}_1, \boldsymbol{H}_2$ and, finally, $\boldsymbol{H}_3 = \boldsymbol{H}_C$. For example, we show the first Hamiltonian $\boldsymbol{H}_1$ satisfies the negative gradient condition and that the gradient is *stable* under perturbation going from $\boldsymbol{H}_1 \to \boldsymbol{H}_2 \to \boldsymbol{H}_3$. Controlling perturbations of the energy gradient is surprisingly challenging, and it is not *a priori* clear why this stability property should hold due to multiple (possibly competing) energy scales, including β^{-1}, τ^{-1}, the spectral gap, and the Bohr-frequency gap.[4] The perturbative errors are not suppressed by the spectral gap of the Hamiltonian as seen in standard settings, but instead by the Bohr-frequency gap, which can be much smaller (see Theorem 4.16.1). These techniques allow us to establish the robustness of energy gradients when perturbing a degenerate Hamiltonian with a sufficiently large Bohr-frequency gap.

We emphasize that while we proved that $\boldsymbol{H}_C$ has no suboptimal local minima when C is a polynomial-size quantum circuit, the same is not true for general local Hamiltonians. Finding the ground state of a local Hamiltonian is a QMA-hard problem; hence, we do not expect it to be solved efficiently by the quantum thermal gradient descent algorithm or by any other quantum algorithm. In the case of a quantum circuit that verifies the witness for a problem in QMA, Kitaev's corresponding local Hamiltonian contains a term, often denoted $\boldsymbol{H}_{in}$, which specifies some of the input qubits and leaves the input qubits corresponding to the witness unspecified, and a term, often denoted $\boldsymbol{H}_{out}$, which checks whether the witness is accepted. Due to the unspecified witness qubits in $\boldsymbol{H}_{in}$, the energy landscape contains a significant number of local minima corresponding to all possible witnesses. Furthermore, most of these local minima correspond to rejected witnesses and are suboptimal because of the energy penalty from $\boldsymbol{H}_{out}$. For these QMA-complete Hamiltonians, quantum thermal gradient descent is likely to remain stuck for a long time at a suboptimal local minimum. In $\boldsymbol{H}_C$, the term $\boldsymbol{H}_{in}$ specifies all input qubits,

[4] Recall that spectral gap is the minimum nonzero difference between energy eigenvalues. Bohr-frequency gap is the minimum nonzero difference between the difference of energy eigenvalues.

and the term $\boldsymbol{H}_{out}$ is absent, which greatly simplifies the energy landscape, enabling quantum thermal gradient descent to find the global minimum efficiently.

Local minima under local unitary perturbations

We present the central idea for proving Proposition 4.2.1 that establishes the classical easiness for finding local minima under local unitary perturbations.

To understand how easy the problem of finding local minima under local unitary perturbations is, we need to characterize the energy landscape. The following lemma provides a universal characterization of the structure of the energy landscape under local unitary perturbations. The formal statement is given in Lemma 4.7.1, and the proof is given in Appendix 4.7.

Lemma 4.4.1 (Barren plateau; informal)**.** *Given any n-qubit local Hamiltonian* $\boldsymbol{H}$*. A random pure n-qubit state* $|\psi\rangle$ *is an approximate local minimum of* $\boldsymbol{H}$ *under local unitary perturbations.*

The proof of the above lemma illustrates the following physical picture: the energy landscape in the pure state space defined in terms of local unitary perturbations consists of a large barren plateau [129] with doubly-exponentially many approximate local minima having exponentially small energy gradient. Additionally, almost all of the local minima have local properties that are exponentially close to that of the maximally mixed state. As a result, while finding ground states is classically hard, finding local minima under local unitary perturbations is classically trivial.

4.5 Appendix:Notations and Preliminaries

Before we begin stating and proving our results formally in the rest of the appendices, we present some notations used throughout the paper. We also give a brief review of key concepts in quantum information theory that we utilize in this work.

Notations

This section recapitulates notations, and the reader may skim through this and return as needed.

$$\boldsymbol{H} := \sum_i E_i |\psi_i\rangle\langle\psi_i| \qquad \text{Hamiltonian and the eigendecomposition}$$

$$\mathrm{Spec}(\boldsymbol{H}) := \{E_i\} \qquad \text{the spectrum of the Hamiltonian}$$

$$\nu \in B(\boldsymbol{H}) := \{E_i - E_j \mid E_i, E_j \in \mathrm{Spec}(\boldsymbol{H})\} \qquad \text{the set of Bohr frequencies}$$

$$\Delta_\nu(\boldsymbol{H}) := \min\{|\nu_1 - \nu_2| : \nu_1 \neq \nu_2 \in B(\boldsymbol{H})\} \qquad \text{the Bohr-frequency gap}$$

$$\boldsymbol{A}(t) := \mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} \qquad \text{Heisenberg evolution for operator } \boldsymbol{A}$$

$$m : \qquad \text{the number of jump operators}$$

$$\{\boldsymbol{A}^a\}_{a=1}^m : \qquad \text{the set of jump operators}$$

$$\boldsymbol{\rho} : \qquad \text{the density matrix}$$

$$\mathcal{L} : \qquad \text{a Lindbladian in the Schrodinger Picture}$$

$$\beta : \qquad \text{the inverse temperature}$$

$$\hat{\boldsymbol{A}}_f(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}\omega t}\boldsymbol{A}(t)\mathrm{d}t \qquad \text{Operator Fourier transform of } \boldsymbol{A} \text{ under } f$$

$$f_\tau(t) := \frac{1}{\sqrt{\tau}} \cdot \mathbb{1}(|t| \leq \tau/2) \qquad \text{the normalized window function with width } \tau$$

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega t} f(t)\mathrm{d}t \qquad \text{Fourier transform of a scalar function } f(t)$$

$$\boldsymbol{A}_\nu := \sum_{E_2 - E_1 = \nu} \boldsymbol{P}_{E_2}\boldsymbol{A}\boldsymbol{P}_{E_1} \qquad \text{operator } \boldsymbol{A} \text{ at exact Bohr frequency } \nu$$

$$\boldsymbol{I} : \qquad \text{the identity operator}$$

$$\|f\|_p := \left(\int_{-\infty}^{\infty} |f(t)|^p \mathrm{d}t\right)^{1/p} \qquad \text{the } p\text{-norm of a function}$$

$$\|\boldsymbol{O}\| := \sup_{|\psi\rangle,|\phi\rangle} \frac{\langle\phi|\boldsymbol{O}|\psi\rangle}{\||\psi\rangle\| \cdot \||\phi\rangle\|} \qquad \text{the operator norm of a matrix } \boldsymbol{O}$$

$$\|\boldsymbol{O}\|_p := (\mathrm{tr}|\boldsymbol{O}|^p)^{1/p} \qquad \text{the Schatten p-norm of a matrix } \boldsymbol{O}$$

$$\|\mathcal{L}\|_{p-p} := \sup_{\boldsymbol{O}} \frac{\|\mathcal{L}[\boldsymbol{O}]\|_p}{\|\boldsymbol{O}\|_p} \qquad \text{the induced } p-p \text{ norm of a superoperator } \mathcal{L}$$

We write scalars, functions and vectors in normal font, and natural constants $\mathrm{e}, \mathrm{i}, \pi$ are particularly in Roman font. We write matrices in bold font $\boldsymbol{O}$ and super-operators in curly font $\mathcal{L}$.

Furthermore, we define the indicator function $\mathbb{1}(S)$ which is 1 if the statement S is true and 0 otherwise. For any orthogonal projector $\boldsymbol{P}$, we denote $\boldsymbol{P}^\perp = \boldsymbol{I} - \boldsymbol{P}$. We

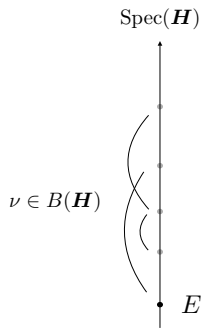


Figure 4.2: The Bohr frequencies $\nu \in B(\boldsymbol{H}) = \{E_i - E_j \mid E_i, E_j \in \mathrm{Spec}(\boldsymbol{H})\}$ are the differences of energy (eigenvalues of the Hamiltonian $\boldsymbol{H}$).

say $\boldsymbol{A} \overset{E}{\approx} \boldsymbol{B}$ when $\|\boldsymbol{A} - \boldsymbol{B}\| \leq E$.

To simplify the notation, we often drop f as a subscript $\hat{\boldsymbol{A}}_f(\omega) \equiv \hat{\boldsymbol{A}}(\omega)$, by which we have chosen the window function $f(t) = f_\tau(t)$.

Lindbladians

Completely Positive Trace-Preserving (CPTP) maps, also called quantum channels and quantum processes in the literature, correspond to all possible physical operations that could transform quantum states into other quantum states. *Lindbladian*s are infinitesimal generators of CPTP maps. That is, they map density operators to density operators (even if the map is tensored with the identity)

$$\mathcal{I} \otimes \mathrm{e}^{\mathcal{L}t}[\cdot] : \mathcal{S} \to \mathcal{S} \quad \text{for each} \quad t \geq 0.$$

In the Schrodinger Picture, a Lindbladian always has the following structure:

$$\mathcal{L}[\boldsymbol{\rho}] = \underbrace{-\mathrm{i}[\boldsymbol{H}, \boldsymbol{\rho}]}_{\text{coherent term}} + \sum_{j \in J} \Bigg(\underbrace{\boldsymbol{L}_j \boldsymbol{\rho} \boldsymbol{L}_j^\dagger}_{\text{transition rate}} - \underbrace{\frac{1}{2}\{\boldsymbol{L}_j^\dagger \boldsymbol{L}_j, \boldsymbol{\rho}\}}_{\text{decay rate}} \Bigg),$$

where the commutator is shorthanded by $[\boldsymbol{A}, \boldsymbol{B}] = \boldsymbol{A}\boldsymbol{B} - \boldsymbol{B}\boldsymbol{A}$ and the anti-commutator by $\{\boldsymbol{A}, \boldsymbol{B}\} = \boldsymbol{A}\boldsymbol{B} + \boldsymbol{B}\boldsymbol{A}$. The operator $\boldsymbol{H}$ can be any Hermitian matrix, and the set of *Lindblad operators* $\{\boldsymbol{L}_j\}_{j \in J}$ can be arbitrary as the second term always ensures trace-preserving.

Thermal Lindbladians

In this section, we describe the basic parameters that define a *thermal Lindbladian*, i.e., Lindbladian originating from generic system-bath interactions under a Markovian,

weak-coupling assumption [135]. Consider an n-qubit quantum system governed by a Hamiltonian $\boldsymbol{H}$ and a heat bath with inverse temperature β and time scale τ. The bath interacts with the system via a set of local interaction terms *acting on the system* $\{\boldsymbol{A}^1, \ldots, \boldsymbol{A}^m\} = \{\boldsymbol{A}^a\}_{a=1}^m$, where each operator $\boldsymbol{A}^a$ acts on a constant number of qubits. Each operator $\boldsymbol{A}^a$ can be arbitrary ($\boldsymbol{A}^a$ does not need to be Hermitian nor unitary), but the set should be closed under Hermitian conjugate,

$$\{\boldsymbol{A}^a\}_{a=1}^m = \{\boldsymbol{A}^{a\dagger}\}_{a=1}^m.$$

Each $\boldsymbol{A}^a$ is referred to as a *jump operator* and induces changes in energy (in the n-qubit system). For simplicity, we will enforce the following normalization for the interaction strengths,

$$\|\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|_\infty \leq 1 \quad \text{for each} \quad a = 1, \ldots, m. \tag{4.2}$$

For example, we may consider $m = 3n$ and $\boldsymbol{A}^1, \ldots, \boldsymbol{A}^m$ to be all single-qubit Pauli observables $\boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{Z}_i$ for $i = 1, \ldots, n$, which have an interaction strength $\|\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|_\infty = 1$.

The above parameters determine the thermal Lindbladian governing the equation of motion for the density operator, also referred to as the *coarse-grained master equation* [135]

$$\frac{\mathrm{d}\boldsymbol{\rho}}{\mathrm{d}t} = -\mathrm{i}[\boldsymbol{H}, \boldsymbol{\rho}] + \sum_{a=1}^m \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}(\boldsymbol{\rho}), \tag{4.3}$$

The term $-\mathrm{i}[\boldsymbol{H}, \boldsymbol{\rho}]$ corresponds to the Hamiltonian dynamics governed by the system Hamiltonian $\boldsymbol{H}$, the (closed system) Schrodinger's equation. The effects of system bath interaction are captured by a weighted average of the thermal Lindbladian $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$, defined by each local jump operator $\boldsymbol{A}^a$, the Hamiltonian $\boldsymbol{H}$, and parameters of the bath β, τ. The weighting is captured by the nonnegative vector $\boldsymbol{\alpha} \in \mathbb{R}_{\geq 0}^m$.

By varying the m-dimensional nonnegative vector $\boldsymbol{\alpha} \in \mathbb{R}_{\geq 0}^m$, the open system dynamics in Eq. (4.3) have the freedom to tune the interaction strengths for the jump operators. Each α_a corresponds to the interaction strength of a jump operator $\boldsymbol{A}^a$ and can be effectively absorbed into the set of jump operators by considering

$$\left\{\sqrt{\alpha_a}\boldsymbol{A}^a\right\}_a.$$

The interaction strength $\alpha_a \geq 0$ determines how much contribution each thermal Lindbladian $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ provides, and can be regarded as a *probabilistic* mixture. For

example, if α_2 is set to 0, one removes the jump operator $\boldsymbol{A}^2$ from the system-bath interaction. This flexibility lets us study a (convex) set of thermal perturbations due to system-bath interaction by considering all $\boldsymbol{\alpha} \in \mathbb{R}^m_{\geq 0}$. As the system is weakly coupled to the bath, $\boldsymbol{\alpha}$ is considered to be a vector with a small $\|\boldsymbol{\alpha}\|_1 = \sum_a \alpha_a$.

For each local interaction term $\boldsymbol{A}^a$, the corresponding thermal Lindbladian $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ is an open system evolution with Lindblad jump operators $\{\hat{\boldsymbol{A}}^a(\omega)\}_\omega$ for all possible energy differences $\omega \in (-\infty, \infty)$. Each Lindblad jump operator $\hat{\boldsymbol{A}}^a(\omega)$ is a restricted version of the system-bath interaction term $\boldsymbol{A}^a$ that only contains transitions between eigenstates of $\boldsymbol{H}$ whose associated eigenvalues, i.e., energies, differ by approximately ω. The inverse temperature β sets the transition weight $\gamma_\beta(\omega)$, which determines the probability of occurrence for each Lindblad operator $\hat{\boldsymbol{A}}^a(\omega)$. For $\beta > 0$, the transition weight $\gamma_\beta(\omega)$ favors cooling ($\omega < 0$) over heating ($\omega > 0$) transitions. The timescale τ sets the resolution ($1/\tau$) at which $\hat{\boldsymbol{A}}(\omega)$ identifies the energy differences between the eigenstates. The exact form of thermal Lindbladians is relatively complex, so we defer further discussion to Appendix 4.10 when needed for the full technical proof.

4.6 Appendix:Local minima in quantum systems

In this appendix, we will introduce local minima in classical optimization, extend the definition to quantum systems, and formalize the problem of finding a local minimum in quantum systems.

Local minima in classical optimization

In this subsection, we describe the definition of local minima in finite-dimensional Euclidean spaces, introduce a direct generalization to geometries with tangent spaces and exponential maps (such as circles and spheres), and discuss the concept of approximate local minima.

Local minima in Euclidean space

In classical optimization, one considers a real-valued function $h(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}$ over a domain $\mathcal{X} \subseteq \mathbb{R}^n$ consisting of n-dimensional vectors, and the goal is to find the global minimum of $h(\boldsymbol{x})$,

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathcal{X}} h(\boldsymbol{x}).$$

Finding the global minimum is already NP-hard even when $h(\boldsymbol{x})$ is a quadratic function [144]. Instead of finding a global minimum, one typically resorts to finding a local minimum $\boldsymbol{x}^\#$, which is the minimum in a neighborhood around $\boldsymbol{x}^\#$. The

definition of a local minimum $\boldsymbol{x}^{\#}$ is that there exists a distance $\delta > 0$, such that

$$h(\boldsymbol{x}^{\#} + \boldsymbol{\alpha}) \geq h(\boldsymbol{x}^{\#}), \quad \text{for all} \quad \|\boldsymbol{\alpha}\| \leq \delta \quad \text{and} \quad \boldsymbol{x}^{\#} + \boldsymbol{\alpha} \in \mathcal{X}. \tag{4.4}$$

Here the vector $\boldsymbol{\alpha}$ is of the same dimension as $\boldsymbol{x}^{\#}$. We will refer to the above as an *exact local minimum* because all points in the neighborhood have to be *at least* $h(\boldsymbol{x}^{\#})$. When there is an $\boldsymbol{\alpha}$ such that $h(\boldsymbol{x}^{\#} + \boldsymbol{\alpha})$ is only lower than $h(\boldsymbol{x}^{\#})$ by an extremely small value, $\boldsymbol{x}^{\#}$ is still not an exact local minimum. We will also define the approximate local minimum that relaxes this in Appendix 4.6.

Local minima in general geometrical spaces

The concept of a local minimum can be directly generalized to any geometry with tangent spaces and exponential maps, such as spheres, density matrices, unitaries, and more general Riemannian manifolds. Consider the tangent space $T_{\boldsymbol{x}}$ and the exponential map $\exp_{\boldsymbol{x}}$ of a point $\boldsymbol{x}$. In a physical picture, the tangent space $T_{\boldsymbol{x}}$ is the space consisting of all vectors $\boldsymbol{\alpha}$ that describe the direction $\hat{\boldsymbol{\alpha}}$ and magnitude $\|\boldsymbol{\alpha}\|$ for a particle moving at point $\boldsymbol{x}$ on a manifold, and the exponential map $\exp_{\boldsymbol{x}}$ is a function that takes in the vector $\boldsymbol{\alpha} \in T_{\boldsymbol{x}}$ encompassing the direction and magnitude and outputs the point after moving $\boldsymbol{x}$ in the direction $\hat{\boldsymbol{\alpha}}$ with a magnitude $\|\boldsymbol{\alpha}\|$.[5] To visualize these concepts, we give two warm-up examples in the following.

Euclidean space: In an m-dimensional Euclidean space, $\forall \boldsymbol{x} \in \mathcal{X} = \mathbb{R}^m$, the tangent space is

$$T_{\boldsymbol{x}} = \{\boldsymbol{\alpha} \in \mathbb{R}^m\}.$$

Given $\boldsymbol{\alpha} \in T_{\boldsymbol{x}}$, when we move $\boldsymbol{x}$ in the direction $\hat{\boldsymbol{\alpha}}$ with a magnitude $\|\boldsymbol{\alpha}\|$, we obtain

$$\exp_{\boldsymbol{x}}(\boldsymbol{\alpha}) = \boldsymbol{x} + \boldsymbol{\alpha}.$$

We can see that this matches our physical picture.

Particle moving counter-clockwise on a circle: As another warm-up, let us consider a unit circle $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^2 \,|\, \|\boldsymbol{x}\| = 1\}$ where a particle can only move counter-clockwise. In this example, the tangent space $T_{\boldsymbol{x}}$ of a unit vector $\boldsymbol{x} \in \mathbb{R}^2$ with $\|\boldsymbol{x}\| = 1$ is the set of one-dimensional rays,

$$T_{\boldsymbol{x}} = \{\alpha \in \mathbb{R} \mid \alpha \geq 0\}.$$

[5] Strictly speaking, to define the exponential map, we need to know how to "transport" the vector α along itself. Fortunately, this is natural for all cases we consider.

The condition $\alpha \geq 0$ comes from the constraint that the particle can only move counter-clockwise (unidirectional rather than bidirectional). When we move $\boldsymbol{x}$ according to $\alpha \in T_{\boldsymbol{x}}$, we obtain

$$\exp_{\boldsymbol{x}}(\alpha) = \exp\left(\begin{pmatrix} 0 & -\alpha \\ \alpha & 0 \end{pmatrix}\right)\boldsymbol{x} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}\boldsymbol{x}.$$

The larger α is, the bigger the rotation is.

Using the language of tangent spaces and exponential maps, an exact local minimum $\boldsymbol{x}^{\#} \in \mathcal{X}$ of a function h is equivalent to the statement that there exists $\delta > 0$, such that

$$h(\exp_{\boldsymbol{x}^{\#}}(\boldsymbol{\alpha})) \geq h(\boldsymbol{x}^{\#}), \quad \text{for all} \quad \boldsymbol{\alpha} \in T_{\boldsymbol{x}^{\#}},\ \|\boldsymbol{\alpha}\| \leq \delta. \tag{4.5}$$

For the case of optimizing over m-dimensional Euclidean space, the condition of Eq. (4.5) becomes the same as Eq. (4.4) noting $\exp_{\boldsymbol{x}^{\#}}(\boldsymbol{\alpha}) = \boldsymbol{x}^{\#} + \boldsymbol{\alpha}$. However, the condition can be quite different when the tangent space changes. For example, consider a 2-dimensional Euclidean space and the function $h(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$. In general, there is only one exact local minimum $\boldsymbol{x}^{\#} = 0$. However, if the particle can only move to the right, the tangent space becomes $T_{\boldsymbol{x}} = \{\boldsymbol{\alpha} \in \mathbb{R}^2 \mid \boldsymbol{\alpha}_1 \geq 0\}$ and every point $\boldsymbol{x}$ with $\boldsymbol{x}_1 \geq 0$ and $\boldsymbol{x}_2 = 0$ is an exact local minimum. Modifying the tangent space changes the definition of neighborhood. Hence, the set of local minima would be changed accordingly. We will consider the most suitable norm $\|\boldsymbol{\alpha}\|$ for each context.

Approximate local minima

While global minima are computationally hard to find, exact local minima are not much easier. If there is an $\boldsymbol{\alpha}$ such that $h(\exp_{\boldsymbol{x}^{\#}}(\boldsymbol{\alpha}))$ is lower than $h(\boldsymbol{x}^{\#})$ by an extremely small value, $\boldsymbol{x}^{\#}$ is not consider to be an exact local minimum. The requirement to resolve an extremely small value in exact local minima leads to the fact that finding an exact local minimum is still computationally hard [5]. Furthermore, exact local minima are very sensitive to small perturbations to the function h. Therefore, it is desirable to define approximate local minima to promote computational efficiency and robustness to small perturbations. We consider the following principle for defining ϵ-approximate local minima: *if a function h^* is very close to h, then an exact local minimum of h^* is an approximate local minimum of $\tilde{h}$.* The formal definition is given below.

Definition 4.6.1. *(ϵ-approximate local minima) Given a space $\mathcal{X}$ with tangent spaces T_x and exponential maps $\exp_x$ for all $x \in \mathcal{X}$, and a function h: $\mathcal{X} \to \mathbb{R}$. $\boldsymbol{x}^{\#}$ is an*

ϵ-approximate local minimum of h if $\boldsymbol{x}^\#$ is the exact local minimum of some function h^*, where $\Delta(\boldsymbol{x}) := h^*(\boldsymbol{x}) - h(\boldsymbol{x})$ satisfies

$$\begin{aligned} |\Delta(\boldsymbol{x})| \le \epsilon \quad &\textit{for each} \quad \boldsymbol{x} \in \mathcal{X}, && \textit{(ϵ-bounded)}, \\ \left|\Delta(\exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha}))\right| \le \epsilon\|\boldsymbol{\alpha}\| \quad &\textit{for each} \quad \boldsymbol{\alpha} \in T_{\boldsymbol{x}^\#}, && \textit{(ϵ-Lipschitz around $x^\#$)}. \end{aligned}$$

A $(\epsilon = 0)$-approximate local minimum of h is an exact local minimum of h.

Under this definition, $\boldsymbol{x}^\#$ is an approximate local minimum of h if there is an $\boldsymbol{x}$ in the neighborhood of $\boldsymbol{x}^\#$ such that $h(\boldsymbol{x})$ is lower than $h(\boldsymbol{x}^\#)$ by an extremely small value. We also give the following equivalent characterization based on looking at the local neighborhood.

Proposition 4.6.1. *(An equivalent characterization of ϵ-approximate local minima) $\boldsymbol{x}^\# \in \mathcal{X}$ is an ϵ-approximate local minimum of the function h if and only if there exists a distance $\delta > 0$,*

$$h(\exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha})) \ge h(\boldsymbol{x}^\#) - \epsilon\|\boldsymbol{\alpha}\| \quad \textit{for each} \quad \boldsymbol{\alpha} \in T_{\boldsymbol{x}^\#}, \|\boldsymbol{\alpha}\| \le \delta, \tag{4.6}$$

i.e., all the neighboring points can at most be $\epsilon\|\boldsymbol{\alpha}\|$ lower than the point $\boldsymbol{x}^\#$.

Proof. For the "only if" statement, we recall the definition of an exact local minimum that there exists $\delta > 0$, such that $h^*(\exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha})) - h^*(\boldsymbol{x}^\#) \ge 0$ for all $\boldsymbol{\alpha} \in T_{\boldsymbol{x}^\#}$ and $\|\boldsymbol{\alpha}\| \le \delta$. From the ϵ-Lipschitz condition around $x^\#$ for the function $\Delta(\boldsymbol{x})$, we have

$$\begin{aligned} 0 \le h^*(\exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha})) - h^*(\boldsymbol{x}^\#) &= h(\exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha})) - h(\boldsymbol{x}^\#) + \Delta(\exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha})) \\ &\le h(\exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha})) - h(\boldsymbol{x}^\#) - \epsilon\|\boldsymbol{\alpha}\|. \end{aligned}$$

This concludes the "only if" statement.

For the "if" statement, consider δ to be of at most 1 and let

$$\Delta(\boldsymbol{x}) := \begin{cases} h(\boldsymbol{x}^\#) - h(\boldsymbol{x}), & \text{if } \boldsymbol{x} = \exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha}) \text{ for some } \boldsymbol{\alpha} \in T_{\boldsymbol{x}^\#}, \|\boldsymbol{\alpha}\| \le \delta, \\ 0, & \text{otherwise.} \end{cases}$$

We have $\boldsymbol{x}^\#$ is an exact local minimum for $h^*(\boldsymbol{x}) := h(\boldsymbol{x}) + \Delta(\boldsymbol{x})$. Furthermore, because $h(\boldsymbol{x}^\#) - h(\exp_{\boldsymbol{x}^\#}(\boldsymbol{\alpha})) \le \epsilon\|\boldsymbol{\alpha}\| \le \epsilon$, both ϵ-bounded and ϵ-Lipschitz around $x^\#$ are satisfied by $\Delta(\boldsymbol{x})$. ∎

Defining local minima in quantum systems

To define local minima, we need to consider the domain $\mathcal{X}$ of elements $\boldsymbol{x} \in \mathcal{X}$, the optimization function $h(\boldsymbol{x})$, the tangent space $T_{\boldsymbol{x}}$ consisting of all possible directions and magnitudes to move an element $\boldsymbol{x}$, where $\boldsymbol{\alpha} \in T_{\boldsymbol{x}}$ encompass the direction $\hat{\boldsymbol{\alpha}}$ and the magnitude $\|\boldsymbol{\alpha}\|$, and the exponential map $\exp_{\boldsymbol{x}}(\boldsymbol{\alpha})$ that describes the resulting element after moving $\boldsymbol{x}$ under $\boldsymbol{\alpha}$.

In the following, we present two settings. The first setting in Appendix 4.6 considers general quantum states that can evolve under thermodynamic processes induced by interacting with a low-temperature heat bath. This setting defines local minima under thermal perturbations. The second setting in Appendix 4.6 considers pure quantum states that can move under any unitary generated by a set of local Hermitian operators (e.g., all two-qubit Pauli observables $\boldsymbol{P}_i \otimes \boldsymbol{Q}_j$, where $\boldsymbol{P}, \boldsymbol{Q} \in \{\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\}$). This setting defines local minima under local unitary perturbations.

Definition based on thermal perturbations

In quantum mechanics, the central optimization problem considers a function h defined by the Hamiltonian $\boldsymbol{H}$ of an n-qubit quantum system,

$$h(\boldsymbol{\rho}) = \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}),$$

which is the average energy of an n-qubit quantum state $\boldsymbol{\rho}$. The ground states $\boldsymbol{\rho}^{(g)}$ of $\boldsymbol{H}$ are the global minima of the optimization over $h(\boldsymbol{\rho}) = \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})$ in the quantum state space, i.e., the set of density operators (trace-one positive semidefinite matrices),

$$\mathcal{S}_{2^n} := \{\boldsymbol{\rho} \in \mathbb{C}^{2^n \times 2^n} \mid \boldsymbol{\rho}^\dagger = \boldsymbol{\rho},\ \boldsymbol{\rho} \geq 0,\ \mathrm{tr}(\boldsymbol{\rho}) = 1\}.$$

When the quantum system is placed in a heat bath with inverse temperature $\beta \in [0, \infty]$, time scale $\tau \in [0, \infty]$, and system-bath interactions based on m local jump operators[6] $\boldsymbol{A}^1, \ldots, \boldsymbol{A}^m$, the system dynamics are effectively described by the *thermal Lindbladians* $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$,

$$\frac{\mathrm{d}\boldsymbol{\rho}(t)}{\mathrm{d}t} = -\mathrm{i}[\boldsymbol{H}, \boldsymbol{\rho}] + \sum_{a=1}^{m} \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}], \tag{4.7}$$

where $\alpha_a \geq 0$ for each a. After time t, the initial quantum state $\boldsymbol{\rho}$ will evolve to

$$\boldsymbol{\rho}(t) = \exp\left(-\mathrm{i}t[\boldsymbol{H}, \cdot] + \sum_{a=1}^{m} t\alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho}).$$

[6] A local operator $\boldsymbol{A}^a$ acts on $O(1)$ qubits, but the set of qubits that $\boldsymbol{A}^a$ acts on may not be geometrically close.

Each term $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ is the thermal Lindbladian associated with a local jump operator $\boldsymbol{A}^a$ (recall that local operator $\boldsymbol{A}^a$ acts on a constant number of qubits). See Appendix 4.5 for a brief review of thermal Lindbladians, and Appendix 4.10 for the exact form of thermal Lindbladians.

The coefficient $\alpha_a \geq 0$ corresponds to the interaction strength of each jump operator $\boldsymbol{A}^a$. As $\alpha_a < 0$ is equivalent to reversing time, we cannot have $\alpha_a < 0$ since thermodynamic processes are irreversible in general. Different interaction strength vector $\boldsymbol{\alpha}$ corresponds to a different system-bath interaction, and the thermodynamics could be different. Because $\boldsymbol{\alpha}$ describes the probability of each jump occurring, the natural norm $\|\boldsymbol{\alpha}\|$ for the interaction strength vector $\boldsymbol{\alpha}$ is $\|\boldsymbol{\alpha}\|_1$. We denote $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_1$ as the unit vector.

The thermodynamics equation in Eq. (4.7) consists of a fast-rotating term $-\mathrm{i}[\boldsymbol{H}, \cdot]$ due to the system Hamiltonian $\boldsymbol{H}$ that keeps the energy $\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})$ invariant and the thermal perturbation term $\sum_a \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ due to the heat bath that cools the system. Because $-\mathrm{i}[\boldsymbol{H}, \cdot]$ keeps the energy constant, only the thermal perturbation term $\sum_a \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ is relevant for minimizing the energy $h(\boldsymbol{\rho}) = \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})$. For notational simplicity, we will only consider contributions from the thermal perturbations and absorb the t dependence in $t\alpha_a$ into α_a since $\boldsymbol{\alpha}$ is an arbitrary nonnegative vector. Together, the thermal perturbation on $\boldsymbol{\rho}$ due to a heat bath with inverse temperature β, time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ can be written as

$$\boldsymbol{\rho} \to \exp\left(\sum_{a=1}^{m} \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho})$$

for a nonnegative vector $\boldsymbol{\alpha} \in \mathbb{R}^m_{\geq 0}$ that combines the interaction strength vector and time t.

A dictionary between all the relevant functions and variables for optimizing $\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})$ in n-qubit quantum systems under a heat bath with inverse temperature β and time scale τ and optimizing $h(\boldsymbol{x})$ in an n-dimensional Euclidean space is given as follows.

$$\begin{aligned}
\mathcal{X} = \mathbb{R}^n \quad &\leftrightarrow \quad \mathcal{X} = \mathcal{S}_{2^n}, && \text{(domain)} \\
\boldsymbol{x} \in \mathbb{R}^n \quad &\leftrightarrow \quad \boldsymbol{\rho} \in \mathcal{S}_{2^n}, && \text{(an element)} \\
h(\boldsymbol{x}) \quad &\leftrightarrow \quad h(\boldsymbol{\rho}) = \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}), && \text{(optimization function)} \\
T_{\boldsymbol{x}} = \{\boldsymbol{\alpha} \in \mathbb{R}^n\} \quad &\leftrightarrow \quad \{\boldsymbol{\alpha} \in \mathbb{R}^m_{\geq 0}\}, && \text{(tangent space)} \\
\exp_{\boldsymbol{x}}(\boldsymbol{\alpha}) = \boldsymbol{x} + \boldsymbol{\alpha} \quad &\leftrightarrow \quad \exp\left(\sum_{a=1}^{m} \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho}) && \text{(exponential map).}
\end{aligned}$$

The formal definition of tangent spaces and exponential maps via Lindbladians is given below.

Definition 4.6.2 (Tangent spaces of quantum states in a heat bath)**.** *Consider an n-qubit quantum state $\boldsymbol{\rho}$, an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_{a=1}^m$, and parameters $\beta, \tau \geq 0$. The tangent space $T_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_{a=1}^m}$ under a heat bath with an inverse temperature β, a time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ is defined as*

$$T_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_{a=1}^m} := \left\{\boldsymbol{\alpha} \in \mathbb{R}_{\geq 0}^m\right\},$$

which is independent of $\beta, \tau, \boldsymbol{H}, \{\boldsymbol{A}^a\}_{a=1}^m$. The exponential map $\exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}$ is defined as

$$\exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}(\boldsymbol{\alpha}) := \exp\left(\sum_{a=1}^m \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho}). \tag{4.8}$$

With the definition of tangent spaces and exponential maps, we can define ϵ-approximate local minimum similar to the classical case in Eq. (4.6). We consider the natural choice of $\|\cdot\|_1$ for the nonnegative vector $\boldsymbol{\alpha}$ encompassing the probability of each jump. Our results remain qualitatively the same for other reasonable vector norms, such as Euclidean norm $\|\cdot\|_2$ or ℓ_p norm $\|\cdot\|_p$.

Definition 4.6.3 (Local minima under thermal perturbations)**.** *Given an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_{a=1}^m$, and parameters $\beta, \tau \geq 0$, an n-qubit state $\boldsymbol{\rho} \in \mathcal{S}_{2^n}$ is an ϵ-approximate local minimum of $\boldsymbol{H}$ under thermal perturbations with an inverse temperature β, a time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ if there is a $\delta > 0$ such that*

$$\operatorname{tr}\left(\boldsymbol{H} \exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}(\boldsymbol{\alpha})\right) \geq \operatorname{tr}(\boldsymbol{H}\boldsymbol{\rho}) - \epsilon\|\boldsymbol{\alpha}\|_1 \quad \textit{for each} \quad \boldsymbol{\alpha} \in \mathbb{R}_{\geq 0}^m, \|\boldsymbol{\alpha}\|_1 \leq \delta,$$

i.e., all the neighboring points can at most be $\epsilon\|\boldsymbol{\alpha}\|_1$ lower than the point $\boldsymbol{\rho}$.

A central concept we will be using for characterizing local minima under thermal perturbations is the energy gradient. The energy gradient at an n-qubit state $\boldsymbol{\rho}$ under thermal perturbation is determined by the following state-independent operator:

$$\text{(energy gradient operator):} \qquad \sum_{a=1}^m \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\hat{\boldsymbol{e}}_a,$$

where we denote $\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}$ to be the Hermitian conjugate of $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$. The energy gradient operator is a vector of Hermitian observables. The terminology stems from

the fact that evaluating the energy gradient operator on a state $\boldsymbol{\rho}$ gives the energy gradient at the state $\boldsymbol{\rho}$,

$$\mathrm{tr}\left(\boldsymbol{H}\exp\left(-\mathrm{i}\sum_{a=1}^{m}\alpha_a\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho})\right)=\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})+\boldsymbol{\alpha}\cdot\sum_{a=1}^{m}\mathrm{tr}\left(\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\boldsymbol{\rho}\right)\hat{\boldsymbol{e}}_a+O(\|\alpha\|^2).$$

In Appendix 4.8, we provide more discussions about the energy gradient.

Thermal perturbations depend on how the quantum system is interacting with the heat bath. Local minima defined above are local minima of the Hamiltonian $\boldsymbol{H}$ under thermal perturbations induced by all system-bath interactions generated by the jump operators $\{\boldsymbol{A}^a\}_a$.

Remark 4.6.0.1 (Thermodynamics at local minima)**.** *Given a specific system-bath interaction, inverse temperature β, and time scale τ, there could still be thermodynamics at a local minimum. For example, when β is not infinitely large, a local minimum could still move to other higher-energy states due to thermal fluctuations. Another example is when the local minimum is on a large and flat plateau, then the local minimum can still perform a random walk on the plateau.*

Definition based on local unitary perturbations

Inspired by variation quantum eigensolvers [35, 77], another natural definition for tangent spaces, exponential maps, and local minima considers pure states and local unitary perturbation. Given m local Hermitian operators $\boldsymbol{h}_1,\ldots,\boldsymbol{h}_m$ with $\|\boldsymbol{h}_a\|_\infty=1$. Here, local means that each operator $\boldsymbol{h}_a$ only acts on a constant number of qubits. We can consider all possible local unitary perturbations formed by performing time evolution under a Hamiltonian generated by the set $\{\boldsymbol{h}_a\}_a$ of local Hermitian operators,

$$\sum_{a=1}^{m}\alpha_a\boldsymbol{h}_a,$$

for any $\boldsymbol{\alpha}\in\mathbb{R}^m$. Since the time evolution under a Hamiltonian is always reversible, there is no additional requirement that $\boldsymbol{\alpha}$ must be in the nonnegative orthant. Similar to thermal perturbations, we will absorb the contribution of evolution time t into the arbitrary vector $\boldsymbol{\alpha}$. Consider the following dictionary between all the relevant functions and variables for optimizing $\langle\psi|\boldsymbol{H}|\psi\rangle$ in n-qubit pure state $|\psi\rangle$ under local

unitary perturbation and optimizing $h(\boldsymbol{x})$ in an n-dimensional Euclidean space.

$$\begin{aligned}
\mathcal{X} = \mathbb{R}^n \quad &\leftrightarrow \quad \mathcal{X} = \{|\psi\rangle \in \mathbb{C}^{2^n} | \langle\psi|\psi\rangle = 1\}, && \text{(domain)} \\
\boldsymbol{x} \in \mathbb{R}^n \quad &\leftrightarrow \quad |\psi\rangle \in \mathbb{C}^{2^n}, \langle\psi|\psi\rangle = 1, && \text{(an element)} \\
h(\boldsymbol{x}) \quad &\leftrightarrow \quad h(|\psi\rangle) = \langle\psi|\boldsymbol{H}|\psi\rangle, && \text{(optimization function)} \\
T_{\boldsymbol{x}} = \{\boldsymbol{\alpha} \in \mathbb{R}^n\} \quad &\leftrightarrow \quad \{\boldsymbol{\alpha} \in \mathbb{R}^m\}, && \text{(tangent space)} \\
\exp_{\boldsymbol{x}}(\boldsymbol{\alpha}) = \boldsymbol{x} + \boldsymbol{\alpha} \quad &\leftrightarrow \quad \exp\left(\sum_{a=1}^{m} \alpha_a \boldsymbol{h}_a\right)|\psi\rangle && \text{(exponential map).}
\end{aligned}$$

The tangent space and the exponential map can be formally defined as follows.

Definition 4.6.4 (Tangent spaces of pure quantum states under local unitaries)**.** *Given an n-qubit pure quantum state $|\psi\rangle$ and m local Hermitian operators $\{\boldsymbol{h}_a\}_a$. The tangent space T_ψ is defined as*

$$T_\psi^{\{\boldsymbol{h}_a\}_a} := \mathbb{R}^m,$$

and the exponential map $\exp_\psi$ is defined as

$$\exp_\psi^{\{\boldsymbol{h}_a\}_a}(\boldsymbol{\alpha}) := \exp\left(-i \sum_a \alpha_a \boldsymbol{h}_a\right)|\psi\rangle.$$

When the set $\{\boldsymbol{h}_a\}_a$ is the set of all two-qubit Pauli observables, the tangent space T_ψ and exponential map $\exp_\psi$ define a Riemannian manifold that connects all n-qubit pure states through unitary evolutions. This Riemannian manifold is the state version of the manifold over quantum unitaries defined in a seminal work on the geometry of quantum computation [140].

The optimization function is $h(|\psi\rangle) = \langle\psi|\boldsymbol{H}|\psi\rangle$, the average energy of the Hamiltonian $\boldsymbol{H}$ for the pure state $|\psi\rangle$. Performing gradient descent on this pure state Riemannian manifold to minimize $\langle\psi|\boldsymbol{H}|\psi\rangle$ is equivalent to performing adaptive variational quantum optimization [77] to minimize the Hamiltonian $\boldsymbol{H}$. The local minima can be defined similarly as before. To be consistent with local minima under thermal perturbations, we consider the ℓ_1-norm $\|\boldsymbol{\alpha}\|_1$. All of our results remain qualitatively the same for other reasonable vector norms, such as the Euclidean norm or ℓ_p norm.

Definition 4.6.5 (local minima under local unitary perturbations)**.** *Given an n-qubit Hamiltonian $\boldsymbol{H}$, and m local Hermitian operators $\{\boldsymbol{h}_a\}_a$. A pure state $|\psi\rangle$ is an*

ϵ-approximate local minimum of $\boldsymbol{H}$ under local unitary perturbations generated by $\{\boldsymbol{h}_a\}_a$ if

$$\exp_{\psi}^{\{\boldsymbol{h}_a\}_a}(\boldsymbol{\alpha})^{\dagger}\boldsymbol{H}\exp_{\psi}^{\{\boldsymbol{h}_a\}_a}(\boldsymbol{\alpha}) \geq \langle\psi|\boldsymbol{H}|\psi\rangle - \epsilon\|\boldsymbol{\alpha}\|_1, \text{ for each } \quad \boldsymbol{\alpha} \in T_{\psi}^{\{\boldsymbol{h}_a\}_a}, \|\boldsymbol{\alpha}\|_1 \leq \delta.$$

for some $\delta > 0$.

This is also a valid definition of local minima in quantum systems. However, we will later show that the optimization landscape defined in this way always has a very large barren plateau. Hence, the problem of finding a local minimum defined in this way will be a trivial problem.

The problem of finding a local minimum in quantum systems

With these definitions of local minima, we can define the task of finding a local minimum in a straightforward manner. To formulate the problem to have purely classical output, we focus on outputting a simple property, such as the expectation value of a local observable $\boldsymbol{O}$, of an approximate local minimum $\boldsymbol{\rho}$. Furthermore, we only consider Hamiltonians $\boldsymbol{H}$ that can be written as a sum of local observables, commonly referred to as local Hamiltonians in the literature.

While there can be many approximate local minima, we consider the algorithm to be successful if the algorithm outputs the property of any one of the local minima.

Definition 4.6.6 (Finding a local minimum under low-temperature thermal perturbations)**.** *Given error $\epsilon > 0$, inverse temperature $\beta \geq 0$, time scale $\tau \geq 0$, an n-qubit local Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_a$, and a local observable $\boldsymbol{O}$ with $\|\boldsymbol{O}\|_\infty \leq 1$. Output a real value $v \in [-1, 1]$, such that v is ϵ-close to $\mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho})$ for an ϵ-approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}$ under thermal perturbations with an inverse temperature β, a time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$.*

Definition 4.6.7 (Finding a local minimum under local unitary perturbations)**.** *Given error $\epsilon > 0$, an n-qubit local Hamiltonian $\boldsymbol{H}$, m local Hermitian operators $\{\boldsymbol{h}_a\}_a$, and a local observable $\boldsymbol{O}$ with $\|\boldsymbol{O}\|_\infty \leq 1$. Output a real value $v \in [-1, 1]$, such that v is ϵ-close to $\langle\psi|\boldsymbol{O}|\psi\rangle$ for an ϵ-approximate local minimum $|\psi\rangle$ of the Hamiltonian $\boldsymbol{H}$ under local unitary perturbations generated by $\{\boldsymbol{h}_a\}_a$.*

Ideally, we would like the two problems to be quantumly easy and classically hard. However, we will show that only the first problem based on thermal perturbations is both quantumly easy and classically hard. The second problem based on local unitary

perturbation is classically trivial due to the presence of too many local minima in an exponentially large barren plateau.

The importance of irreversible perturbations

Suppose that the perturbations $\mathcal{P}_{\boldsymbol{\alpha}}$ parameterized by a polynomial-size vector $\boldsymbol{\alpha}$ are reversible $\mathcal{P}_{-\boldsymbol{\alpha}} = \mathcal{P}_{\boldsymbol{\alpha}}^{-1}$ and are smooth. The following argument shows that the energy landscape must have doubly-exponentially many approximate local minima.

Given any n-qubit state $\boldsymbol{\rho}$ and any n-qubit Hamiltonian $\boldsymbol{H}$ with $\|\boldsymbol{H}\|_\infty = \text{poly}(n)$. Consider a gradient descent algorithm that starts at $\boldsymbol{\rho}$. Because $\|\boldsymbol{H}\|_\infty = \text{poly}(n)$, after a polynomial number of steps T, the gradient descent algorithm can find an approximate local minimum $\boldsymbol{\rho}^{\#}$ of $\boldsymbol{H}$,

$$\boldsymbol{\rho}^{\#} = \mathcal{P}_{\boldsymbol{\alpha}_T} \ldots \mathcal{P}_{\boldsymbol{\alpha}_1}(\boldsymbol{\rho}).$$

From the reversibility of the perturbations, we have

$$\boldsymbol{\rho} = \mathcal{P}_{-\boldsymbol{\alpha}_1} \ldots \mathcal{P}_{-\boldsymbol{\alpha}_T}(\boldsymbol{\rho}^{\#}).$$

Consider a covering net $\mathcal{N}$ for the set of approximate local minima of $\boldsymbol{H}$. Because the packing net for all n-qubit states is of size

$$\exp(\exp(\Omega(n))),$$

and the covering net for the perturbations is of size

$$\exp(\text{poly}(n)).$$

We have the following relationship,

$$\exp(T \cdot \text{poly}(n))|\mathcal{N}| = \exp(\exp(\Omega(n))).$$

Hence, we can see that

$$|\mathcal{N}| = \exp(\exp(\Omega(n)) - \text{poly}(n)) = \exp(\exp(\Omega(n)))$$

since $\exp(\Omega(n))$ grows much faster than $\text{poly}(n)$.

4.7 Appendix:Characterizing local minima under local unitary perturbations

Now that we have defined local minima in quantum systems, we present a set of results characterizing properties of local minima in quantum systems in this and the next appendix. These results provide a further understanding of local minima

in quantum systems and are essential to establishing the main theorems given in Appendix 4.9.

We begin by looking at the energy landscape defined by local unitary perturbations. We will prove a central lemma portraying the energy landscapes defined by local unitary perturbations for pure quantum systems. The lemma states that most pure quantum states $|\psi\rangle$ are local minima under local unitary perturbations with an expectation value close to $\mathrm{tr}(\boldsymbol{O})/2^n = \mathrm{tr}(\boldsymbol{O}(\boldsymbol{I}/2^n))$ for any local observable $\boldsymbol{O}$. Furthermore, the proof shows that the gradient at a randomly sampled local minimum $|\psi\rangle$ is exponentially close to zero.

Lemma 4.7.1 and its proof provide the following physical picture. In the energy landscape defined by local unitary perturbations, there is an overwhelmingly large barren plateau consisting of local minima with almost equal energy as their neighbors. Furthermore, these local minima behave like the maximally mixed state $\boldsymbol{I}/2^n$, which makes the task of predicting properties for a local minimum under local unitary perturbations classically trivial to solve.

Given m local Hermitian operators $\{\boldsymbol{h}_a\}_a$ and $\boldsymbol{\alpha} \in \mathbb{R}^m$. By applying Taylor's theorem in Prop. 4.8.1 to the one-dimensional function

$$g(t) = \exp_{\psi}^{\{\boldsymbol{h}_a\}_a}(t\hat{\boldsymbol{\alpha}})^{\dagger} \boldsymbol{H} \exp_{\psi}^{\{\boldsymbol{h}_a\}_a}(t\hat{\boldsymbol{\alpha}})$$

for $0 \leq t \leq \|\boldsymbol{\alpha}\|_1$ and $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_1$, we can obtain the following proposition.

Proposition 4.7.1 (Taylor's theorem for local unitary perturbations)**.** *Given an n-qubit Hamiltonian $\boldsymbol{H}$, $\boldsymbol{\alpha} \in \mathbb{R}^n$, m local Hermitian operators $\{\boldsymbol{h}_a\}_a$, and an n-qubit pure state $|\psi\rangle$. We have*

$$\exp_{\psi}(\boldsymbol{\alpha})^{\dagger} \boldsymbol{H} \exp_{\psi}(\boldsymbol{\alpha}) = \langle\psi|\boldsymbol{H}|\psi\rangle - \mathrm{i}\langle\psi|\left[\boldsymbol{H}, \sum_{a=1}^{m} \alpha_a \boldsymbol{h}_a\right]|\psi\rangle \\ - \frac{1}{2}\sum_{a=1}^{m}\sum_{a'=1}^{m} \alpha_a \alpha_{a'} \exp_{\psi}(\eta\hat{\boldsymbol{\alpha}})^{\dagger}[[\boldsymbol{H}, \boldsymbol{h}_a], \boldsymbol{h}_{a'}] \exp_{\psi}(\eta\hat{\boldsymbol{\alpha}}),$$

for some $0 \leq \eta \leq \|\boldsymbol{\alpha}\|_1$.

Lemma 4.7.1 (A random state is a local minimum under local unitary perturbations; Restatement of Lemma 4.4.1)**.** *Consider a large problem size n. Given error $\epsilon \geq 1/2^{n/4}$, an n-qubit local Hamiltonian $\boldsymbol{H}$ with $\|\boldsymbol{H}\|_{\infty} = \mathrm{poly}(n)$, m local Hermitian operators $\{\boldsymbol{h}_a\}_a$ with $m = \mathrm{poly}(n)$ and $\|\boldsymbol{h}_a\|_{\infty} = 1$, and a local observable*

$\boldsymbol{O}$ with $\|\boldsymbol{O}\|_\infty \leq 1$. With probability at least $1 - 1/2^{2^{n/4}}$, an n-qubit state $|\psi\rangle$ sampled uniformly at random is an ϵ-approximate local minimum of $\boldsymbol{H}$ under local unitary perturbations generated by $\{\boldsymbol{h}_a\}_a$ and $\langle\psi|\boldsymbol{O}|\psi\rangle$ is ϵ-close to $\mathrm{tr}(\boldsymbol{O})/2^n$.

Proof. From Lemma III.5 in [87], for any Pauli operator $\boldsymbol{Q} \in \{\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\}^{\otimes n} \setminus \{\boldsymbol{I}^{\otimes n}\}$ and a random n-qubit pure state $|\psi\rangle$ sampled uniformly, we have

$$\Pr_{|\psi\rangle}[|\langle\psi|\boldsymbol{Q}|\psi\rangle| > \delta] \leq 2\exp\left(-\frac{2^n\delta^2}{10}\right),$$

for any $0 \leq \delta \leq 1$. Let $\delta = 1/2^{n/3}$. Then, we have

$$\Pr_{|\psi\rangle}\left[|\langle\psi|\boldsymbol{Q}|\psi\rangle| > \frac{1}{2^{n/3}}\right] \leq 2\exp\left(-\frac{2^{n/3}}{10}\right).$$

Recall that any Hermitian operator has a unique Pauli decomposition:

$$\begin{aligned}
\boldsymbol{H} &= \sum_{\boldsymbol{P}\in\{\boldsymbol{I},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}^{\otimes n}} \alpha_{\boldsymbol{P}}(\boldsymbol{H})\boldsymbol{P}, \\
\boldsymbol{O} &= \sum_{\boldsymbol{P}\in\{\boldsymbol{I},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}^{\otimes n}} \alpha_{\boldsymbol{P}}(\boldsymbol{O})\boldsymbol{P}, \\
\boldsymbol{h}^a &= \sum_{\boldsymbol{P}\in\{\boldsymbol{I},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}^{\otimes n}} \alpha_{\boldsymbol{P}}(\boldsymbol{h}^a)\boldsymbol{P},
\end{aligned}$$

where the Pauli coefficients $\alpha_{\boldsymbol{P}}(\cdot)$ satisfy

$$\sum_{\boldsymbol{P}\in\{\boldsymbol{I},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}^{\otimes n}} \alpha^2_{\boldsymbol{P}}(\boldsymbol{H}) \leq \|\boldsymbol{H}\|^2_\infty = \mathrm{poly}(n), \tag{4.9}$$

$$\sum_{\boldsymbol{P}\in\{\boldsymbol{I},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}^{\otimes n}} \alpha^2_{\boldsymbol{P}}(\boldsymbol{O}) \leq \|\boldsymbol{O}\|^2_\infty = 1. \tag{4.10}$$

$$\sum_{\boldsymbol{P}\in\{\boldsymbol{I},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}^{\otimes n}} \alpha^2_{\boldsymbol{P}}(\boldsymbol{h}_a) \leq \|\boldsymbol{h}_a\|^2_\infty = 1. \tag{4.11}$$

Let S_0 be the set of Pauli operator $\boldsymbol{P}$ with nonzero Pauli coefficients $\alpha_{\boldsymbol{P}}$ in the Pauli decompositions of either $\boldsymbol{H}$ or $\boldsymbol{O}$,

$$S_0 = \left\{\boldsymbol{P} \in \{\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\}^{\otimes n} \setminus \{\boldsymbol{I}^{\otimes n}\} \mid \alpha_{\boldsymbol{P}}(\boldsymbol{H}) \neq 0 \text{ or } \alpha_{\boldsymbol{P}}(\boldsymbol{O}) \neq 0\right\},$$

and S_E be the set of Pauli operator $\boldsymbol{P}$ with nonzero Pauli coefficients $\alpha_{\boldsymbol{P}}$ in the Pauli decompositions of $\boldsymbol{h}_a$ for some a,

$$S_E = \left\{\boldsymbol{P} \in \{\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\}^{\otimes n} \setminus \{\boldsymbol{I}^{\otimes n}\} \mid \exists 1 \leq a \leq m, \alpha_{\boldsymbol{P}}(\boldsymbol{h}_a) \neq 0\right\}.$$

Because $\boldsymbol{H}$ is a local Hamiltonian and $\boldsymbol{O}$ is a local observable, we have $|S_0| = \text{poly}(n)$. Because $\boldsymbol{h}_a$ is a local observable, we have $|S_E| = O(m) = \text{poly}(n)$. We then define

$$S = \left\{ \boldsymbol{P}' \in \{\boldsymbol{I}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}\}^{\otimes n} \setminus \{\boldsymbol{I}^{\otimes n}\} \mid \exists \boldsymbol{Q} \in S_0, \boldsymbol{P} \in S_E, \text{tr}(\boldsymbol{P}'[\boldsymbol{Q}, \boldsymbol{P}]) \neq 0 \right\} \cup S_0.$$

Because $|S_0| = \text{poly}(n)$ and $[\boldsymbol{Q}, \boldsymbol{P}]$ is another Pauli observable up to a phase, we have $|S| \leq |S_E||S_1| = \text{poly}(n)$. The union bound yields the following probabilistic statement:

$$1 - \Pr_{|\psi\rangle}\left[|\langle\psi|\boldsymbol{Q}|\psi\rangle| < \frac{1}{2^{n/3}}, \ \forall \boldsymbol{Q} \in S \right] \leq 2|S| \exp\left(-\frac{2^{n/3}}{10}\right) \leq \frac{\text{poly}(n)}{2^{2^{n/3}/10}} < \frac{1}{2^{2^{n/4}}},$$

where the last inequality holds for any large n since $2^{2^{n/3}/10 - 2^{n/4}}$ grows much faster than any polynomial of n. We condition on the event for the random state $|\psi\rangle$ that

$$|\langle\psi|\boldsymbol{Q}|\psi\rangle| < \frac{1}{2^{n/3}} \quad \text{for all} \quad \boldsymbol{Q} \in S,$$

referred to as event E^*. We can obtain the following from Cauchy-Schwarz inequality:

$$\begin{aligned} |\langle\psi|[\boldsymbol{H}, \boldsymbol{h}_a]|\psi\rangle| &\leq \sum_{\boldsymbol{Q},\boldsymbol{P} \in \{\boldsymbol{I},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}^{\otimes n}} \left|\alpha_{\boldsymbol{Q}}(\boldsymbol{H})\right| |\alpha_{\boldsymbol{P}}(\boldsymbol{h}_a)| |\langle\psi|[\boldsymbol{Q}, \boldsymbol{P}]|\psi\rangle| \\ &\leq \sqrt{\sum_{\boldsymbol{Q},\boldsymbol{P} \in \{\boldsymbol{I},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}^{\otimes n}} \alpha_{\boldsymbol{Q}}^2(\boldsymbol{H}) \alpha_{\boldsymbol{P}}^2(\boldsymbol{h}_a)} \sqrt{\frac{|S|}{2^{2n/3}}} \leq \frac{\text{poly}(n)}{2^{n/3}}, \end{aligned} \tag{4.12}$$

where the second inequality uses the conditioning on event E^* and $[\boldsymbol{Q}, \boldsymbol{P}] \neq 0 \implies \boldsymbol{Q}, \boldsymbol{P} \neq \boldsymbol{I}^{\otimes n}$, and the third inequality uses $|S| = \text{poly}(n)$ and Eq. (4.9), (4.11). Similarly, we also have

$$|\langle\psi|[[\boldsymbol{H}, \boldsymbol{h}_a], \boldsymbol{h}_{a'}]|\psi\rangle| \leq \frac{\text{poly}(n)}{2^{n/3}}.$$

Using Eq. (4.10) instead of Eq. (4.9), we can similarly obtain

$$|\langle\psi|\boldsymbol{O}|\psi\rangle - \alpha_{I^{\otimes n}}(\boldsymbol{O})| = \left|\langle\psi|\boldsymbol{O}|\psi\rangle - \frac{\text{tr}(O)}{2^n}\right| \leq \frac{\text{poly}(n)}{2^{n/3}} < \frac{1}{2^{n/4}} \leq \epsilon \tag{4.13}$$

for any large problem size n since $2^{n/3}$ grows much faster than any polynomial in n. We now show that $|\psi\rangle$ is an ϵ-approximate local minimum of $\boldsymbol{H}$ under local unitary perturbations. To establish this claim, from Def. 4.6.5, we need to prove that

$$\exp_\psi(\boldsymbol{\alpha})^\dagger \boldsymbol{H} \exp_\psi(\boldsymbol{\alpha}) \geq \langle\psi|\boldsymbol{H}|\psi\rangle - \epsilon\|\boldsymbol{\alpha}\|_1, \ \text{for each} \quad \boldsymbol{\alpha} \in T_\psi, \|\boldsymbol{\alpha}\|_1 \leq \delta$$

for some $\delta > 0$. Recall from Lemma 4.7.1 based on Taylor's theorem (Prop. 4.8.1), we have

$$\exp_\psi(\boldsymbol{\alpha})^\dagger \boldsymbol{H} \exp_\psi(\boldsymbol{\alpha}) = \langle\psi|\boldsymbol{H}|\psi\rangle - \mathrm{i}\langle\psi|\left[\boldsymbol{H}, \sum_{a=1}^m \alpha_a \boldsymbol{h}_a\right]|\psi\rangle \\ - \frac{1}{2}\sum_{a=1}^m\sum_{a'=1}^m \alpha_a\alpha_{a'} \exp_\psi(\eta\hat{\boldsymbol{\alpha}})^\dagger [[\boldsymbol{H}, \boldsymbol{h}_a], \boldsymbol{h}_{a'}] \exp_\psi(\eta\hat{\boldsymbol{\alpha}}),$$

for some $0 \le \eta \le \|\boldsymbol{\alpha}\|_1$. For the linear term, from Eq. (4.12) bounding $|\langle\psi|[\boldsymbol{H}, \boldsymbol{h}_a]|\psi\rangle|$, we have

$$\left|-\mathrm{i}\langle\psi|\left[\boldsymbol{H}, \sum_{a=1}^m \alpha_a \boldsymbol{h}_a\right]|\psi\rangle\right| \le \sum_{a=1}^m |\alpha_a| \frac{\mathrm{poly}(n)}{2^{n/3}} \le \frac{\mathrm{poly}(n)}{2^{n/3}}\|\boldsymbol{\alpha}\|_1,$$

where the last inequality uses $m = \mathrm{poly}(n)$. For the quadratic residual term, we have

$$\left|\frac{1}{2}\sum_{a=1}^m\sum_{a'=1}^m \alpha_a\alpha_{a'} \exp_\psi(\eta\hat{\boldsymbol{\alpha}})^\dagger [[\boldsymbol{H}, \boldsymbol{h}_a], \boldsymbol{h}_{a'}] \exp_\psi(\eta\hat{\boldsymbol{\alpha}})\right| \le 2\|\boldsymbol{\alpha}\|_1^2\|\boldsymbol{H}\|_\infty = \mathrm{poly}(n)\|\boldsymbol{\alpha}\|_1^2.$$

Together, we can combine the inequalities to get

$$\exp_\psi(\boldsymbol{\alpha})^\dagger \boldsymbol{H} \exp_\psi(\boldsymbol{\alpha}) \ge \langle\psi|\boldsymbol{H}|\psi\rangle - \frac{\mathrm{poly}(n)}{2^{n/3}}\|\boldsymbol{\alpha}\|_1 - \mathrm{poly}(n)\|\boldsymbol{\alpha}\|_1^2 \\ \ge \langle\psi|\boldsymbol{H}|\psi\rangle - 0.5\epsilon\|\boldsymbol{\alpha}\|_1 - \mathrm{poly}(n)\|\boldsymbol{\alpha}\|_1^2,$$

where the second inequality holds for any large problem size n since $\epsilon \ge 1/2^{n/4}$ decays much slower than $\mathrm{poly}(n)/2^{n/3}$. For any $\|\boldsymbol{\alpha}\| < \delta := 0.5\epsilon/\mathrm{poly}(n)$, we have

$$\exp_\psi(\boldsymbol{\alpha})^\dagger \boldsymbol{H} \exp_\psi(\boldsymbol{\alpha}) \ge \langle\psi|\boldsymbol{H}|\psi\rangle - 0.5\epsilon\|\boldsymbol{\alpha}\|_1 - 0.5\epsilon\|\boldsymbol{\alpha}\|_1 = \langle\psi|\boldsymbol{H}|\psi\rangle - \epsilon\|\boldsymbol{\alpha}\|_1,$$

which shows that $|\psi\rangle$ is an ϵ-approximate local minimum of $\boldsymbol{H}$ under local unitary perturbations. Finally, because the event E^* occurs with probability at least 0.99, by combining Eq. (4.13) and the above, we establish the claim that, with high probability, a random n-qubit state $|\psi\rangle$ sampled uniformly is an ϵ-approximate local minimum of $\boldsymbol{H}$ under local unitary perturbations and $\langle\psi|\boldsymbol{O}|\psi\rangle$ is ϵ-close to $\mathrm{tr}(O)/2^n$. ■

4.8 Appendix:Characterizing local minima under thermal perturbations

In this appendix, we characterize local minima under thermal perturbations. In particular, we will focus on the gradients of the energy landscape, conditions of local minima, and conditions on the Hamiltonian $\boldsymbol{H}$ that ensure approximate local minima are approximate global minima, i.e., there are no suboptimal local minima in the energy landscape.

Energy gradients

The energy landscape is much more nontrivial when defined under thermal perturbations. We can study the energy landscape by looking at the energy gradients. Recall the exponential map $\exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}$ in Eq. (4.8) and consider the one-dimensional function

$$g(t) = \mathrm{tr}\left(\boldsymbol{H} \exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}(t\hat{\boldsymbol{\alpha}})\right) = \mathrm{tr}\left(\boldsymbol{H} \exp\left(\sum_{a=1}^{m} t\hat{\alpha}_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho})\right)$$

for $0 \leq t \leq \|\boldsymbol{\alpha}\|_1, \hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_1$. We have the following derivatives:

$$\frac{dg}{dt}(t) = \mathrm{tr}\left(\boldsymbol{H} \sum_a \hat{\alpha}_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\left[\exp\left(\sum_{a=1}^{m} t\hat{\alpha}_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho})\right]\right),$$

$$\frac{d^2g}{dt^2}(t) = \mathrm{tr}\left(\boldsymbol{H} \sum_a \sum_{a'} \hat{\alpha}_a \hat{\alpha}_{a'} \mathcal{L}_{a'}^{\beta,\tau,\boldsymbol{H}}\left[\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\left[\exp\left(\sum_{a=1}^{m} t\hat{\alpha}_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right)(\boldsymbol{\rho})\right]\right]\right).$$

Recall Taylor's theorem with the Lagrange form of the remainder from standard single-variate calculus. By applying Taylor's theorem in Prop. 4.8.1 to $g(t)$, we can obtain Prop. 4.8.2.

Proposition 4.8.1 (Taylor's theorem)**.** *Let $g(t) : \mathbb{R} \to \mathbb{R}$ be twice differentiable on the open interval between 0 and t and $g'(t)$ continuous on the closed interval between 0 and t. Then*

$$g(t) = g(0) + g'(0)t + \frac{1}{2}g''(\eta)t^2,$$

for some real number η between 0 and t.

Proposition 4.8.2 (Taylor's theorem for thermal perturbations)**.** *Given an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_a$, parameters $\beta, \tau \geq 0$, $\boldsymbol{\alpha} \in \mathbb{R}^m_{\geq 0}$, and an n-qubit state $\boldsymbol{\rho} \in \mathcal{S}_{2^n}$.*

$$\mathrm{tr}\left(\boldsymbol{H} \exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}(\boldsymbol{\alpha})\right) = \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}) + \sum_a \alpha_a \,\mathrm{tr}(\boldsymbol{H}\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}])$$
$$+ \frac{1}{2}\sum_a \sum_{a'} \alpha_a \alpha_{a'} \,\mathrm{tr}\left(\boldsymbol{H}\mathcal{L}_{a'}^{\beta,\tau,\boldsymbol{H}}[\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}(\eta\hat{\boldsymbol{\alpha}})]]\right)$$

for some $0 \leq \eta \leq \|\boldsymbol{\alpha}\|_1$.

We define the energy gradients as follows. We separately consider a positive and a negative energy gradient. The motivation of the definition is that the positive

(negative) energy gradient should determine the direction of the thermodynamics that causes the energy of the state to increase (decrease). Because our goal is to understand local minima, we will focus on the negative energy gradient. When one studies local maxima, one will focus on the positive energy gradient.

Definition 4.8.1 (Energy gradients of a state under thermal perturbations)**.** *Given an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_a$, and parameters $\beta, \tau \geq 0$, the gradients of an n-qubit state $\boldsymbol{\rho} \in \mathcal{S}_{2^n}$ under thermal perturbations with inverse temperature β, time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ is defined as*

$$\boldsymbol{\nabla}^{+}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho}) := \sum_{a=1}^{m} \max\Big(+\operatorname{tr}\Big(\boldsymbol{H}\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]\Big), 0\Big)\hat{\boldsymbol{e}}_a, \quad \textit{(positive energy gradient)}$$

$$\boldsymbol{\nabla}^{-}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho}) := \sum_{a=1}^{m} \max\Big(-\operatorname{tr}\Big(\boldsymbol{H}\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]\Big), 0\Big)\hat{\boldsymbol{e}}_a, \quad \textit{(negative energy gradient)}$$

$$\boldsymbol{\nabla}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho}) := \sum_{a=1}^{m} \operatorname{tr}\Big(\boldsymbol{H}\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]\Big)\hat{\boldsymbol{e}}_a, \quad \textit{(energy gradient)}$$

where $\hat{\boldsymbol{e}}_a$ is the unit vector along the a-th coordinate.

Since the set of jump operators $\{\boldsymbol{A}^a\}_a$ will be fixed, we will sometimes drop the dependence on $\{\boldsymbol{A}^a\}_a$ for notational simplicity. The positive/negative energy gradient belongs to the tangent space $\mathbb{R}^m_{\geq 0}$, but the energy gradient

$$\boldsymbol{\nabla}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho}) = \boldsymbol{\nabla}^{+}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho}) - \boldsymbol{\nabla}^{-}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho})$$

may not be in the tangent space due to negative values. So, in general, one could not move in the direction of the energy gradient. However, one could move in the direction of the positive or negative energy gradient. It is instructive to think about the Heisenberg picture and define the *energy gradient operator*.

Definition 4.8.2 (Energy gradient operator)**.** *Given an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_a$, inverse temperature $\beta \geq 0$, and time scale $\tau \geq 0$, the energy gradient operators under thermal perturbations is*

$$\sum_{a=1}^{m} \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{H}]\ \hat{\boldsymbol{e}}_a,$$

which is a vector of n-qubit Hermitian operators.

We can provide an upper and lower bound to the energy gradients by combining Prop. 4.10.2 and Prop. 4.10.3 to obtain the following proposition.

Proposition 4.8.3 (Bound on the energy gradients)**.** *Given an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_a$, inverse temperature $\beta \geq 0$, and time scale $\tau \geq 0$,*

$$\|\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\|_\infty \leq 3\|\boldsymbol{H}\|_\infty$$

for all $a = 1, \ldots, m$.

The $\beta, \tau \to \infty$ limit (zero temperature heat bath with an infinite time scale) recovers the Davies' generator $\mathcal{L}_a^{\infty,\infty,\boldsymbol{H}}$. The Davies' generator takes an energy eigenvector $|\psi_j\rangle\langle\psi_j|$ of $\boldsymbol{H}$ to energy eigenvectors with equal or lower energy, i.e., for any $t \geq 0$,

$$\langle\psi_k| \exp\left(t\mathcal{L}_a^{\infty,\infty,\boldsymbol{H}}\right)\left(|\psi_j\rangle\langle\psi_j|\right)|\psi_k\rangle = 0 \quad \text{for any} \quad j,k \quad \text{such that} \quad E_k > E_j.$$

We can use the above to obtain the following proposition.

Proposition 4.8.4 (Vanishing positive energy gradient)**.** *For $\beta = \tau = \infty$, we have*

$$\mathcal{L}_a^{\infty,\infty,\boldsymbol{H}^\dagger}[\boldsymbol{H}] \preceq 0, \quad \textit{for each} \quad a.$$

Hence, the positive energy gradient vanishes $\boldsymbol{\nabla}^+_{\infty,\infty}(\boldsymbol{H},\boldsymbol{\rho}) = 0$ and $\boldsymbol{\nabla}_{\infty,\infty}(\boldsymbol{H},\boldsymbol{\rho}) = -\boldsymbol{\nabla}^-_{\infty,\infty}(\boldsymbol{H},\boldsymbol{\rho})$ for all Hamiltonian $\boldsymbol{H}$ and state $\boldsymbol{\rho}$.

This proposition illustrates that thermal perturbations induced by a zero-temperature heat bath with an infinite time scale should only absorb energy from the quantum system and not cause the energy to increase. Hence, the positive energy gradient must vanish.

A sufficient condition and a necessary condition of local minima

Using the negative gradient, we can show a necessary condition and a sufficient condition for local minima under thermal perturbations. They differ only slightly ($<$ vs $\leq$). From the conditions, we can see that local minima are well characterized by the negative energy gradient. Recall that $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$ is the ℓ_∞ norm for a finite-dimensional vector $\boldsymbol{x}$.

Lemma 4.8.1 (A sufficient condition for local minima under thermal perturbations)**.** *Given $\epsilon > 0$, an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_a$, and parameters $\beta, \tau \geq 0$, an n-qubit state $\boldsymbol{\rho}$ with a small negative energy gradient,*

$$\|\boldsymbol{\nabla}^-_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho})\|_\infty < \epsilon, \tag{4.14}$$

is an ϵ-approximate local minimum $\boldsymbol{\rho}$ of the n-qubit Hamiltonian $\boldsymbol{H}$ under thermal perturbations with inverse temperature β, time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$.

Proof. Consider $C_L = \max_a \|\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\|_{1-1} > 0$ and $C_H = \|\boldsymbol{H}\|_\infty$. Given $\boldsymbol{\alpha} \in \mathbb{R}^m_{\geq 0}$, we have

$$\left|\sum_a \sum_{a'} \alpha_a \alpha_{a'} \operatorname{tr}(\boldsymbol{H}\mathcal{L}_{a'}^{\beta,\tau,\boldsymbol{H}}[\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\sigma}]])\right| \leq C_L^2 C_H \|\boldsymbol{\alpha}\|_1^2,$$

for any state $\boldsymbol{\sigma}$. Let $\epsilon_0 := \|\boldsymbol{\nabla}^-_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho})\|_\infty < \epsilon$. From $\alpha_a \geq 0$ and Cauchy-Schwarz inequality,

$$\sum_a \alpha_a \operatorname{tr}(\boldsymbol{H}\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]) \geq -\sum_a |\alpha_a| \max\left(-\operatorname{tr}\left(\boldsymbol{H}\mathcal{L}_i^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]\right), 0\right) \geq -\|\boldsymbol{\alpha}\|_1 \epsilon_0.$$

Together, Taylor's theorem for thermal perturbations (Prop. 4.8.2) implies

$$\operatorname{tr}\left(\boldsymbol{H}\exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}(\boldsymbol{\alpha})\right) \geq \operatorname{tr}(\boldsymbol{H}\boldsymbol{\rho}) - \|\boldsymbol{\alpha}\|_1\epsilon_0 - \frac{\|\boldsymbol{\alpha}\|_1^2}{2} C_L^2 C_H,$$

for any $\boldsymbol{\alpha} \in \mathbb{R}^m_{\geq 0}$. From the above, we see that for any $\|\boldsymbol{\alpha}\|_1^2 < \delta := \frac{2(\epsilon-\epsilon_0)}{C_L^2 C_H}$,

$$\begin{aligned}\operatorname{tr}\left(\boldsymbol{H}\exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}(\boldsymbol{\alpha})\right) &\geq \operatorname{tr}(\boldsymbol{H}\boldsymbol{\rho}) - \epsilon\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\alpha}\|_1\left((\epsilon-\epsilon_0) - \frac{C_L^2 C_H}{2}\|\boldsymbol{\alpha}\|_1^2\right) \\ &\geq \operatorname{tr}(\boldsymbol{H}\boldsymbol{\rho}) - \epsilon\|\boldsymbol{\alpha}\|_1.\end{aligned}$$

So, $\boldsymbol{\rho}$ is an ϵ-approximate local minimum of $\boldsymbol{H}$ under thermal perturbations with inverse temperature β, time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$. ∎

Lemma 4.8.2 (A necessary condition for local minima under thermal perturbations)**.** *Given $\epsilon > 0$, an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_a$, and parameters $\beta, \tau \geq 0$, an ϵ-approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}$ under thermal perturbations with inverse temperature β, time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ satisfies*

$$\|\boldsymbol{\nabla}^-_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho})\|_\infty \leq \epsilon,$$

which differs only slightly from the condition in Eq. (4.14).

Proof. Recall that $\boldsymbol{\nabla}^-_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho}) \in \mathbb{R}^m_{\geq 0}$. Let $a^* = \arg\max_a \left(\nabla^-_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho})_a\right)$. If the negative energy gradient vector $\mathrm{tr}(\boldsymbol{H}\mathcal{L}^{\beta,\tau,\boldsymbol{H}}_{a^*}[\boldsymbol{\rho}])$ is zero, then the claim holds. Hence, we only need to consider the case when the negative energy gradient vector is nonzero. In this case,

$$0 < \|\boldsymbol{\nabla}^-_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho})\|_\infty = \nabla^-_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho})_{a^*} = -\mathrm{tr}(\boldsymbol{H}\mathcal{L}^{\beta,\tau,\boldsymbol{H}}_{a^*}[\boldsymbol{\rho}]).$$

Consider $\hat{\boldsymbol{\alpha}} := \hat{\boldsymbol{e}}_{a^*} \in \mathbb{R}^m_{\geq 0}$, which satisfies $\|\hat{\boldsymbol{\alpha}}\|_1 = 1$. We have

$$\begin{aligned}\lim_{t\to 0^+} \frac{\mathrm{tr}(\boldsymbol{H}\exp^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}_{\boldsymbol{\rho}}(t\hat{\boldsymbol{\alpha}})) - \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})}{t} &= \mathrm{tr}(\boldsymbol{H}\mathcal{L}^{\beta,\tau,\boldsymbol{H}}_{a^*}[\boldsymbol{\rho}]) \\ &= -\|\boldsymbol{\nabla}^-_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho})\|_\infty.\end{aligned}$$

At the same time, for any $t > 0$, we also have

$$\frac{\mathrm{tr}(\boldsymbol{H}\exp^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}_{\boldsymbol{\rho}}(t\hat{\boldsymbol{\alpha}})) - \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})}{t} \geq -\epsilon\|\hat{\boldsymbol{\alpha}}\|_1 = -\epsilon.$$

Together, we obtain the desired claim. ∎

Hamiltonians without suboptimal local minima

An important concept in classical optimization is to understand when all local minima are global minima. For example, in convex optimization, checking the convexity of the objective function $h(\boldsymbol{x})$ ensures that all local minima are global minima. When all local minima are global minima, it is commonly referred to as having no suboptimal local minima in the optimization landscape. For optimizing quantum Hamiltonians, we can define a similar concept. Let us begin with a definition of approximate global minimum.

Definition 4.8.3 (Approximate global minimum of Hamiltonians)**.** *Given $\epsilon, \delta > 0$ and an n-qubit Hamiltonian $\boldsymbol{H}$ with minimum energy E_0. Let $\boldsymbol{P}_{G+\epsilon}(\boldsymbol{H})$ be the projector to the subspace of energy eigenstates of $\boldsymbol{H}$ with energy at most $E_0 + \epsilon$. An n-qubit state $\boldsymbol{\rho}$ is an ϵ-approximate global minimum of $\boldsymbol{H}$ with failure probability $\leq \delta$ if* $\mathrm{tr}(\boldsymbol{P}_{G+\epsilon}(\boldsymbol{H})\boldsymbol{\rho}) \geq 1 - \delta$.

Definition 4.8.4 (No suboptimal local minima)**.** *Given $\epsilon > 0$. We say an n-qubit Hamiltonian $\boldsymbol{H}$ has no suboptimal ϵ-approximate local minima with failure probability δ if any ϵ-approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}$ is an ϵ-approximate global minimum of $\boldsymbol{H}$ with failure probability $\leq \delta$, i.e.,* $\mathrm{tr}(\boldsymbol{P}_{G+\epsilon}(\boldsymbol{H})\boldsymbol{\rho}) \geq 1 - \delta$.

While the above definitions apply to any Hamiltonian $\boldsymbol{H}$, in this work, we will focus on Hamiltonians with a gap $\Delta > 0$ between the minimum energy and the second minimum energy, also known as the spectral gap. By definition of $\boldsymbol{P}_{G+\epsilon}(\boldsymbol{H})$ and spectral gap Δ, we have

$$\epsilon < \Delta \implies \boldsymbol{P}_{G+\epsilon}(\boldsymbol{H}) = \boldsymbol{P}_G(\boldsymbol{H}).$$

As we will almost always consider $\epsilon < \Delta$, any ϵ-approximate global minimum is an 0-approximate global minimum or *exact* global minimum.

Convexity implies all local minima are global in classical optimization. In the following lemma, we present a sufficient condition for ensuring that all local minima are global in quantum systems. As we can see, all we need is to check the negative gradient operator is sufficiently positive in the non-ground-state space $\boldsymbol{I} - \boldsymbol{P}_G$. We will refer to this as the *negative gradient condition*.

Lemma 4.8.3 (A sufficient condition ensuring all local minima are global)**.** *Given $\epsilon, \delta > 0$, an n-qubit Hamiltonian $\boldsymbol{H}$, m local jump operators $\{\boldsymbol{A}^a\}_a$, and parameters $\beta, \tau \geq 0$. Let $\boldsymbol{P}_G(\boldsymbol{H})$ be the projection onto the ground state space of $\boldsymbol{H}$. If there exists $\boldsymbol{\alpha} \in \mathbb{R}^m_{\geq 0}$ with $\|\boldsymbol{\alpha}\|_1 = 1$, such that the negative gradient operator satisfies*

$$\textit{(negative gradient condition):} \quad -\sum_a \alpha_a \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{H}] \geq \frac{2\epsilon}{\delta}(\boldsymbol{I} - \boldsymbol{P}_G(\boldsymbol{H})) - \epsilon \boldsymbol{I}, \tag{4.15}$$

then any ϵ-approximate local minimum $\boldsymbol{\rho}$ of the n-qubit Hamiltonian $\boldsymbol{H}$ under thermal perturbations with inverse temperature β, time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ is an exact global minimum with failure probability $\leq \delta$. That is, $\mathrm{tr}(\boldsymbol{P}_G(\boldsymbol{H})\boldsymbol{\rho}) \geq 1 - \delta$.

Proof. From the necessary condition for local minima in Lemma 4.8.2, any ϵ-approximate local minimum $\boldsymbol{\rho}$ of the n-qubit Hamiltonian $\boldsymbol{H}$ under thermal perturbations with inverse temperature β, time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ satisfies

$$-\mathrm{tr}(\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{H}]\boldsymbol{\rho}) \leq \epsilon \quad \text{for each} \quad a = 1, \ldots, m.$$

Hence, from Eq. (4.15), we have

$$\epsilon \geq -\sum_a \alpha_a \,\mathrm{tr}(\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{H}]\boldsymbol{\rho}) \geq \frac{2\epsilon}{\delta}(1 - \mathrm{tr}(\boldsymbol{P}_G(\boldsymbol{H})\boldsymbol{\rho})) - \epsilon.$$

This immediately implies that $\mathrm{tr}(\boldsymbol{P}_G(\boldsymbol{H})\boldsymbol{\rho}) \geq 1 - \delta$. ∎

4.9 Appendix:Complexity of finding a local minimum in quantum systems

In this appendix, we formally present the main results of this paper shown earlier in section 4.2 regarding the computational complexity of finding a local minimum in quantum systems. We separate the results into two parts. First, we look at the problem of finding a local minimum under local unitary perturbations (Def. 4.6.7), showing that the problem is classically trivial to solve. Next, we look at the problem of finding a local minimum under low-temperature thermal perturbations (Def. 4.6.6). We will see that this problem is quantumly easy but classically hard to solve, establishing a promising candidate problem for quantum advantage.

Finding a local minimum under local unitary perturbations

We begin with the first main result stating the problem of a local minimum under local unitary perturbations is classically trivial. The main issue is that there is a large barren plateau (which consists of many local minima with high energy) in the quantum optimization landscape. Hence, a classical algorithm can efficiently estimate the properties of a single local minimum.

Proposition 4.9.1 (Classically easy to find a local minimum under local unitary perturbations; Restatement of Proposition 4.2.1)**.** *Consider a large problem size n. There is a trivial classical algorithm that guarantees the following. Given error $\epsilon = 1/\mathrm{poly}(n)$, an n-qubit local Hamiltonian $\boldsymbol{H}$ with $\|\boldsymbol{H}\|_\infty = \mathrm{poly}(n)$, m local Hermitian operators $\{\boldsymbol{h}^a\}_{a=1}^m$ with $m = \mathrm{poly}(n)$ and $\|\boldsymbol{h}_a\|_\infty = 1$, and a local observable $\boldsymbol{O}$ with $\|\boldsymbol{O}\|_\infty \leq 1$.*

The classical algorithm runs in time $O(1)$ and outputs a real value $v \in [-1, 1]$, such that v is ϵ-close to $\langle\psi|\boldsymbol{O}|\psi\rangle$ for an ϵ-approximate local minimum $|\psi\rangle$ of the Hamiltonian $\boldsymbol{H}$ under local unitary perturbations generated by $\{\boldsymbol{h}^a\}_a$.

Proof. From Lemma 4.7.1 given in Appendix 4.7 characterizing local minima of $\boldsymbol{H}$ under local unitary perturbations, with high probability, a state $|\psi\rangle$ sampled uniformly at random from the space of pure states is an ϵ-approximate local minimum $|\psi\rangle$ of the local Hamiltonian $\boldsymbol{H}$ under local unitary perturbations, and $\langle\psi|\boldsymbol{O}|\psi\rangle$ is ϵ-close to $\mathrm{tr}(\boldsymbol{O})/2^n$. Hence, there exists an ϵ-approximate local minimum $|\psi\rangle$ of $\boldsymbol{H}$ under local unitary perturbations, and $\langle\psi|\boldsymbol{O}|\psi\rangle$ is ϵ-close to $\mathrm{tr}(\boldsymbol{O})/2^n$.

This characterization of local minima gives rise to the following trivial classical algorithm. Given a local observable $\boldsymbol{O}$, represented by the subset S of qubits $\boldsymbol{O}$ acts on with $|S| = O(1)$ and a $2^{|S|\times|S|}$ Hermitian matrix $\boldsymbol{O}^*$. A classical algorithm can

compute $\mathrm{tr}(\boldsymbol{O})/2^n$ efficiently by computing the trace of $\boldsymbol{O}^*$ and dividing by $2^{|S|}$. This trivial classical algorithm runs in time $O(1)$. ■

Finding a local minimum under thermal perturbations

We now turn to the second main result of this work, which shows that finding a local minimum under low-temperature thermal perturbations is easy with a quantum computer. This is in contrast to the task of finding the ground state (global minimum), which is hard on quantum computers. The formal statement is given below in Theorem 4.9.1.

Theorem 4.9.1 (Quantumly easy to find a local minimum under thermal perturbations; Restatement of Theorem 4.2.1)**.** *Let n be the problem size. There is a $\mathrm{poly}(n)$-time quantum algorithm that guarantees the following. Given error $\epsilon = 1/\mathrm{poly}(n)$, inverse temperature $0 \leq \beta \leq \mathrm{poly}(n)$, time scale $\tau = \mathrm{poly}(n)$, an n-qubit local Hamiltonian $\boldsymbol{H}$ with $\|\boldsymbol{H}\|_\infty = \mathrm{poly}(n)$, m local jump operators $\{\boldsymbol{A}^a\}_{a=1}^m$ with $m = \mathrm{poly}(n)$, and a local observable $\boldsymbol{O}$ with $\|\boldsymbol{O}\|_\infty \leq 1$.*

The quantum algorithm outputs a real value $v \in [-1, 1]$, such that v is ϵ-close to $\mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho})$ for an ϵ-approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}$ under thermal perturbations with an inverse temperature β, a time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$.

Proof idea. We consider a version of gradient descent, which we refer to as *Quantum thermal gradient descent*, that mimics how Nature cools the quantum system when the system is interacting locally and weakly with a low-temperature heat bath. The algorithm starts with an arbitrary initial state $\boldsymbol{\rho}_0$. For each step $t = 0, 1, 2, \ldots$, the algorithm considers the current state $\boldsymbol{\rho}_t$ and proposes the next state $\boldsymbol{\rho}_{t+1}$. The tangent space $T_{\boldsymbol{\rho}_t}^{\beta,\tau,\boldsymbol{H}}$ at $\boldsymbol{\rho}_t$ is high dimensional with many possible directions/dynamics depending on the system-bath interaction. The algorithm chooses a direction that lowers the energy as fast as possible by computing the gradient of the energy and proposes $\boldsymbol{\rho}_{t+1}$ by performing gradient descent. As long as the current state $\boldsymbol{\rho}_t$ is not an ϵ-approximate local minimum of $\boldsymbol{H}$ under thermal perturbations, the energy will decrease by a sufficiently large amount

$$\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}_{t+1}) < \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}_t) - \frac{1}{\mathrm{poly}(n)}.$$

Because the energy is bounded from below, there are, at most, a polynomial number of steps $t \leq \mathrm{poly}(n)$ until the algorithm arrives at an ϵ-approximate local minimum

of $\boldsymbol{H}$ under thermal perturbations. The detailed proof of Theorem 4.9.1 is given in Appendix 4.11. ■

Finally, we turn to the third main result establishing the difficulty of finding a local minimum under thermal perturbations using a classical computer. To establish this result, we consider a class of geometrically local Hamiltonians $\{\boldsymbol{H}_C\}_C$ on 2D lattices. Each Hamiltonian $\boldsymbol{H}_C$ corresponds to a 2D circuit $\boldsymbol{U}_C = \boldsymbol{U}_T \cdots \boldsymbol{U}_2 \boldsymbol{U}_1$ acting on n qubits with $T = 2t_0 + L = \text{poly}(n)$ gates as constructed in Fig. 1 of [142] with the additional padding to the construction in [142] such that the first and last $t_0 = cL^2$ gates being the identity gates for a constant $c = O(1)$. The construction in [142] has the property that each gate of the 2D circuit $\boldsymbol{U}_C$ is geometrically adjacent to the subsequent gate.

Given the 2D circuit $\boldsymbol{U}_C$ on n qubits with T gates. The geometrically local Hamiltonian $\boldsymbol{H}_C$ acts on $n+T$ qubits on a 2D lattice and has a highly-entangled unique ground state that encodes the quantum computation based on the 2D circuit $\boldsymbol{U}_C$,

$$|\eta_{\boldsymbol{0}}\rangle = \sum_{t=0}^{T} \sqrt{\xi_t} \left(\boldsymbol{U}_t \cdots \boldsymbol{U}_1 |0^n\rangle\right) \otimes \left|1^t 0^{T-t}\right\rangle, \qquad \text{where} \quad \xi_t := \frac{1}{2^T} \binom{T}{t}.$$

We present the detailed construction of the 2D Hamiltonian $\boldsymbol{H}_C$ in Definition 4.14.1 in Appendix 4.14. We have the following proposition for estimating single-qubit observables on the ground state of $\boldsymbol{H}_C$.

Proposition 4.9.2 (BQP-hardness for estimating properties of the ground state of $\boldsymbol{H}_C$)**.** *If there is a classical algorithm that can estimate any single-qubit observable on the unique ground state of the geometrically local Hamiltonian $\boldsymbol{H}_C$ in time polynomial in the number of qubits in $\boldsymbol{H}_C$ to error $1/4$ for any $\boldsymbol{H}_C$ in the class, then* BPP = BQP.

Proof. Consider the single-qubit observable $\boldsymbol{Z}_j$ and let T_j be the last time that qubit j is acted by a gate in the circuit C. The ground state expectation of $\boldsymbol{Z}_j$ is

$$\begin{aligned}
\langle \eta_{\boldsymbol{0}} | \boldsymbol{Z}_j | \eta_{\boldsymbol{0}} \rangle &= \sum_{t=T_j+1}^{T} \xi_t \langle 0^n | \boldsymbol{U}_1^\dagger \cdots \boldsymbol{U}_t^\dagger \boldsymbol{Z}_j \boldsymbol{U}_t \cdots \boldsymbol{U}_1 | 0^n \rangle + \sum_{t=0}^{T_j} \xi_t \langle 0^n | \boldsymbol{U}_1^\dagger \cdots \boldsymbol{U}_t^\dagger \boldsymbol{Z}_j \boldsymbol{U}_t \cdots \boldsymbol{U}_1 | 0^n \rangle \\
&= \langle 0^n | \boldsymbol{U}_1^\dagger \cdots \boldsymbol{U}_T^\dagger \boldsymbol{Z}_j \boldsymbol{U}_T \cdots \boldsymbol{U}_1 | 0^n \rangle P_{t>T_j} + \epsilon_j, \\
\text{where} \quad P_{t>T_j} &:= \sum_{t>T_j} \xi_t, \quad \epsilon_j := \sum_{t \le T_j} \xi_t \langle 0^n | \boldsymbol{U}_1^\dagger \cdots \boldsymbol{U}_t^\dagger \boldsymbol{Z}_j \boldsymbol{U}_t \cdots \boldsymbol{U}_1 | 0^n \rangle.
\end{aligned}$$

We have used the fact that U_t for $t > T_j$ acts like identity on the j-th qubit. Note that

$$|\epsilon_j| \leq 1 - P_{t>T_j} =: P_{t\leq T_j}.$$

We can make ϵ_j arbitrarily small using a tail bound on the binomial distribution. Given any circuit, one could always pad more identity gates to form an L-gate circuit, such that the last 3/4 of the L gates are identity. Recall that $T = 2t_0 + L = 2cL^2 + L$. Then $T_j \leq cL^2 + L/4$ and we have $|\epsilon_j| \leq P_{t\leq T_j} \leq P_{t\leq cL^2+L/4}$.

Using Hoeffding's inequality, we can bound the probability of sampling a time t, such that $t \leq cL^2 + L/4$, according to the Binomial distribution $\{\xi_t\}_{t=0}^{T}$. This yields

$$|\epsilon_j| \leq \exp\left[-2T\left(\frac{1}{2} - \frac{cL^2 + L/4}{2cL^2 + L}\right)^2\right] = \mathrm{e}^{-\frac{L}{8+16cL}}.$$

By choosing a small constant $c \leq 1/(16\ln 18) - 1/(2L)$, we have $|\epsilon_j| \leq 1/18$ and $P_{t>T_j} \geq 17/18$. Because of the bounds on the error $|\epsilon_j|$ and the probability $P_{t>T_j}$, a classical algorithm satisfying the assumption of the proposition can determine whether

$$\langle 0^n|\boldsymbol{U}_1^\dagger \cdots \boldsymbol{U}_T^\dagger Z_j \boldsymbol{U}_T \cdots \boldsymbol{U}_1|0^n\rangle > 1/3 \quad \text{or} \quad \langle 0^n|\boldsymbol{U}_1^\dagger \cdots \boldsymbol{U}_T^\dagger Z_j \boldsymbol{U}_T \cdots \boldsymbol{U}_1|0^n\rangle < -1/3,$$

for any 2D circuit $\boldsymbol{U}_C$ with $T = 2t_0 + L = 2cL^2 + L$ gates, where the first t_0 and and the last $(3/4)L + t_0$ gates are identity. Because one could think of the circuit $\boldsymbol{U}_C$ as having $L/4$ gates for any $L = \mathrm{poly}(n)$, this immediately implies that a polynomial-time classical algorithm can decide whether the expectation value of Z_j on the output state is $> 1/3$ or $< -1/3$ for any polynomial-size 2D circuit where all consecutive gates are adjacent in the 2D geometry.

A 2D circuit $\boldsymbol{U}_C$ such that any gate is adjacent to the subsequent gate can be constructed from any quantum circuit without the 2D constraint, such that a single-qubit observable Z_i on the output of the original quantum circuit corresponds to a single-qubit observable Z_j on the output of the 2D circuit. As a result, any polynomial-time classical algorithm that can determine whether the expectation value of Z_j on $\boldsymbol{U}_t \cdots \boldsymbol{U}_1|0^n\rangle$ is greater than $1/3$ or smaller than $-1/3$ can be used to simulate any polynomial-time quantum algorithm for solving decision problems in classical polynomial time. Hence, BPP = BQP. ■

Using a series of mathematical techniques presented in Appendix 4.12 for characterizing whether all local minima are global minima in a many-body Hamiltonian,

we prove that all local minima $\boldsymbol{H}_C$ are close to the unique ground state $|\eta_{\mathbf{0}}\rangle$ in Theorem 4.9.2. This theorem is the most involved technical contribution of this work. Intuitively, one can think of the energy landscape of the 2D Hamiltonian $\boldsymbol{H}_C$ over the space of n-qubit density matrices under low-temperature thermal perturbations to have a good bowl shape. This is in stark contrast to the energy landscape under local unitary perturbations, where the landscape always contains an overwhelmingly large barren plateau causing the problem of finding local minima to be classically easy. Furthermore, this theorem shows that a low-temperature cooling can always find a state close to the ground state irrespective of where we initialize the state in the exponentially large quantum state space.

Theorem 4.9.2 (All local minima are global in BQP-hard Hamiltonians; Restatement of Theorem 4.2.2)**.** *Let $\boldsymbol{P}_G(\boldsymbol{H}_C) = |\eta_{\mathbf{0}}\rangle\langle\eta_{\mathbf{0}}|$ be the ground state of the 2D Hamiltonian $\boldsymbol{H}_C$ acting on $n+T = \mathrm{poly}(n)$ qubits. There is a choice of $m = \mathrm{Poly}(n)$ two-qubit jump operators $\{\boldsymbol{A}^a\}_a$ satisfying the following.*

Given $0 < \delta < 1$. For any small error $\epsilon = 1/\mathrm{Poly}(n, 1/\delta)$, any ϵ-approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}_C$ under thermal perturbations with a large inverse temperature $\beta = \mathrm{Poly}(n, 1/\delta)$, a large time scale $\tau = \mathrm{Poly}(n, 1/\delta)$, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ is an exact global minimum with high probability, i.e., we have $\mathrm{tr}(\boldsymbol{P}_G(\boldsymbol{H}_C)\boldsymbol{\rho}) \geq 1 - \delta$.

The proof of Theorem 4.9.2 is given in Appendix 4.14. To show that the landscape has a good bowl shape, we utilize the negative gradient condition given in Appendix 4.8. However, the negative energy gradient operator is not easy to study. To establish this strong claim, we give a series of techniques in Appendix 4.12 for characterizing negative energy gradient operator in few-qubit systems, in commuting Hamiltonians, and in perturbed Hamiltonians. These technical tools can also be used to understand the energy landscape in other interacting many-body Hamiltonians.

While finding a local minimum under local unitary perturbations is classically easy, the characterization of the energy landscape in these BQP-hard Hamiltonians $\boldsymbol{H}_C$ implies that finding a local minimum under thermal perturbations is *universal for quantum computation* and is hence classically hard if BPP $\neq$ BQP. Recall that BPP = BQP implies that all single-qubit measurements of all polynomial-size quantum circuits can be simulated in polynomial time on a classical computer. Since one expects some quantum circuits to be hard to simulate on a classical computer,

Theorem 4.9.3 implies that finding a local minimum under thermal perturbations is classically hard.

Theorem 4.9.3 (Classically hard to find a local minimum under thermal perturbations; Restatement of Theorem 4.2.3)**.** *Let n be the problem size. Suppose there is a* $\mathrm{poly}(n)$*-time classical algorithm guaranteeing the following. Given error* $\epsilon = 1/\mathrm{poly}(n)$*, inverse temperature* $0 \leq \beta \leq \mathrm{poly}(n)$*, time scale* $0 \leq \tau \leq \mathrm{poly}(n)$*, an* n*-qubit local Hamiltonian* $\boldsymbol{H}$ *with* $\|\boldsymbol{H}\|_\infty = \mathrm{poly}(n)$*,* m *local jump operators* $\{\boldsymbol{A}^a\}_{a=1}^m$ *with* $m = \mathrm{poly}(n)$*, and a single-qubit observable* $\boldsymbol{O}$ *with* $\|\boldsymbol{O}\|_\infty \leq 1$*.*

The classical algorithm outputs a real value $v \in [-1, 1]$*, such that* v *is* ϵ*-close to* $\mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho})$ *for an* ϵ*-approximate local minimum* $\boldsymbol{\rho}$ *of the Hamiltonian* $\boldsymbol{H}$ *under thermal perturbations with an inverse temperature* β*, a time scale* τ*, and system-bath interactions generated by* $\{\boldsymbol{A}^a\}_a$*. Then* BPP = BQP*.*

Proof. Assuming the existence of a polynomial-time classical algorithm that satisfies the properties stated in the theorem. Apply this classical algorithm to the 2D Hamiltonian $\boldsymbol{H}_C$ considered in Theorem 4.9.2 with a sufficiently small approximation error ϵ, such that any ϵ-approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}_C$ under thermal perturbations with polynomially-large β, τ and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$ is an exact global minimum with high probability, i.e.,

$$\langle \eta_{\boldsymbol{0}} | \boldsymbol{\rho} | \eta_{\boldsymbol{0}} \rangle = \mathrm{tr}(\boldsymbol{P}_G(\boldsymbol{H}_C)\boldsymbol{\rho}) \geq 1 - \frac{1}{16^2},$$

where $|\eta_{\boldsymbol{0}}\rangle$ is the unique ground state of $\boldsymbol{H}_C$. We further consider ϵ to be small enough such that

$$\epsilon < \frac{1}{8}. \tag{4.16}$$

Let $\boldsymbol{\rho}$ be an ϵ-approximate local minimum of the Hamiltonian $\boldsymbol{H}$ under thermal perturbations. Consider the observable $\boldsymbol{O}_j = \boldsymbol{Z}_j$ from the proof of Proposition 4.9.2. Using the Fuchs–van de Graaf inequalities, we have

$$\|\boldsymbol{\rho} - |\eta_{\boldsymbol{0}}\rangle\langle\eta_{\boldsymbol{0}}|\|_1 \leq \frac{1}{8}. \tag{4.17}$$

Because the classical algorithm can estimate $\mathrm{tr}(\boldsymbol{O}_j\boldsymbol{\rho})$ to error ϵ, from Eq. (4.16) and (4.17), the classical algorithm can estimate $\langle \eta_{\boldsymbol{0}} | \boldsymbol{O}_j | \eta_{\boldsymbol{0}} \rangle$ to error $1/4$ in time polynomial in the number of qubits in $\boldsymbol{H}_C$. From Prop. 4.9.2, this implies that BPP = BQP. ∎

In the following, we use the previous theorem to show that quantum machines can improve over any efficient classical algorithm that variationally optimizes a

classical ansatz that can efficiently predict local properties by performing low-temperature cooling. Examples of the classical ansatz include tensor networks with efficient tensor contraction algorithms and neural-network quantum states with fast sampling algorithms. This result provides a physically-relevant problem that yields an advantage in minimizing the energy of a geometrically-local Hamiltonian.

Corollary 4.9.1 (Quantum advantage over variationally optimized classical ansatz)**.** *Under the conjecture that* BPP $\neq$ BQP*, there exists a class of n-qubit geometrically-local Hamiltonian $\boldsymbol{H}$ on a two-dimensional lattice with $\|\boldsymbol{H}\|_\infty = O(n)$ that satisfies the following. Given any classical ansatz of n-qubit state $\boldsymbol{\rho}$ that can estimate the expectation value of single-qubit observables to $1/\mathrm{poly}(n)$ error in $\mathrm{poly}(n)$-time on classical computers, any $\mathrm{poly}(n)$-time classical algorithm for minimizing the energy $\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho})$ using the classical ansatz, and samples of the state $\boldsymbol{\rho}$ represented by the optimized classical ansatz. A quantum machine can find a state $\boldsymbol{\rho}^{\#}$ with strictly lower energy than $\boldsymbol{\rho}$ in $\mathrm{poly}(n)$ time by running a quantum thermal gradient descent based on low-temperature cooling.*

Proof. The central claim is that the state $\boldsymbol{\rho}$ found by an efficient classical algorithm cannot be an ϵ-approximate local minimum under low-temperature thermal perturbations. We establish this claim by contradiction. Suppose that the classical ansatz for $\boldsymbol{\rho}$ found by the efficient classical algorithm is an ϵ-approximate local minimum. Then the classical algorithm can use the classical ansatz to predict the expectation values of single-qubit observables of an ϵ-approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}$ to ϵ error. From Theorem 4.9.3, this implies that BPP = BQP, which is a contradiction.

Because $\boldsymbol{\rho}$ is not an ϵ-approximate local minimum under low-temperature thermal perturbations, a quantum machine can use samples of $\boldsymbol{\rho}$ to initialize at the state $\boldsymbol{\rho}$ and perform one gradient descent step based on low-temperature cooling. From Lemma 4.8.2 on the necessary condition for local minima, there exists $a \in \{1, \ldots, m\}$ such that $\mathrm{tr}(\boldsymbol{H}\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]) < -\epsilon$. From Lemma 4.11.1 on cooling by gradient descent, one can show that a single gradient descent step yields a state $\boldsymbol{\rho}^{(\mathrm{next})}$ with a strictly lower energy than the state $\boldsymbol{\rho}$. Hence, one establishes the desired claim. ∎

4.10 Appendix:Details of thermal Lindbladians

In the rest of the appendices, we give the full detailed proofs of Theorems 4.9.1 and 4.9.2 that are central to establishing the computational complexity of finding local minima under thermal perturbations in the previous appendix. To that end,

we need to provide the techincal details of thermal Lindbladians that generate such perturbations.

We have previously presented a high-level introduction to thermal Lindbladians in Appendix 4.5. This has been sufficient for defining local minima and analyzing some basic properties, but not enough for proving Theorems 4.9.1 and 4.9.2. In this appendix, we present the exact form of thermal Lindbladians, their properties, and the algorithmic primitives for simulating quantum thermodynamics.

Exact form

The exact form of the thermal Lindbladian depends on a few physical concepts due to the microscopic derivation from a system-bath interaction [135]. For each jump $\boldsymbol{A}^a$, we have

$$\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}(\boldsymbol{\rho}) := -\mathrm{i}[\boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}}, \boldsymbol{\rho}] + \underbrace{\int_{-\infty}^{\infty} \gamma_\beta(\omega)\left[\hat{\boldsymbol{A}}^a(\omega)\boldsymbol{\rho}\hat{\boldsymbol{A}}^a(\omega)^\dagger - \frac{1}{2}\{\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega), \boldsymbol{\rho}\}\right]\mathrm{d}\omega}_{:=\mathcal{D}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]}, \tag{4.18}$$

where $\mathcal{D}_a^{\beta,\tau,\boldsymbol{H}}$ is the purely dissipative part of the thermal Lindbladian. Implicitly, the operator $\hat{\boldsymbol{A}}^a(\omega)$ also depends on the Hamiltonian $\boldsymbol{H}$ and the time scale τ. We now unpack the physical concepts that form the building blocks of this expression.

Transition weight. At a fixed inverse temperature β, the *transition weight* $\gamma_\beta(\omega)$ tells us how strong the rate of a transition/jump should be, depending on the energy difference ω. In particular, the transition weight satisfies the following *Kubo-Martin-Schwinger (KMS) condition* and convenient normalization

$$\gamma_\beta(\omega)/\gamma_\beta(-\omega) = \mathrm{e}^{-\beta\omega} \quad \text{and} \quad 0 \le \gamma_\beta(\omega) \le 1 \quad \text{for any} \quad \beta \ge 0 \quad \text{and any} \quad \omega \in \mathbb{R}, \tag{4.19}$$

which is reminiscent of how detailed balance is enforced in classical Markov chains. We remark that any $\gamma_\beta(\omega)$ obeying the above KMS condition and normalization also satisfies the following tail bound:

$$\max_{\omega\ge\Delta} \omega\gamma_\beta(\omega) \le \max_{\omega\ge\Delta} \omega\mathrm{e}^{-\beta\omega} = \frac{1}{\beta}\max_{x\ge\beta\Delta} x\mathrm{e}^{-x} \le \frac{1}{\beta}\max_{x\ge\beta\Delta} \mathrm{e}^{-x/2} = \frac{\mathrm{e}^{-\beta\Delta/2}}{\beta}. \tag{4.20}$$

For concreteness, we usually adopt the common choice of γ_β corresponding to *Glauber dynamics*, with a cut-off frequency Λ_0 to regulate the inverse Fourier transform:

$$\gamma_\beta(\omega) = \frac{1}{2 + \ln(1+\beta\Lambda_0)} \frac{\mathrm{e}^{-\omega^2/2\Lambda_0^2}}{1+\mathrm{e}^{\beta\omega}}. \tag{4.21}$$

In the zero temperature regime ($\beta = \infty$), the function $(1 + e^{\beta\omega})^{-1}$ gives a step function (one for negative ω and zero for positive ω). Based on the choice of bath phenomenology, there are plenty of options for the transition weight, such as ohmic heating $\gamma_\beta(\omega) = \frac{1}{\omega_0}\frac{\omega e^{-\omega^2/2\Lambda_0^2}}{1-e^{-\beta\omega}}$. However, for simplicity, we will stick to the Glauber dynamics. Unless otherwise stated, we will also choose the cut-off frequency to be

$$\Lambda_0 = 1,$$

since each local jumps $\boldsymbol{A}^a$ changes the energy by at most $O(1)$ for our usage (and is generally true for local Hamiltonians with bounded degree interaction graph and bounded-norm terms). We do not expect our main conclusion to change under other reasonable choices of $\gamma_\beta(\omega)$.

Operator Fourier transform. Given a jump operator $\boldsymbol{A}^a$, we consider the *operator Fourier Transform* [39] for the Heisenberg-evolved jump operator $\boldsymbol{A}^a(t)$ characterized by a time scale $\tau \in \mathbb{R}$ of the heat bath

$$\hat{\boldsymbol{A}}^a(\omega) := \frac{1}{\sqrt{2\pi\tau}} \int_{-\tau/2}^{\tau/2} \underbrace{e^{i\boldsymbol{H}t}\boldsymbol{A}^a e^{-i\boldsymbol{H}t}}_{=:\boldsymbol{A}^a(t)} e^{-i\omega t} dt.$$

The operator $\hat{\boldsymbol{A}}^a(\omega)$ corresponds to matrix elements in $\boldsymbol{A}^a$ that induce jumps between energy eigenstates with an energy difference approximately $\omega \pm O(\frac{1}{\tau})$. The bigger τ is, the more precise ω corresponds to the true energy difference; see Appendix 4.15 for further details. Physically, τ is related to microscopic parameters of the bath (the bath correlation time and the weak-coupling strength [135]), but our discussion only requires the single time scale τ that sets the Fourier transform energy uncertainty.

Lamb-shift. The interaction with the heat bath induces an additional correction term in the coherent Hamiltonian dynamics of the n-qubit system, known as the *Lamb-shift*. Given a jump operator $\boldsymbol{A}^a$, we have the following Lamb-shift Hamiltonian that depends on the bath correlation function $c_\beta(t)$ and the time scale τ:

$$\boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}} := \frac{i}{2\sqrt{2\pi}\tau} \int_{-\tau/2}^{\tau/2} \int_{-\tau/2}^{\tau/2} \mathrm{sgn}(t_1 - t_2) c_\beta(t_2 - t_1) \boldsymbol{A}^a(t_2) \boldsymbol{A}^a(t_1) dt_2 dt_1 \quad (4.22)$$

While the Lamb-shift term is physically important, for our purposes, it is largely treated as a source of error; the energy gradient contribution comes from the dissipative part $\mathcal{D}_a^{\beta,\tau,\boldsymbol{H}}$.

Bath correlation function. In the Lamb-shift term, the bath correlation function $c_\beta(t)$ is the Fourier transform of the transition weight $\gamma_\beta(\omega)$,

$$c_\beta(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \gamma_\beta(\omega) \mathrm{e}^{+\mathrm{i}\omega t} \mathrm{d}t.$$

The prefactor in Eq. (4.21) is chosen such that (see Proposition 4.10.1)

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left| c_\beta(t) \right| \mathrm{d}t \leq 1. \tag{4.23}$$

This normalization sets the strength of $\| \boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}} \|$ to be bounded by $O(1)$.

Absolute zero $\beta = \infty$. It is instructive to consider the case of zero temperature $\beta = \infty$ and infinite time scale $\tau = \infty$. In this case, the transition weight $\gamma_\beta(\omega)$ is a step function (1 for $\omega < 0$ and 0 for $\omega > 0$) and $\hat{\boldsymbol{A}}^a(\omega)$ measures the energy difference perfectly. Thus, all heating transitions ($|E\rangle \to |E + \omega\rangle$ for $\omega > 0$) are forbidden, and all cooling transitions ($|E\rangle \to |E + \omega\rangle$ for $\omega < 0$) will remain. Hence, in the case when $\beta = \tau = \infty$, the thermal Lindbladian only lowers the energy. This matches our physical intuition that a zero-temperature bath only absorbs energy from the system.

Multiple jumps. The thermal Lindbladian $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ considers merely a single jump operator $\boldsymbol{A}^a$ in the system-bath interaction. When there are multiple jump operators, the total thermal Lindbladian is a weighted sum of the individual thermal Lindbladian $\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$,

$$\mathcal{L}^{\beta,\tau,\boldsymbol{H}} = \sum_{a=1}^{m} \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}},$$

where $\alpha_a \geq 0$ is a nonnegative weight.

Again, the interaction strength vector $\boldsymbol{\alpha} \in \mathbb{R}_{\geq 0}^m$ weights the contribution of each thermal Lindbladian. Thus, the total equation of motion under multiple jumps reads

$$\begin{aligned} \frac{d\boldsymbol{\rho}}{dt} &= -\mathrm{i}[\boldsymbol{H}, \boldsymbol{\rho}] + \mathcal{L}^{\beta,\tau,\boldsymbol{H}}(\boldsymbol{\rho}) \\ &= -\mathrm{i}\left[\boldsymbol{H} + \sum_{a=1}^{m} \alpha_a \boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}}, \boldsymbol{\rho} \right] + \sum_{a=1}^{m} \alpha_a \mathcal{D}_a^{\beta,\tau,\boldsymbol{H}}(\boldsymbol{\rho}), \end{aligned} \tag{4.24}$$

which consists of a coherent part and a purely dissipative part.

Calculation for normalization of $c_\beta(t)$. We now give a supplemental calculation that shows our choice of $\gamma_\beta(\omega)$ in Equation 4.21 satisfies the condition in Equation 4.23.

Proposition 4.10.1. *For*

$$\hat{f}(\omega) := \frac{e^{-\omega^2/2\Lambda_0^2}}{1+e^{\beta\omega}},$$

we have that

$$\frac{1}{\sqrt{2\pi}}\|f\|_1 \leq 2 + \ln(1+\beta\Lambda_0).$$

Proof. We want to bound the 1-norm of $f(t)$ in the time domain. To do so, we bound the moments in the time domains

$$\begin{aligned}
\sqrt{2\pi}\|f\|_\infty \leq \|\hat{f}\|_1 &\leq \int_{-\infty}^{\infty} e^{-\omega^2/2\Lambda_0^2} d\omega = \Lambda_0\sqrt{2\pi}. \\
\sqrt{2\pi}\|tf(t)\|_\infty &\leq \|\frac{d}{d\omega}\hat{f}\|_1 = 2\cdot\|\hat{f}\|_\infty \leq 2. \\
\sqrt{2\pi}\|t^2 f(t)\|_\infty &\leq \|\frac{d^2}{d\omega^2}\hat{f}\|_1 \leq 4\cdot\|\frac{d}{d\omega}\hat{f}\|_\infty \leq T.
\end{aligned}$$

The second line uses the fact that $\hat{f}$ is increasing and then decreasing (from $-\infty$ to ∞). The third line evaluates the derivative

$$\begin{aligned}
4\left|\frac{d}{d\omega}\left(\frac{e^{-\omega^2/2\Lambda_0^2}}{1+e^{\beta\omega}}\right)\right| &= 4\left|\frac{-e^{-\omega^2/2\Lambda_0^2}\omega/\Lambda_0^2}{1+e^{\beta\omega}} - \frac{e^{-\omega^2/2\Lambda_0^2}\beta e^{\beta\omega}}{(1+e^{\beta\omega})^2}\right| \\
&\leq 4\left(\frac{1}{\sqrt{e}\Lambda_0} + \beta\right) =: T.
\end{aligned}$$

Thus, we may partition into three integrals to optimize the bound

$$\begin{aligned}
\|f\|_1 &= \left(\int_{|t|\leq\Lambda_0^{-1}} + \int_{T\geq|t|\geq\Lambda_0^{-1}} + \int_{|t|\geq T}\right)|f(t)|dt \\
&\leq \int_{|t|\leq\Lambda_0^{-1}} \Lambda_0 dt + \frac{1}{\sqrt{2\pi}}\int_{T\geq|t|\geq\Lambda_0^{-1}} \frac{2}{|t|}dt + \frac{1}{\sqrt{2\pi}}\int_{|t|\geq T}\frac{T}{t^2}dt \\
&\leq 2 + \frac{4}{\sqrt{2\pi}}\ln(\Lambda_0 T) + \frac{2}{\sqrt{2\pi}} \\
&\leq \frac{2+2\sqrt{2\pi}+4\ln(\frac{4}{\sqrt{e}}+4\beta\Lambda_0)}{\sqrt{2\pi}} \\
&\leq \frac{2+2\sqrt{2\pi}+8\ln(2)+4\ln(1+\beta\Lambda_0)}{\sqrt{2\pi}} \leq \sqrt{2\pi}(2+\ln(1+\beta\Lambda_0)),
\end{aligned}$$

where in the last line, we used $1/\sqrt{e} \leq 1$ among other numerical bounds. ∎

Properties of thermal Lindbladians

From the exact forms of the thermal Lindbladians, we have the following propositions.

Proposition 4.10.2 (Norm for the dissipative part [39])**.** *Any purely dissipative Lindbladian* $\sum_a \mathcal{D}_a^{\beta,\tau,\boldsymbol{H}}$ *defined in Eq.* (4.18) *for any set of jump operators* $\{\boldsymbol{A}^a\}_{a=1}^m$ *and any transition weight satisfying Eq.* (4.19) *have bounded superoperator norms*

$$\left\|\sum_{a=1}^m \alpha_a \mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}}\right\|_{\infty-\infty} = \left\|\sum_{a=1}^m \alpha_a \mathcal{D}_a^{\beta,\tau,\boldsymbol{H}}\right\|_{1-1} \leq 2\|\sum_{a=1}^m \alpha_a \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|.$$

The first equality is the duality between the $1-1$ *and* $\infty-\infty$ *superoperator norms.*

Proposition 4.10.3 (Properties of the Lamb-shift term [39])**.** *The sum of Lamb-shift term* (4.22) *for any set of jump operators* $\{\boldsymbol{A}^a\}_{a=1}^m$ *under a normalized bath correlation function* $c_\beta(t)$ *given by Eq.* (4.23) *satisfies that*[7]

$$\left\|\sum_{a=1}^m \alpha_a \boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}}\right\| \leq \frac{1}{2}\|\sum_{a=1}^m \alpha_a \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|$$
$$\|\sum_{a=1}^m \alpha_a [\boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}}, \boldsymbol{H}]\| \leq O\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}}\|\sum_{a=1}^m \alpha_a \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\right).$$

For large enough τ*, the Lamb-shift term almost commutes with the Hamiltonian.*

From Prop. 4.10.2 and Prop. 4.10.3, we have the following norm bound for thermal Lindbladians.

Proposition 4.10.4 (Norm of thermal Lindbladians)**.** *Given a Hamiltonian* $\boldsymbol{H}$*, an inverse temperature* $\beta \geq 0$*, a time scale* $\tau \geq 0$*,* m *local jump operators* $\{\boldsymbol{A}^a\}_{a=1}^m$*, a transition weight* $\gamma_\beta(\omega)$ *satisfying Eq.* (4.19)*, a normalized bath correlation function* $c_\beta(t)$ *satisfying Eq.* (4.23)*. The associated thermal Lindbladian* $\sum_{a=1}^m \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ *has bounded superoperator norms*

$$\left\|\sum_{a=1}^m \alpha_a \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}\right\|_{\infty-\infty} = \left\|\sum_{a=1}^m \alpha_a \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}\right\|_{1-1} \leq 3\|\sum_{a=1}^m \alpha_a \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|,$$

which is controlled by the interaction strength vector $\boldsymbol{\alpha}$ *under the normalization of* $\boldsymbol{A}^a$ *in Eq.* (4.2).

[7] Implicitly, the Lamb-shift term has units of energy yet do not scale with $\|\boldsymbol{H}\|$.

Algorithmic primitives for simulating thermal Lindbladians

In this subsection, we review existing algorithmic primitives for simulating thermal Lindbladians [39], estimating energy and expectation value of observables using block-encoding and quantum singular value transform (QSVT). See [126] for a tutorial on block encoding and QSVT. We begin with a definition of a block-encoding for Hermitian matrices, i.e., observables.

Definition 4.10.1 (Block-encoding for Hermitian matrices)**.** *We say that a unitary $\boldsymbol{U}$ is a block-encoding for a Hermitian matrix $\boldsymbol{O}$ if*

$$(\langle 0^d| \otimes \boldsymbol{I}) \cdot \boldsymbol{U} \cdot (|0^d\rangle \otimes \boldsymbol{I}) = \boldsymbol{O} \quad \textit{for} \quad d \in \mathbb{Z}^+.$$

Recall the following result stating that expectation values can be estimated using block-encoding.

Proposition 4.10.5 (Measuring observable using block-encoding)**.** *Given a block-encoding $\boldsymbol{U_O}$ for a Hermitian matrix $\boldsymbol{O}$ and samples of a state $\boldsymbol{\rho}$. One could estimate* $\mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho})$ *to small error* $0 < \epsilon < 0.5$ *using only* $\tilde{O}(1/\epsilon^2)$ *queries to the unitary $\boldsymbol{U_O}$.*

Proof. Consider $|0\rangle\langle 0| \otimes \boldsymbol{\rho}$ and apply the Hadamard test to sample $\mathrm{tr}[\boldsymbol{U}|0\rangle\langle 0| \otimes \boldsymbol{\rho}] = \mathrm{tr}[\boldsymbol{O}\boldsymbol{\rho}]$[8]. ∎

Linear combinations of unitaries allow us to make efficient block-encodings of Hamiltonians presented as a sum of local terms. This fact results in the following proposition.

Proposition 4.10.6 (Block-encoding for Hamiltonian; see [19, 45, 126])**.** *Any n-qubit Hamiltonian $\boldsymbol{H}$ has an efficient block-encoding $\boldsymbol{U}_{\boldsymbol{H}/\lambda_1}$ for some scalar λ_1 being the 1-norm of Pauli expansion coefficients.*

From [39], we have the following for the Lamb-shift term $\boldsymbol{H}_{LS}$ from Eq. (4.22). Conveniently, the Lamb-shift term is already normalized (Proposition 4.10.3).

Proposition 4.10.7 (Block-encoding for Lamb-shift term; see [39])**.** *The Lamb-shift term $\boldsymbol{H}_{LS}$ has an efficient block-encoding $\boldsymbol{U}_{LS}$.*

We define the block-encoding for a Lindbladian without the coherent commutator term $-\mathrm{i}[\boldsymbol{H}, \boldsymbol{\rho}]$.

[8] We thank Yu Tong for discussions on this argument.

Definition 4.10.2 (Block-encoding for Lindblad operators [39]). *Given a purely irreversible Lindbladian*

$$\mathcal{L}[\rho] := \sum_{j\in J}\left(\boldsymbol{L}_j\rho\boldsymbol{L}_j^\dagger - \frac{1}{2}\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\rho - \frac{1}{2}\rho\boldsymbol{L}_j^\dagger\boldsymbol{L}_j\right),$$

we say that a unitary $\boldsymbol{U}$ is a block-encoding for Lindblad operators $\{\boldsymbol{L}_j\}_{j\in J}$ if [9]

$$(\langle 0^b| \otimes \boldsymbol{I})\cdot\boldsymbol{U}\cdot(|0^c\rangle\otimes\boldsymbol{I}) = \sum_{j\in J}|j\rangle\otimes\boldsymbol{L}_j \quad \textit{for} \quad b\le c\in\mathbb{Z}^+.$$

Theorem 4.10.1 (Linear-time Lindbladian simulation [39]). *Suppose the jumps $\boldsymbol{A}^a$ can be block-encoded by a unitary $\boldsymbol{V}_{jump}$ using $c\in\mathbb{Z}$ ancillas qubits. Then, we can simulate the map $e^{t\mathcal{L}}$ for (4.24) to $\epsilon\le 1/2$ precision in the diamond norm using*

$$\begin{aligned}
&\tilde{O}((c+1)) && \textit{resettable ancilla},\\
&\tilde{O}((t+1)\tau) && \textit{controlled Hamiltonian simulation time},\\
&\tilde{O}((t+1)(c+1)) && \textit{other two-qubit gates},\\
\textit{and}\quad &\tilde{O}(t+1) && \textit{queries to } \boldsymbol{W},\ \boldsymbol{prep}_{c_\beta(t)}, \boldsymbol{Prep}'_{c_\beta(\bar{t})},\ \textit{and } \boldsymbol{V}_{jump}
\end{aligned}$$

where $\tilde{O}(\cdot)$ absorbed poly-logarithmic dependences on $t, \|\boldsymbol{H}\|, \epsilon, \tau, \beta$. Furthermore, a block-encoding of the purely irreversible Lindbladian $\mathcal{D}^{\beta,\tau,\boldsymbol{H}}$ with discretized frequency labels can be implemented efficiently.

The above uses the following circuit components required for implementation: the controlled Hamiltonian simulation

$$\sum_{\bar{t}\in S_{t_0}}|\bar{t}\rangle\langle\bar{t}|\otimes e^{\pm i\bar{t}\boldsymbol{H}},$$

the unitary gates for preparing the bath correlation function in superposition

$$\boldsymbol{Prep}_{c_\beta(\bar{t})}:|\bar{0}\rangle\to\sum_{\bar{t}\in S_{t_0}}\sqrt{|c_\beta(\bar{t})|}|\bar{t}\rangle \quad\text{and}\quad \boldsymbol{Prep}'_{c_\beta(\bar{t})}:|\bar{0}\rangle\to\sum_{\bar{t}\in S_{t_0}}\frac{c_\beta(\bar{t})}{\sqrt{|c_\beta(\bar{t})|}}|\bar{t}\rangle,$$

and the controlled rotation for transition weights

$$\boldsymbol{W} := \sum_{\bar{\omega}\in S_{\omega_0}}\begin{pmatrix}\sqrt{\gamma(\bar{\omega})} & -\sqrt{1-\gamma(\bar{\omega})}\\ \sqrt{1-\gamma(\bar{\omega})} & \sqrt{\gamma(\bar{\omega})}\end{pmatrix}\otimes|\bar{\omega}\rangle\langle\bar{\omega}|.$$

[9] In the first register, we could use any orthonormal basis, sticking to computational basis elements $|j\rangle$ is just for ease of presentation. Intuitively one can think about b as the number of ancilla qubits used for implementing the operators $\boldsymbol{L}_j$, while typically $a-b\approx\log|J|$.

Indeed, the above implementation uses discrete labels for the time $\bar{t} \in S_{t_0}$ and frequencies $\bar{\omega} \in S_{\omega_0}$ corresponding to $\mathrm{d}t$ and $\mathrm{d}\omega$; these dominate the ancilla use. For conceptual simplicity, we focus on the continuous integral everywhere else and emphasize that the discretization is merely for implementation and introduces a negligible error; see [39].

The controlled Hamiltonian simulation can be implemented efficiently for any n-qubit local Hamiltonian $\boldsymbol{H}$ [19, 45, 126]. The other operations $\boldsymbol{W}$, $\boldsymbol{prep}_{c_\beta(t)}$,$\boldsymbol{Prep}'_{c_\beta(\bar{t})}$ can all be implemented efficiently [39] with the physically-motivated choice considered in Appendix 4.10.

Proposition 4.10.8 (Gradient of an observable under Lindbladian evolution; adapted from [39])**.** *Given block-encodings $\boldsymbol{U}$ for a purely irreversible Lindbladian (Def. 4.10.2) and $\boldsymbol{U_O}$ for a Hermitian observable $\boldsymbol{O}$, we get a block-encoding of*

$$\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{O} \boldsymbol{L}_j \quad \textit{via} \quad \boldsymbol{V} := (\boldsymbol{Y}_{\frac{1}{2}} \otimes \boldsymbol{U}^\dagger \otimes \boldsymbol{I}_d) \cdot \left(2|0^{b+1}\rangle\langle 0^{b+1}| \otimes \boldsymbol{I} - \boldsymbol{I}\right) \otimes \boldsymbol{U_O} \cdot (\boldsymbol{Y}_{\frac{1}{2}} \otimes \boldsymbol{U} \otimes \boldsymbol{I}_d),$$

where $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$ *and* $\boldsymbol{Y}_{\frac{1}{2}} := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

Proof. We calculate

$$\begin{aligned}
&(\langle 0^{c+1}| \otimes \boldsymbol{I} \otimes \langle 0^d|) \cdot \boldsymbol{V} \cdot (|0^{c+1}\rangle \otimes \boldsymbol{I} \otimes |0^d\rangle) \\
&= \left(\langle -| \otimes (\langle 0^c| \otimes \boldsymbol{I})\boldsymbol{U}^\dagger \otimes \langle 0^d|\right) \cdot \left(2|0^{b+1}\rangle\langle 0^{b+1}| \otimes \boldsymbol{I} - \boldsymbol{I}\right) \otimes \boldsymbol{U_O} \cdot \left(|+\rangle \otimes \boldsymbol{U}(|0^c\rangle \otimes \boldsymbol{I}) \otimes |0^d\rangle\right) \\
&= \left(\langle -| \otimes (\langle 0^c| \otimes \boldsymbol{I})\boldsymbol{U}^\dagger \otimes \langle 0^d|\right) \cdot \left(2|0^{b+1}\rangle\langle 0^{b+1}| \otimes \boldsymbol{I}\right) \otimes \boldsymbol{U_O} \cdot \left(|+\rangle \otimes \boldsymbol{U}(|0^c\rangle \otimes \boldsymbol{I}) \otimes |0^d\rangle\right) \\
&= (\langle 0^c| \otimes \boldsymbol{I}) \cdot \boldsymbol{U}^\dagger \cdot (|0^b\rangle\langle 0^b| \otimes \boldsymbol{I}) \otimes \boldsymbol{O} \cdot (|0^c\rangle \otimes \boldsymbol{I}) \\
&= \left(\sum_{j\in J} \langle j| \otimes \boldsymbol{L}_j^\dagger\right) \boldsymbol{I} \otimes \boldsymbol{O} \left(\sum_{j'\in J} |j'\rangle \otimes \boldsymbol{L}_{j'}\right) = \sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{O} \boldsymbol{L}_j.
\end{aligned}$$

■

Corollary 4.10.1 (Block-encoding the gradient of the Hamiltonian)**.** *Given a block-encoding for a purely irreversible Lindbladian $\mathcal{L}$ and a Hamiltonian $\boldsymbol{H}$, there is an efficient block-encoding for*

$$\frac{1}{2}\mathcal{L}^\dagger[\boldsymbol{H}],$$

which is a Hermitian operator corresponding to the gradient of $\boldsymbol{H}$ under $\mathcal{L}$.

Proof. Apply Proposition 4.10.8 for Lindbladian $\mathcal{L}$, Hermitian observables $\boldsymbol{H}$ and $\boldsymbol{I}$ to obtain block-encodings for $\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{H} \boldsymbol{L}_j$ and $\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j$. Then, use quantum singular value transform (QSVT) for products and sums of block-encoding to obtain the block-encoding for

$$\frac{1}{2}\mathcal{L}^\dagger[\boldsymbol{H}] = \frac{1}{2}\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{H} \boldsymbol{L}_j - \frac{1}{4}\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j \boldsymbol{H} - \frac{1}{4}\boldsymbol{H}\sum_{j\in J} \boldsymbol{L}_j^\dagger \boldsymbol{L}_j$$

at high precision. ∎

From all of the above propositions, corollaries, and theorems, we can obtain the following.

Lemma 4.10.1 (Measuring energy gradient)**.** *Given an n-qubit Hamiltonian $\boldsymbol{H}$, inverse temperature $\beta \geq 0$, time scale $\tau \geq 0$, samples of an n-qubit state $\boldsymbol{\rho}$, a thermal Lindbladian $\mathcal{L}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}$ from Eq. (4.24). The energy gradient*

$$\mathrm{tr}(\boldsymbol{H}\mathcal{L}^{\beta,\tau,\boldsymbol{H}}(\boldsymbol{\rho})) = \mathrm{tr}(\mathcal{L}^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\boldsymbol{\rho})$$

can be estimated to error ϵ using time and samples of $\boldsymbol{\rho}$ polynomial in $n, 1/\epsilon, \|\boldsymbol{H}\|, \beta, \tau$.

Proof. From the form of thermal Lindbladians (4.24) and dropping the scripts $\mathcal{L}^{\beta,\tau,\boldsymbol{H}} = \mathcal{L}, \boldsymbol{H}_{LS}^{\beta,\tau,\boldsymbol{H}} = \boldsymbol{H}_{LS}, \mathcal{D}^{\beta,\tau,\boldsymbol{H}} = \mathcal{D}$ we have

$$\mathcal{L}^\dagger(\boldsymbol{H}) = \mathrm{i}[\boldsymbol{H}_{LS}, \boldsymbol{H}] + \mathcal{D}^\dagger(\boldsymbol{H}).$$

Our goal is to create the block-encoding for $\mathcal{L}^\dagger(\boldsymbol{H})$. First, we use quantum singular value transform (QSVT) for products and sums of block-encoding to obtain the block-encoding for $\mathrm{i}[\boldsymbol{H}_{LS}, \boldsymbol{H}]$ from block-encoding for $\boldsymbol{H}$ and $\boldsymbol{H}_{LS}$ in Propositions 4.10.6 and 4.10.7. Next, using the block-encoding for the purely irreversible Lindbladian $\mathcal{D}$ from Theorem 4.10.1 and the block-encoding for $\boldsymbol{H}$, we can apply Corollary 4.10.1 to obtain efficient block-encoding for $\mathcal{D}^\dagger(\boldsymbol{H})$. To obtain the block-encoding for $\mathcal{L}^\dagger(\boldsymbol{H})$, we use QSVT for sums of block-encoding to add $\mathrm{i}[\boldsymbol{H}_{LS}, \boldsymbol{H}]$ and $\mathcal{D}^\dagger(\boldsymbol{H})$. Finally, using Prop. 4.10.5, we can estimate $\mathrm{tr}(\mathcal{L}^\dagger(\boldsymbol{H})\boldsymbol{\rho})$ efficiently. All the above QSVT manipulations operate at high precision, and the discrete Fourier transform well-approximates the continuum at poly-logarithmic costs [39]. ∎

4.11 Appendix:A polynomial-time quantum algorithm for finding a local minimum under thermal perturbations (Proof of Theorem 4.9.1)

In this appendix, we present the proof of Theorem 4.9.1 by giving a polynomial-time quantum algorithm for finding local minima under thermal perturbations. We

refer to the efficient quantum algorithm as *Quantum thermal gradient descent* as the algorithm performs gradient descent using thermal Lindbladians induced by a heat bath. The algorithm uses the properties of thermal Lindbladians presented in Appendix 4.10.

Cooling by gradient descent

The central idea of quantum thermal gradient descent is the following. When we are not at a local minimum under thermal perturbations, the negative energy gradient will be sufficiently large, and we can decrease the energy by following a direction with a negative energy gradient. This is characterized by the following lemma. We will use this lemma to design the gradient descent algorithm for finding a local minimum.

Lemma 4.11.1 (Cooling by gradient descent)**.** *Given parameters* $0 < \tilde{\epsilon} < 0.5$, $B \geq 1$, $\beta, \tau \geq 0$, *an n-qubit Hamiltonian* $\boldsymbol{H}$ *with* $\|\boldsymbol{H}\|_\infty \leq B$, *and m local jump operators* $\{\boldsymbol{A}^a\}_a$. *Consider* $a = 1, \ldots, m$ *with an approximate energy gradient* g_a *satisfying*

$$\left| g_a - \mathrm{tr}\left(\boldsymbol{H} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}^{(t-1)}] \right) \right| < 0.01\tilde{\epsilon}.$$

Suppose there exist $a^* \in \{1, \ldots, m\}$ *with sufficiently negative approximate energy gradient,*

$$g_{a^*} < -0.99\tilde{\epsilon}.$$

The state after evolving $\boldsymbol{\rho}$ *along the direction* $\hat{\boldsymbol{e}}_{a^*}$ *for a small step* $s = |g_{a^*}|/(9B^2) > 0$,

$$\boldsymbol{\rho}^{(\mathrm{next})} := \exp_{\boldsymbol{\rho}}^{\beta,\tau,\boldsymbol{H},\{\boldsymbol{A}^a\}_a}(s\hat{\boldsymbol{e}}_{a^*})$$

guarantees the following energy decrease,

$$\mathrm{tr}\left(\boldsymbol{H}\boldsymbol{\rho}^{(\mathrm{next})} \right) < \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}) - \frac{\tilde{\epsilon}^2}{20B^2}.$$

Proof. From Prop. 4.8.2 on Taylor's theorem, we have the following identity

$$\mathrm{tr}\left(\boldsymbol{H}\boldsymbol{\rho}^{(\mathrm{next})} \right) = \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}) + s\,\mathrm{tr}(\boldsymbol{H}\mathcal{L}_{a^*}^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]) + \frac{s^2}{2}\mathrm{tr}(\boldsymbol{H}\mathcal{L}_{a^*}^{\beta,\tau,\boldsymbol{H}}[\mathcal{L}_{a^*}^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\sigma}]])$$

for some n-qubit state $\boldsymbol{\sigma}$. We will separately control the linear term and the quadratic term.

Linear term. From the definition of the energy gradient vector $\boldsymbol{\nabla}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H}, \boldsymbol{\rho})$, we have

$$\mathrm{tr}(\boldsymbol{H}\mathcal{L}_{a^*}^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\rho}]) < g_{a^*} + 0.01\tilde{\epsilon} < \frac{98}{99}g_{a^*} = -\frac{98}{99}|g_{a^*}|.$$

The second inequality follows from $g_{a^*} < -0.99\tilde{\epsilon}$, hence $0.01\tilde{\epsilon} < -(1/99)g_{a^*}$.

Quadratic term. We can bound the quadratic term as follows:

$$\frac{1}{2}\operatorname{tr}(\boldsymbol{H}\mathcal{L}_{a^*}^{\beta,\tau,\boldsymbol{H}}[\mathcal{L}_{a^*}^{\beta,\tau,\boldsymbol{H}}[\boldsymbol{\sigma}]]) \le \frac{1}{2}\left\|\mathcal{L}_{a^*}^{\dagger\beta,\tau,\boldsymbol{H}}\right\|_{\infty-\infty}\|\mathcal{L}_{a^*}^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\|_\infty.$$

From Prop. 4.8.3 and Prop. 4.10.4 that bounds the norm of these objects, we have

$$\frac{1}{2}\left\|\mathcal{L}_{a}^{\dagger\beta,\tau,\boldsymbol{H}}\right\|_{\infty-\infty}\|\mathcal{L}_{a'}^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\|_\infty \le 4.5\|\boldsymbol{H}\|_\infty^2 \le 4.5B^2.$$

Combining the linear and quadratic terms with $s = |g_{a^*}|/(9B^2) > 0$, we have

$$\operatorname{tr}\left(\boldsymbol{H}\boldsymbol{\rho}^{(\text{next})}\right) \le \operatorname{tr}(\boldsymbol{H}\boldsymbol{\rho}) - \left(\frac{98}{99\times 9} - \frac{1}{18}\right)\frac{|g_{a^*}|^2}{B^2} < \operatorname{tr}(\boldsymbol{H}\boldsymbol{\rho}) - 0.054\frac{|g_{a^*}|^2}{B^2}.$$

We can use $|g_{a^*}|^2 > 0.99^2\tilde{\epsilon}^2$ to obtain the desired claim. ∎

Quantum thermal gradient descent

Given error $\epsilon = 1/\text{poly}(n)$, norm bound $B = \text{poly}(n)$, inverse temperature $0 \le \beta \le \text{poly}(n)$, time scale $\tau = \text{poly}(n)$, an n-qubit local Hamiltonian $\boldsymbol{H}$ with $\|\boldsymbol{H}\|_\infty \le B$, m local jump operators $\{\boldsymbol{A}^a\}_a$ with $m = \text{poly}(n)$, and a local observable $\boldsymbol{O}$ with $\|\boldsymbol{O}\|_\infty \le 1$.

We consider a coordinate-wise gradient descent algorithm that implements the following. The initial state $\boldsymbol{\rho}^{(0)}$ is arbitrary as long as copies of the state can be prepared on the quantum computer. For example, we can set $\boldsymbol{\rho}^{(0)}$ to be the maximally mixed state $\frac{\boldsymbol{I}}{2^n}$. The total number of steps is

$$T := \frac{42B^3}{\epsilon^2}. \tag{4.27}$$

For each time step t from 1 to T, the algorithm does the following.

1. For each direction $a = 1, \ldots, m$, estimate an approximate energy gradient $g_a^{(t)}$ satisfying
$$\left|g_a^{(t)} - \operatorname{tr}\left(\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\boldsymbol{\rho}^{(t-1)}\right)\right| < 0.0099\epsilon. \tag{4.28}$$
The energy gradient can be estimated efficiently using Lemma 4.10.1 given copies of $\boldsymbol{\rho}^{(t-1)}$ prepared through Eq. (4.29) and Theorem 4.10.1. From the bound on energy gradients in Prop. 4.8.3, we have $|g_a^{(t)}| \le 3B + 0.0099\epsilon$. If $g_a^{(t)} < -0.99\epsilon$, set $a^{(t)} := a$ and *terminate* the for-loop over a.

2. If $a^{(t)}$ is not found, set $\boldsymbol{\rho}^{(T)} := \boldsymbol{\rho}^{(t-1)}$ and *terminate* the for-loop over t. Otherwise evolve $\boldsymbol{\rho}^{(t-1)}$ under the direction $\hat{\boldsymbol{e}}_{a^{(t)}}$ for a small step $s^{(t)} :=$

$|g_a^{(t)}|/(9B^2)$,

$$\boldsymbol{\rho}^{(t)} := \exp\left(s^{(t)}\mathcal{L}_{a^{(t)}}^{\beta,\tau,\boldsymbol{H}}\right)\left(\boldsymbol{\rho}^{(t-1)}\right) = \prod_{t'=1}^{t}\exp\left(s^{(t)}\mathcal{L}_{a^{(t)}}^{\beta,\tau,\boldsymbol{H}}\right)\left(\boldsymbol{\rho}^{(0)}\right). \tag{4.29}$$

Because $0 \leq s^{(t)} \leq 1/(2B)$, a single copy of $\boldsymbol{\rho}^{(t)}$ can be prepared in polynomial-time using the thermal Lindbladian simulation algorithm in [39]; see Theorem 4.10.1.

We will show that the state $\boldsymbol{\rho}^{(T)}$ created by the gradient descent algorithm is an ϵ-approximate local minimum of $\boldsymbol{H}$ under thermal perturbations. Furthermore, using the thermal Lindbladian simulation algorithm, a quantum machine can efficiently create many copies of $\boldsymbol{\rho}^{(T)}$.

Proof of Theorem 4.9.1

The central idea in the proof of Theorem 4.9.1 is the following lemma. The lemma combines the key results characterizing local minima in Appendix 4.8.

Lemma 4.11.2 (Gradient descent finds a local minimum)**.** *$\boldsymbol{\rho}^{(T)}$ from Eq. (4.29) is an ϵ-approximate local minimum of $\boldsymbol{H}$ under thermal perturbations with inverse temperature β, time scale τ, and system-bath interactions generated by $\{\boldsymbol{A}^a\}_a$.*

Proof. Suppose the algorithm terminates at some time step $t < T$, then $g_a^{(t)} \geq -0.99\epsilon$. From Eq. (4.28), we have $\mathrm{tr}\left(\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}(\boldsymbol{H})\boldsymbol{\rho}^{(t-1)}\right) \geq -0.9999\epsilon$. Hence,

$$\|\boldsymbol{\nabla}^{-}_{\beta,\tau,\{\boldsymbol{A}^a\}_a}(\boldsymbol{H},\boldsymbol{\rho}^{(t-1)})\|_\infty \leq 0.9999\epsilon < \epsilon.$$

From the sufficient condition for local minima given in Lemma 4.8.1, we have $\boldsymbol{\rho}^{(T)} = \boldsymbol{\rho}^{(t-1)}$ is an ϵ-approximate local minimum $\boldsymbol{\rho}$ of the n-qubit Hamiltonian $\boldsymbol{H}$ under thermal perturbations.

We now show by contradiction that the algorithm must terminate early. Assume that the algorithm did not terminate early. Then, we can use Lemma 4.11.1 with $\tilde{\epsilon} = 0.99\epsilon$ for cooling by gradient descent to obtain

$$\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}^{(T)}) \leq \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}^{(T-1)}) - \frac{0.99^2\epsilon^2}{20B^2} \leq \ldots \leq \mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}^{(0)}) - \frac{0.99^2\epsilon^2}{20B^2}T \leq \|\boldsymbol{H}\|_\infty - \frac{0.99^2\epsilon^2}{20B^2}T.$$

From the definition of T in Eq. (4.27) and $\|\boldsymbol{H}\|_\infty \leq B$, we have

$$\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}^{(T)}) \leq \|\boldsymbol{H}\|_\infty - \frac{0.99^2\epsilon^2}{20B^2}\frac{42B^3}{\epsilon^2} \leq \|\boldsymbol{H}\|_\infty - 2.05B \leq -1.05B.$$

At the same time, because $\|\boldsymbol{\rho}^{(T)}\|_1 = 1$,

$$\mathrm{tr}(\boldsymbol{H}\boldsymbol{\rho}^{(T)}) \geq -\|\boldsymbol{H}\|_\infty \geq -B.$$

This is a contradiction. Hence the algorithm must terminate early. ■

The polynomial-time quantum algorithm for establishing Theorem 4.9.1 is as follows. The algorithm runs quantum thermal gradient descent to find a local minimum $\boldsymbol{\rho}^{(T)}$ of $\boldsymbol{H}$ under thermal perturbations. Recall that B is the upper bound on $\|\boldsymbol{H}\|_\infty$, and is equal to $\mathrm{poly}(n)$, and $1/\epsilon = \mathrm{poly}(n)$. Because every step can be done in polynomial time, and there are at most $T = 42B^3/\epsilon^2 = \mathrm{poly}(n)$ time steps, quantum thermal gradient descent runs in time polynomial in n.

Now, given any observable $\boldsymbol{O}$. The quantum algorithm prepares $O(1/\epsilon^2) = \mathrm{poly}(n)$ copies of $\boldsymbol{\rho}^{(T)}$ in $\mathrm{poly}(n)$ time, then measures $\boldsymbol{O}$ on the $O(1/\epsilon^2)$ copies of $\boldsymbol{\rho}^{(T)}$ to estimate $\mathrm{tr}(\boldsymbol{O}\boldsymbol{\rho}^{(T)})$ to ϵ error. This concludes the proof of Theorem 4.9.1.

4.12 Appendix:Characterizing energy gradients in low-temperature heat bath

Recall from Appendix 4.8 on certifying Hamiltonians without suboptimal local minima, if there exists $\boldsymbol{\alpha} \in \mathbb{R}^m_{\geq 0}$ with $\|\boldsymbol{\alpha}\|_1 = 1$, such that the negative gradient condition holds,

$$-\sum_a \alpha_a \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{H}] \geq \frac{2\epsilon}{\delta}(\boldsymbol{I} - \boldsymbol{P}_G(\boldsymbol{H})) - \epsilon\boldsymbol{I},$$

then any ϵ-approximate local minimum $\boldsymbol{\rho}$ of the n-qubit Hamiltonian $\boldsymbol{H}$ under thermal perturbations is an exact global minimum of $\boldsymbol{H}$ with failure probability $\leq \delta$, i.e., $\mathrm{tr}(\boldsymbol{P}_G(\boldsymbol{H})\boldsymbol{\rho}) \geq 1-\delta$, where $\boldsymbol{P}_G(\boldsymbol{H})$ is the projection onto the ground state space. To understand when the above condition holds, it is imperative to characterize the energy gradients, $\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{H}]$.

In this appendix, we present various lemmas and theorems characterizing the energy gradients, which will be used in our proof of Theorem 4.9.2 in section 4.14 for showing that a certain family of Hamiltonians has no suboptimal local minima. We remark that the proofs of many formal statements in this appendix require concepts and results that won't be shown till later in Appendices 4.15 and 4.16, and we recommend the first-time reader to freely skip the proofs and return later.

For simplicity, we will focus on the nonnegative vector $\boldsymbol{\alpha}$ being uniform over a subset S for the remaining appendices. We will show that this is sufficient for our purposes

even though having the ability to choose $\boldsymbol{\alpha}$ is more powerful. We define the following Lindbladian with uniform weights over a subset $S \subseteq \{1, \ldots, m\}$:

$$\mathcal{L} := \sum_{a \in S} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}. \tag{4.30}$$

Recall from Appendix 4.10 that each $\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}$ corresponds to a jump operator $\boldsymbol{A}^a$ satisfying the normalization condition $\|\boldsymbol{A}^a\|_\infty \leq 1$. If we let $r := 2m\epsilon/\delta, \epsilon' = m\epsilon$ and $S = \{1, \ldots, m\}$, then the negative gradient condition becomes

$$\text{(negative gradient condition)}: \quad -\mathcal{L}^\dagger[\boldsymbol{H}] \geq r(\boldsymbol{I} - \boldsymbol{P}_G) - \epsilon'\boldsymbol{I}. \tag{4.31}$$

This will be the central inequality we would like to establish for the remaining appendices. Throughout the proofs, we will consider different subsets S and show a relation similar to Eq. (4.31) for subset S.

Basic properties of energy gradients in low-temperature bath

We show a few basic properties of energy gradients under a low-temperature, long-time-scale bath. First, we show that the energy gradient, at large β (i.e., low temperatures), is negative semi-definite up to controllable error. Intuitively, this can be seen from the KMS condition in Eq. (4.19), $\gamma_\beta(\omega) = \gamma_\beta(-\omega)\mathrm{e}^{-\beta\omega}$: the heating transition is suppressed by the Boltzmann weight, allowing energy to increase by $\omega \sim \beta^{-1}$. Another source of error is the uncertainty in energy τ^{-1}.

Lemma 4.12.1 (Almost negative gradients)**.** *Consider the thermal Lindbladian $\mathcal{L} = \sum_{a\in S} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ with jump operators $\{\boldsymbol{A}^a\}_{a\in S}$ where $\|\boldsymbol{A}^a\| \leq 1$, and $\gamma_\beta(\omega)$ satisfying Eq. (4.19). Then,*

$$\mathcal{L}^\dagger[\boldsymbol{H}] \leq O\left(|S|\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta}\right)\right) \cdot \boldsymbol{I}.$$

Proof. Rewrite the energy gradient with an error controlled by Proposition 4.10.3 and Lemma 4.16.1 gives

$$\begin{aligned}
\mathcal{L}^\dagger[\boldsymbol{H}] &\approx \sum_{a\in S} \int_{-\infty}^{\infty} \gamma_\beta(\omega)\omega \hat{\boldsymbol{A}}^a(\omega)^\dagger \hat{\boldsymbol{A}}^a(\omega) \mathrm{d}\omega \\
&= \sum_{a\in S} \int_{0}^{\infty} \gamma_\beta(\omega)\omega \hat{\boldsymbol{A}}^a(\omega)^\dagger \hat{\boldsymbol{A}}^a(\omega) \mathrm{d}\omega + \sum_{a\in S} \int_{-\infty}^{0} \gamma_\beta(\omega)\omega \hat{\boldsymbol{A}}^a(\omega)^\dagger \hat{\boldsymbol{A}}^a(\omega) \mathrm{d}\omega
\end{aligned}$$

and bound the positive operator

$$\|\sum_{a\in S}\int_0^\infty \gamma_\beta(\omega)\omega\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega)\mathrm{d}\omega\|_\infty \le |S|\max_{\omega\ge 0}\gamma_\beta(\omega)\omega \le \frac{|S|}{\beta}.$$

The second inequality used the tail bound in Eq. (4.20) with $\Delta = 0$. ∎

Second, we show that the energy gradient operator is nearly diagonal in the energy basis. The intuition is that for any operator $\boldsymbol{A}$, the product

$$\hat{\boldsymbol{A}}^\dagger(\omega)\hat{\boldsymbol{A}}(\omega)$$

is nearly diagonal in the energy basis for large τ.

Lemma 4.12.2 (Energy gradient is almost diagonal)**.** *In the setting of Lemma 4.12.1, assume that for any two well-isolated energy eigensubspaces $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ such that the two sets of eigenvalues have at least distance δ. Then,*

$$\|\boldsymbol{P}_1\mathcal{L}^\dagger[\boldsymbol{H}]\boldsymbol{P}_2\| \le O\left(|S|\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{\|\theta_\beta\|_\infty}{\sqrt{\delta\tau}}\right)\right),$$

where $\theta_\beta(\omega) := \gamma_\beta(\omega)\omega$.

Proof. Formally, approximate the energy gradient by dropping the Lamb-shift term (Proposition 4.10.3) and applying Lemma 4.16.1,

$$\mathcal{L}^\dagger[\boldsymbol{H}] \approx \sum_{a\in S}\int_{-\infty}^\infty \gamma_\beta(\omega)\omega\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega)\mathrm{d}\omega.$$

We then apply the secular approximation for $\mu = \delta/2$ (Corollary 4.15.3) such that the transition amplitudes vanishes between the subspaces

$$\boldsymbol{P}_1\hat{\boldsymbol{S}}^a_\mu(\omega)^\dagger\hat{\boldsymbol{S}}^a_\mu(\omega)\boldsymbol{P}_2 = 0 \quad \text{for each} \quad \omega\in\mathbb{R} \quad \text{and} \quad a\in A.$$

Combining the errors in each of the approximations leads to the claimed result. ∎

Next, we show that the finite-τ Lindbladian can be approximated by the infinite-τ version under certain conditions. The latter, known as the Davies' generator [57], has a simpler form that is more amenable for analysis in some situations.

Lemma 4.12.3 (Recovering Davies' generator)**.** *Consider the dissipative part of the thermal Lindbladian $\mathcal{D}_a^{\beta,\tau,\boldsymbol{H}}$ with the jump operators $\boldsymbol{A}^a$ where $\|\boldsymbol{A}^a\| \le 1$, and any γ_β such that $\|\gamma_\beta\|_\infty \le 1$. Suppose the Bohr-frequency gap is $\Delta_\nu(\boldsymbol{H})$, then*

$$\|\mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}} - \mathcal{D}_a^{\dagger\beta,\infty,\boldsymbol{H}}\|_{\infty-\infty} \le O\left(\max_\nu\left|\gamma_\beta(\nu) - \int_{-\infty}^\infty \gamma_\beta(\omega)\left|\hat{f}(\omega-\nu)\right|^2\mathrm{d}\omega\right| + \frac{1}{\sqrt{\Delta_\nu(\boldsymbol{H})\tau}}\right).$$

Therefore, the Bohr-frequency gap sets a timescale $\sim \Delta_\nu^{-1}$ such that the map $\mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}}$ stabilized.

Proof. It suffices to consider $\mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{O}]$ acting on arbitrary operator $\boldsymbol{O}$ such that $\|\boldsymbol{O}\| = 1$:

$$\begin{aligned}
\mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{O}] &= \int_{-\infty}^{\infty} \gamma_\beta(\omega)\left[\hat{\boldsymbol{A}}^a(\omega)^\dagger \boldsymbol{O}\hat{\boldsymbol{A}}^a(\omega) - \frac{1}{2}\{\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega), \boldsymbol{O}\}\right]\mathrm{d}\omega \\
&\overset{E_1}{\approx} \int_{-\infty}^{\infty} \gamma_\beta(\omega)\left[\hat{\boldsymbol{S}}^a(\omega)^\dagger \boldsymbol{O}\hat{\boldsymbol{S}}^a(\omega) - \frac{1}{2}\{\hat{\boldsymbol{S}}^a(\omega)^\dagger\hat{\boldsymbol{S}}^a(\omega), \boldsymbol{O}\}\right]\mathrm{d}\omega \\
&= \sum_{\nu,\nu'\in B(\boldsymbol{H})} \left(\boldsymbol{A}_{\nu'}^{a\dagger}\boldsymbol{O}\boldsymbol{A}_\nu^a - \frac{1}{2}\{\boldsymbol{A}_{\nu'}^{a\dagger}\boldsymbol{A}_\nu^a, \boldsymbol{O}\}\right)\int_{-\infty}^{\infty}\gamma_\beta(\omega)\hat{f}_\mu^*(\omega-\nu')\hat{f}_\mu(\omega-\nu)\mathrm{d}\omega \\
&= \sum_{\nu\in B(\boldsymbol{H})} \left(\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{O}\boldsymbol{A}_\nu^a - \frac{1}{2}\{\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{A}_\nu^a, \boldsymbol{O}\}\right)\int_{-\infty}^{\infty}\gamma_\beta(\omega)\left|\hat{f}_\mu(\omega-\nu)\right|^2\mathrm{d}\omega \\
&\overset{E_2}{\approx} \sum_{\nu\in B(\boldsymbol{H})} \left(\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{O}\boldsymbol{A}_\nu^a - \frac{1}{2}\{\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{A}_\nu^a, \boldsymbol{O}\}\right)\gamma_\beta(\nu).
\end{aligned}$$

The approximation errors are bounded by

$$\begin{aligned}
E_1 &\le 2\|\boldsymbol{O}\|\|\gamma_\beta\|_\infty\|\boldsymbol{A}^a\|\|f_\tau\cdot(1-\hat{s}_\mu)\|_2\|f_\tau\|_2 = O\left(\frac{1}{\sqrt{\mu\tau}}\right), \\
E_2 &\le 2\|\boldsymbol{O}\|\|\boldsymbol{A}^a\|\max_\nu\left|\gamma_\beta(\nu) - \int_{-\infty}^{\infty}\gamma_\beta(\omega)\left|\hat{f}_\mu(\omega-\nu)\right|^2\mathrm{d}\omega\right| \\
&\le O\left(\max_\nu\left|\gamma_\beta(\nu) - \int_{-\infty}^{\infty}\gamma_\beta(\omega)\left|\hat{f}(\omega-\nu)\right|^2\mathrm{d}\omega\right| + \frac{\|\gamma_\beta\|_\infty}{\mu\tau}\right).
\end{aligned}$$

Combine the error bounds to conclude the proof. Note that since the bound becomes vacuous at $\mu\tau = \Omega(1)$, we have that $O(1/\sqrt{\mu\tau} + 1/(\mu\tau)) = O(1/\sqrt{\mu\tau})$. ∎

Lemma 4.12.4. *For $\gamma_\beta(\omega)$ defined in Eq. (4.21) with $\Lambda_0 = \Theta(1)$, we have that*

$$\max_{\nu\in\mathbb{R}}\left|\gamma_\beta(\nu) - \int_{-\infty}^{\infty}\gamma_\beta(\omega)\left|\hat{f}(\omega-\nu)\right|^2\mathrm{d}\omega\right| \le O\left(\frac{1+\beta}{\tau}\ln\tau\right).$$

Proof. Recall the integration-by-part trick for expectation integral

$$\int_0^\infty f(x)p(x)\mathrm{d}x = -[f(x)P(x)]_{x=0}^{x=\infty} + \int_0^\infty f'(x)P(x)\mathrm{d}x \quad \text{where} \quad P(x) := \int_x^\infty p(y)\mathrm{d}y.$$

Then,

$$\begin{aligned}
\int_{-\infty}^{\infty}\gamma_\beta(\omega)\left|\hat{f}(\omega-\nu)\right|^2\mathrm{d}\omega &= \int_{-\infty}^{\infty}\gamma_\beta(\nu+x)\left|\hat{f}(x)\right|^2\mathrm{d}x = \int_0^\infty \frac{\gamma_\beta(\nu+x)+\gamma_\beta(\nu-x)}{2}\underbrace{2\left|\hat{f}(x)\right|^2}_{=:p(x)}\mathrm{d}x \\
&= \gamma_\beta(\nu) + \int_0^\infty \frac{\gamma_\beta'(\nu+x)+\gamma_\beta'(x-\nu)}{2}P(x)\mathrm{d}x,
\end{aligned}$$

where in the last line we applied the integration-by-part and used $\gamma(\pm\infty) = 0$ and $P(0) = 1$. The error term can be bounded as follows:

$$\begin{aligned}
&\left|\int_0^\infty \frac{\gamma'_\beta(\nu+x)+\gamma'_\beta(x-\nu)}{2}P(x)\mathrm{d}x\right| \\
&= \left|\int_0^{1/\tau} \frac{\gamma'_\beta(\nu+x)+\gamma'_\beta(\nu-x)}{2}P(x)\mathrm{d}x + \int_{1/\tau}^\infty \frac{\gamma'_\beta(\nu+x)+\gamma'_\beta(\nu-x)}{2}P(x)\mathrm{d}x\right| \\
&= O(\frac{1+\beta}{\tau}) + O(\frac{1+\beta}{\tau}\ln\tau).
\end{aligned}$$

In the last line, we control the first term by $P(x) \le 1$ and noting by the product rule we have

$$\begin{aligned}
\left|\gamma'_\beta(\omega)\right| &= O\left(\left|\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{\mathrm{e}^{-\omega^2/2\Lambda_0^2}}{1+\mathrm{e}^{\beta\omega}}\right)\right|\right) = O\left(\left|-\frac{\mathrm{e}^{-\omega^2/2\Lambda_0^2}\omega/\Lambda_0^2}{1+\mathrm{e}^{\beta\omega}} - \frac{\mathrm{e}^{-\omega^2/2\Lambda_0^2}\beta\mathrm{e}^{\beta\omega}}{(1+\mathrm{e}^{\beta\omega})^2}\right|\right) \\
&\le O(\frac{1}{\Lambda_0}+\beta).
\end{aligned}$$

The second term uses the tail bound $P(x) \le \frac{4}{\pi x\tau}$ from Eq. (4.93) and that $\gamma'_\beta(\nu \pm x)$ are each rapidly decaying outside an $x \in [\mp\nu - \Lambda_0, \mp\nu + \Lambda_0]$ window so that the integral over $\frac{1}{x}\mathrm{d}x$ only contributes at most $O(\int_{1/\tau}^{\Lambda_0}\frac{1}{x}\mathrm{d}x) = O(\log(\tau\Lambda_0))$. ∎

Relating subspace and local gradients to global gradients

As a method of proof, we will often analyze a Lindbladian by its constituents, and here we present a few useful relations. First, when studying gradients, the gradient acting on a subspace is often conceptually simpler. The following lemma relates the energy gradient in a subspace and the full energy gradient. This is a direct consequence of Lemma 4.12.1 and Lemma 4.12.2 above.

Lemma 4.12.5 (Subspace gradient and global gradient)**.** *In the setting of Lemma 4.12.2, suppose $\boldsymbol{P}$ projects onto a set of eigenstates of $\boldsymbol{H}$ separated by the rest by a gap of at least δ. Then,*

$$-\mathcal{L}^\dagger[\boldsymbol{H}] \succeq -\boldsymbol{P}\mathcal{L}^\dagger[\boldsymbol{H}]\boldsymbol{P} - O\left(|S|\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\beta} + \frac{1}{\tau} + \frac{\|\theta_\beta\|_\infty}{\sqrt{\delta\tau}}\right)\right)\cdot\boldsymbol{I}.$$

Proof. Let $\boldsymbol{L} = \mathcal{L}^\dagger[\boldsymbol{H}]$. We have,

$$\boldsymbol{L} = \boldsymbol{P}\boldsymbol{L}\boldsymbol{P} + \boldsymbol{P}^\perp\boldsymbol{L}\boldsymbol{P}^\perp + \boldsymbol{P}\boldsymbol{L}\boldsymbol{P}^\perp + \boldsymbol{P}^\perp\boldsymbol{L}\boldsymbol{P}.$$

Using Lemma 4.12.1 establishing the almost negativity of the energy gradient,

$$-\boldsymbol{L} \succeq -O\left(|S|\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta}\right)\right)\boldsymbol{I},$$

we have

$$-\boldsymbol{P}^{\perp}\boldsymbol{L}\boldsymbol{P}^{\perp} \geq -O\left(|S|\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta}\right)\right)\boldsymbol{P}^{\perp}. \qquad \text{(Lemma 4.12.1)}$$

$$\|\boldsymbol{P}\boldsymbol{L}\boldsymbol{P}^{\perp} + \boldsymbol{P}^{\perp}\boldsymbol{L}\boldsymbol{P}\| \leq O\left(|S|\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{\|\theta_\beta\|_\infty}{\sqrt{\delta\tau}}\right)\right). \qquad \text{(Lemma 4.12.2)}$$

Putting the bounds together yields the advertised result. ∎

Next, we provide a lemma that gives a simplified expression of the energy gradient operator when restricted to a subspace of low-energy eigenstates.

Lemma 4.12.6 (Gradient in a subspace)**.** *In the setting of Lemma 4.12.1, suppose* $\boldsymbol{H}$ *has a subspace of low-energy eigenstates with corresponding projector* $\boldsymbol{Q}$ *that is separated from the higher energy eigenstates by an excitation gap* $\Delta_{\boldsymbol{Q}}$*. Let* $\Delta_\nu = \min_{\nu_1 \neq \nu_2 \in B(\boldsymbol{H}|_{\boldsymbol{Q}})} |\nu_1 - \nu_2|$ *be the Bohr-frequency gap of* $\boldsymbol{H}$ *restricted to the subspace. Assuming* $\Delta_\nu/2 < \Delta_{\boldsymbol{Q}}$*, then the energy gradient operator in the subspace can be approximated using*

$$\left\| \boldsymbol{Q}\mathcal{L}^{\dagger}[\boldsymbol{H}]\boldsymbol{Q} - \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H}|_{\boldsymbol{Q}})} \boldsymbol{Q}\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{Q}\boldsymbol{A}_\nu^{a}\boldsymbol{Q}\int_{-\infty}^{0}\gamma_\beta(\omega)\omega|\hat{f}_\mu(\omega-\nu)|^2\mathrm{d}\omega \right\| \leq \epsilon$$

where $\mu = \Delta_\nu/2$ *and*

$$\epsilon \leq |S|O\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta} + \frac{\|\omega\gamma_\beta(\omega)\|_\infty}{\sqrt{\Delta_\nu\tau}}\right).$$

Proof. We invoke a series of approximations to rewrite in terms of the exact Bohr frequencies on the subspace $\boldsymbol{P}_{\mathtt{I}}$.

$$\begin{aligned}
\mathcal{L}^{\dagger}[\boldsymbol{H}] &\overset{E_1}{\approx} \mathcal{D}^{\dagger}[\boldsymbol{H}] \\
&\overset{E_2}{\approx} \sum_{a\in S}\int_{-\infty}^{\infty}\gamma_\beta(\omega)\omega\boldsymbol{A}^a(\omega)^{\dagger}\boldsymbol{A}^a(\omega)\mathrm{d}\omega \\
&\overset{E_3}{\approx} \sum_{a\in S}\int_{-\infty}^{0}\gamma_\beta(\omega)\omega\boldsymbol{A}^a(\omega)^{\dagger}\boldsymbol{A}^a(\omega)\mathrm{d}\omega \\
&\overset{E_4}{\approx} \sum_{a\in S}\int_{-\infty}^{0}\gamma_\beta(\omega)\omega\boldsymbol{S}^a(\omega)^{\dagger}\boldsymbol{S}^a(\omega)\mathrm{d}\omega \\
&= \sum_{a\in S}\sum_{\nu',\nu\in B(\boldsymbol{H})}\int_{-\infty}^{0}\gamma_\beta(\omega)\omega\boldsymbol{A}_{\nu'}^{a\dagger}\boldsymbol{A}_\nu^a\hat{f}_\mu^{\,*}(\omega-\nu')\hat{f}_\mu(\omega-\nu)\mathrm{d}\omega =: \boldsymbol{X}
\end{aligned} \qquad (4.44)$$

The errors are $E_1 = O(|S|\|\boldsymbol{H}\|^{3/4}/\tau^{1/4})$, $E_2 = O(|S|/\tau)$, $E_3 = O(|S|/\beta)$, and $E_4 = O(|S| \times \|\omega\gamma_\beta(\omega)\|_\infty/\sqrt{\mu\tau})$. In particular, E_3 arises from dropping the positive integral range with error bounded by $\max_{\omega\geq 0} \omega\gamma_\beta(\omega) \leq 1/\beta$. In other words, our choice of $\gamma_\beta(\omega)$ allows the energy to increase by at most $O(1/\beta)$.

Sandwiching Eq. (4.12) with $\boldsymbol{Q}$ further simplifies the expression as it restricts to transitions in the subspace. Specifically, we have

$$\begin{aligned}\boldsymbol{Q}\boldsymbol{X}\boldsymbol{Q} &= \sum_{a\in S} \sum_{\nu',\nu\in B(\boldsymbol{H})} \int_{-\infty}^{0} \gamma_\beta(\omega)\omega \boldsymbol{Q}\boldsymbol{A}_{\nu'}^{a\dagger}\boldsymbol{Q}\boldsymbol{A}_{\nu}^{a}\boldsymbol{Q}\hat{f}_\mu^{\,*}(\omega-\nu')\hat{f}_\mu(\omega-\nu)\mathrm{d}\omega \\ &= \sum_{a\in S} \sum_{\nu\in B(\boldsymbol{H}|_{\boldsymbol{Q}})} \int_{-\infty}^{0} \gamma_\beta(\omega)\omega \boldsymbol{Q}\boldsymbol{A}_{\nu}^{a\dagger}\boldsymbol{Q}\boldsymbol{A}_{\nu}^{a}\boldsymbol{Q}\left|\hat{f}_\mu(\omega-\nu)\right|^2\mathrm{d}\omega.\end{aligned}$$

The first line inserts an additional projector $\boldsymbol{Q}$ between $\boldsymbol{A}_{\nu'}^{a\dagger}$ and $\boldsymbol{A}_{\nu}^{a}$ because any transition to excited states require $\nu, \nu' > \Delta_{\boldsymbol{Q}}$, but this is forbidden by the restrictions that $\omega \leq 0$ (from the integral) and that $|\nu - \omega|, |\nu' - \omega| < \mu < \Delta_{\boldsymbol{Q}}$ (from the secular approximation). In the second line, since the Bohr frequencies in $B(\boldsymbol{H}|_{\boldsymbol{Q}})$ are at least $\Delta_\nu = 2\mu$ apart, we must have that

$$\hat{f}_\mu^*(\omega-\nu')\hat{f}_\mu(\omega-\nu) = 0 \quad \text{for all} \quad \omega\in\mathbb{R}, \quad \text{unless} \quad \nu' = \nu.$$

Combining the above with Eq. (4.12) to conclude the proof. ∎

When the Hamiltonian is local, thinking about the gradient "locally" is sometimes useful. The following lemma gives a sufficient condition that guarantees a global gradient. Since the consequence is strong, the premise is also more stringent; it is only helpful when the Hamiltonian is frustration-free.

Lemma 4.12.7 (Local-to-global gradient condition)**.** *Suppose $\boldsymbol{H} = \sum_i \boldsymbol{h}_i$, where each term $\boldsymbol{h}_i \geq 0$. Then for any (not necessarily thermal) Lindbladian $\mathcal{L}$,*

$$-\mathcal{L}^\dagger[\boldsymbol{h}_i] \geq r_i\boldsymbol{h}_i \qquad \Longrightarrow \qquad -\mathcal{L}^\dagger[\boldsymbol{H}] \geq r\boldsymbol{H},$$

where $r = \min_i r_i$.

Proof. By linearity, we have $-\mathcal{L}^\dagger[\boldsymbol{H}] = \sum_i -\mathcal{L}^\dagger[\boldsymbol{h}_i] \geq \sum_i r_i\boldsymbol{h}_i$. Since $r_i\boldsymbol{h}_i \geq r\boldsymbol{h}_i$, we have $-\mathcal{L}^\dagger[\boldsymbol{H}] \geq r\sum_i \boldsymbol{h}_i = r\boldsymbol{H}$, concluding the proof. ∎

Gradients for commuting Hamiltonians

When we are given a commuting Hamiltonian, the energy gradient induced by any local jump operator can be understood by restricting the system to its neighborhood. In this situation, the negative gradient condition for the overall Hamiltonian can be decomposed into conditions that can be checked locally. This gives an efficient method to show a commuting Hamiltonian has a negative gradient for all its excited states, which we elucidate in this section of the appendix.

Recall the thermal Lindbladian $\mathcal{L} := \mathcal{L}^{\beta,\tau,\boldsymbol{H}}$ defined in Eq. (4.18) for a local jump operator $\boldsymbol{A}^a$, whose Heisenberg picture is

$$\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{O}] = \mathrm{i}[\boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}}, \boldsymbol{O}] + \mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{O}],$$

where

$$\mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{O}] = \int_{-\infty}^{\infty} \gamma_\beta(\omega)\Big[\hat{\boldsymbol{A}}^a(\omega)^\dagger \boldsymbol{O} \hat{\boldsymbol{A}}^a(\omega) - \frac{1}{2}\{\hat{\boldsymbol{A}}^a(\omega)^\dagger \hat{\boldsymbol{A}}^a(\omega), \boldsymbol{O}\}\Big]\mathrm{d}\omega.$$

Note $\hat{\boldsymbol{A}}^a(\omega)$ is the operator Fourier transform of $\boldsymbol{A}^a(t) = \mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}$, and $\boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}}$ is a Lamb-shift term defined in Eq. (4.22).

When $\boldsymbol{H}$ is a commuting Hamiltonian (e.g., [31, 96]), an important observation is that $\boldsymbol{A}^a(t)$ only depends on the part of $\boldsymbol{H}$ that does not commute with $\boldsymbol{A}^a$. In particular, the energy gradient for each jump operator only depends on the neighborhood of $\boldsymbol{A}^a$.

Lemma 4.12.8 (Commuting Hamiltonian and localized Lindblad operators)**.** *Suppose $\boldsymbol{H} = \sum_e \boldsymbol{h}_e$ is a commuting Hamiltonian. For any jump operator $\boldsymbol{A}^a$, the associated energy gradient simplifies to*

$$\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{H}] = \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}_{\ni a}}[\boldsymbol{H}_{\ni a}],$$

where $\boldsymbol{H}_{\ni a} = \sum_{e:\,[\boldsymbol{h}_e,\boldsymbol{A}^a]\neq 0} \boldsymbol{h}_e$ is the part of $\boldsymbol{H}$ does not commute with $\boldsymbol{A}^a$.

Proof. When $\boldsymbol{H}$ is commuting, we have $\boldsymbol{A}^a(t) = \mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} = \mathrm{e}^{\mathrm{i}\boldsymbol{H}_{\ni a}t}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\boldsymbol{H}_{\ni a}t}$, so the Lindbladian superoperator only depends on $\boldsymbol{H}_{\ni a}$, i.e., $\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}} = \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}_{\ni a}}$.

Let $\boldsymbol{H}_{\not\ni a} = \boldsymbol{H} - \boldsymbol{H}_{\ni a}$ be the part of $\boldsymbol{H}$ that commutes with $\boldsymbol{A}^a$. Since $[\boldsymbol{H}_{\not\ni a}, \boldsymbol{H}_{\ni a}] = 0$, we have $[\boldsymbol{A}^a(t), \boldsymbol{H}_{\not\ni a}] = 0$ for each t, which implies $[\boldsymbol{H}_{LS,a}^{\beta,\tau,\boldsymbol{H}}, \boldsymbol{H}_{\not\ni a}] = [\hat{\boldsymbol{A}}^a(\omega), \boldsymbol{H}_{\not\ni a}] = 0$. Thus we have $\mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}}[\boldsymbol{H}_{\not\ni a}] = 0$. ∎

Negative gradient condition under perturbations to Hamiltonians

We next look at how the negative energy gradient condition changes under perturbations to the n-qubit Hamiltonian $\boldsymbol{H}$. See Appendix 4.16 for the proof of the following theorem.

Theorem 4.12.1 (Monotonicity of gradient under level splitting)**.** *Consider a highly degenerate Hamiltonian $\boldsymbol{H} = \sum_{\bar{E}} \bar{E} \boldsymbol{P}_{\bar{E}}$ with Bohr-frequency gap $\Delta_\nu := \min_{\nu_1 \neq \nu_2 \in B(\boldsymbol{H})} |\nu_1 - \nu_2|$ of $\boldsymbol{H}$, and add a perturbation $\boldsymbol{H}' := \boldsymbol{H} + \boldsymbol{V}$. Let $\boldsymbol{P} = \boldsymbol{P}_{\bar{E}}$ be a projector to an energy subspace and $\boldsymbol{P}'$ the corresponding perturbed subspace. Suppose the perturbation is weaker than the Bohr-frequency gap, $\|\boldsymbol{V}\| \leq \frac{1}{8}\Delta_\nu$. For any $\beta, \tau > 0$, let $\mathcal{L} = \sum_{a \in S} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$, $\mathcal{L}' = \sum_{a \in S} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}'}$ be thermal Lindbladians with jumps $\{\boldsymbol{A}^a\}_{a \in S}$, where $\|\boldsymbol{A}^a\| \leq 1$ and the transition weight $\gamma_\beta(\omega)$ is given by Eq. (4.21). Then we have the monotone property that*

$$-\mathcal{L}^\dagger[\boldsymbol{H}] \geq r(\boldsymbol{I} - \boldsymbol{P}) - \epsilon \boldsymbol{I} \quad \textit{implies} \quad -\mathcal{L}'^\dagger[\boldsymbol{H}'] \geq r(\boldsymbol{I} - \boldsymbol{P}') - \epsilon' \boldsymbol{I},$$

where

$$\epsilon' \leq \epsilon + |S| \cdot O\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu \tau}} + \frac{e^{-\beta\Delta_\nu/4}}{\beta} + \left(1 + \frac{\Lambda_0 + r}{\Delta_\nu}\right)\|\boldsymbol{V}\|\right). \tag{4.48}$$

Finally, we look at how the negative energy gradient condition changes when restricted to a subspace. See Appendix 4.16 for proofs of the following two corollaries.

Corollary 4.12.1 (Monotonicity of gradient on a subspace)**.** *Consider a Hamiltonian $\boldsymbol{H} = \sum_{\bar{E}} \bar{E} \boldsymbol{P}_{\bar{E}}$ and its perturbation $\boldsymbol{H}' := \boldsymbol{H} + \boldsymbol{V}$. Let $\boldsymbol{P}$ be the ground space projector for $\boldsymbol{H}$ and $\boldsymbol{P}'$ be the corresponding perturbed eigensubspace of $\boldsymbol{H}'$. Let $\boldsymbol{Q}$ be a low-energy eigensubspace projector of $\boldsymbol{H}$ (i.e., $\boldsymbol{Q} = \sum_{E \leq E_{\boldsymbol{Q}}} \boldsymbol{P}_E$ for $E_{\boldsymbol{Q}} \in Spec(\boldsymbol{H})$) with excitation gap $\Delta_{\boldsymbol{Q}}$. Assume $\frac{\|\boldsymbol{V}\|\|\boldsymbol{H}\|}{\Delta_{\boldsymbol{Q}}} \leq \frac{1}{144}\Delta_\nu$ where $\Delta_\nu := \min_{\nu_1 \neq \nu_2 \in B(\boldsymbol{H}|_{\boldsymbol{Q}})} |\nu_1 - \nu_2|$ is the Bohr-frequency gap of $\boldsymbol{H}$ within the subspace $\boldsymbol{Q}$. For any $\beta, \tau > 0$, let $\mathcal{L} = \sum_{a \in S} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$, $\mathcal{L}' = \sum_{a \in S} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}'}$ be thermal Lindbladians with jumps $\{\boldsymbol{A}^a\}_{a \in S}$, where $\|\boldsymbol{A}^a\| \leq 1$ and the transition weight $\gamma_\beta(\omega)$ is given by Eq. (4.21). Then we have the monotone property that*

$$-\boldsymbol{Q}\mathcal{L}^\dagger[\boldsymbol{H}]\boldsymbol{Q} \geq r\boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{P}) - \epsilon \boldsymbol{I} \quad \textit{implies} \quad -\boldsymbol{Q}'\mathcal{L}'^\dagger[\boldsymbol{H}']\boldsymbol{Q}' \geq r\boldsymbol{Q}'(\boldsymbol{I} - \boldsymbol{P}') - \epsilon' \boldsymbol{I},$$

where $\boldsymbol{Q}'$ projects onto the perturbed eigensubspace of $\boldsymbol{H}'$ identified with $\boldsymbol{Q}$, and

$$\epsilon' \le \epsilon + |S| \cdot O\bigg(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu \tau}} + \frac{\Lambda_0}{\sqrt{\Delta_{\boldsymbol{Q}} \tau}} + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta} + \frac{\mathrm{e}^{-\beta\Delta_{\boldsymbol{Q}}/4}}{\beta} + \bigg(1 + \frac{\Lambda_0}{\Delta_\nu}\bigg)\frac{\|\boldsymbol{V}\|\|\boldsymbol{H}\|}{\Delta_{\boldsymbol{Q}}} + r\bigg(\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}} + \frac{\|\boldsymbol{V}\|}{\Delta_\nu}\bigg)\bigg). \tag{4.49}$$

Corollary 4.12.2 (Monotonicity of gradient on a subspace under off-block-diagonal perturbation)**.** *In the setting of Corollary 4.12.1, instead assume* $\frac{\|\boldsymbol{V}\|}{\Delta_\nu}, \frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}} \le (const.)$*, and that the perturbation is off-block-diagonal, i.e.,* $\boldsymbol{Q}\boldsymbol{V}\boldsymbol{Q} = (\boldsymbol{I} - \boldsymbol{Q})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{Q}) = 0$*. Then,*

$$-\boldsymbol{Q}\mathcal{L}^\dagger[\boldsymbol{H}]\boldsymbol{Q} \succeq r\boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{P}) - \epsilon\boldsymbol{I} \quad \textit{implies} \quad -\boldsymbol{Q}'\mathcal{L}'^\dagger[\boldsymbol{H}']\boldsymbol{Q}' \succeq r\boldsymbol{Q}'(\boldsymbol{I} - \boldsymbol{P}') - \epsilon'\boldsymbol{I},$$

where

$$\epsilon' \le \epsilon + |S| \cdot O\bigg(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu \tau}} + \frac{\Lambda_0}{\sqrt{\Delta_{\boldsymbol{Q}} \tau}} + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta} + \frac{\mathrm{e}^{-\beta\Delta_{\boldsymbol{Q}}/4}}{\beta} + \frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}} + \|\boldsymbol{H}_{\boldsymbol{Q}}\| \cdot \bigg(\frac{\|\boldsymbol{H}_{\boldsymbol{Q}}\|\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}\Delta_\nu} + \frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}\Delta_\nu}\bigg) + r\bigg(\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}} + \frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}\Delta_\nu}\bigg)\bigg).$$

4.13 Appendix:Energy landscape of an Ising chain

In this appendix, we take a brief aside to characterize the energy landscape of the one-dimensional ferromagnetic Ising chain under thermal perturbations. This provides a basic example on how the definition of local minima under thermal perturbations is related to the physical picture. We will see that this system has many suboptimal local minima in the absence of an external field with a lifetime polynomial in the system size. Once an external field is added, however, the system essentially has no suboptimal local minima and can quickly cool to the ground state where all spins are aligned. This observation corresponds to the following physical phenomena: when there is no external magnetic field, a ferromagnetic system will often be stuck in a configuration with many domain walls, and an externally applied magnetic field can quickly magnetize the system.

The Hamiltonian for the ferromagnetic Ising chain on a periodic boundary condition is

$$\boldsymbol{H} = -\sum_{j=1}^{n} \boldsymbol{Z}_j \boldsymbol{Z}_{j+1} - h \sum_{j=1}^{n} \boldsymbol{Z}_j,$$

where we identify $\boldsymbol{Z}_{n+1} \equiv \boldsymbol{Z}_1$. Intuitively, this system energetically favors configurations where adjacent spins are aligned. When $h = 0$, we have two degenerate ground states, $|00\cdots0\rangle$ and $|11\cdots1\rangle$, which are the global minima. This degeneracy is broken when $h \neq 0$, and these two states split by energy $2nh$. The system also has many excited states with *domain walls*, i.e., locations where adjacent spins are anti-aligned such as $|01\rangle$ and $|10\rangle$. In what follows, we study the energy landscape of the above system under thermal perturbations with jump operator $\{\boldsymbol{A}^j = \boldsymbol{X}_j\}_{j=1}^n$, setting $\tau = \infty$ for simplicity. We analyze three cases.

Case 1: no external field ($h = 0$). In this case, we will see that any bit string state with domain walls sufficiently far from each other, e.g. $|\cdots0001111000\cdots\rangle$ is a suboptimal local minimum. Indeed, there is no local operation to strictly decrease the energy of such states; the jump operators $\{\boldsymbol{X}_j\}$ can only displace the domain walls by one site, which does not change the energy.

We can see this more formally by computing the energy gradient operator. Since $\boldsymbol{H}$ is a commuting Hamiltonian, we may apply Lemma 4.12.8 and study the gradient induced by a single jump operator $\boldsymbol{X}_j$ by restricting the Hamiltonian to its neighborhood, i.e.,

$$\boldsymbol{H}_{\ni j} = -\boldsymbol{Z}_{j-1}\boldsymbol{Z}_j - \boldsymbol{Z}_j\boldsymbol{Z}_{j+1}.$$

Observe $\boldsymbol{H}_{\ni j}$ has three degenerate eigenspaces $\boldsymbol{P}_j^E$ with energy E as follows:

$$\boldsymbol{P}_j^{-2} = \sum_{s\in\{000,111\}} |s\rangle\langle s|_{j-1,j,j+1}, \qquad \boldsymbol{P}_j^{2} = \sum_{s\in\{010,101\}} |s\rangle\langle s|_{j-1,j,j+1},$$
$$\text{and} \qquad \boldsymbol{P}_j^{0} = \sum_{s\in\{001,100,011,110\}} |s\rangle\langle s|_{j-1,j,j+1}.$$

Then the negative Bohr-frequencies and the associated jumps are

$$\boldsymbol{A}^j_{\nu_1} = \boldsymbol{P}_j^{-2}\boldsymbol{X}_j\boldsymbol{P}_j^{2} = (|000\rangle\langle 010| + |111\rangle\langle 101|)_{j-1,j,j+1}, \qquad \nu_1 = -4,$$
$$\boldsymbol{A}^j_{\nu_2} = \boldsymbol{P}_j^{0}\boldsymbol{X}_j\boldsymbol{P}_j^{2} + \boldsymbol{P}_j^{-2}\boldsymbol{X}_j\boldsymbol{P}_j^{0} = 0, \qquad \nu_2 = -2.$$

Hence, the energy gradient operator associated with jump $\boldsymbol{X}_j$ is

$$\begin{aligned}\mathcal{D}_j^{\dagger\beta,\infty,\boldsymbol{H}}[\boldsymbol{H}] = \mathcal{D}_j^{\dagger\beta,\infty,\boldsymbol{H}_{\ni j}}[\boldsymbol{H}_{\ni j}] &= \textstyle\sum_{\nu\in B(\boldsymbol{H}_{\ni j})} \nu\gamma_\beta(\nu)\boldsymbol{A}_\nu^{j\dagger}\boldsymbol{A}_\nu^{j} \\ &= \theta_0\cdot(|010\rangle\langle 010| + |101\rangle\langle 101|)_{j-1,j,j+1} + O(e^{-4\beta}),\end{aligned}$$

where $\theta_0 = -4\gamma_\beta(-4) = -\Omega(1)$. As we can see, the energy gradient is essentially 0 when the domain walls are more than distance 1 apart, and only becomes significant when two domain walls are next to each other, as in $|\cdots 010\cdots\rangle$ or $|\cdots 101\cdots\rangle$. This implies the presence of exponentially many suboptimal local minima; for example, choose whether or not to have a domain wall every 2 sites.

Despite the presence of many suboptimal local minima, we now argue that they have a lifetime polynomial in the system size n when the system evolves under thermal perturbations. We may understand the dynamics of the system as a random walk of domain walls, and two domain walls annihilate each other when they meet. Since two domain walls at distance ℓ apart moving under diffusive dynamics take $O(\ell^2)$ time to meet, a suboptimal local minimum with k domain walls decays to a lower energy state after approximately $O(n^2/k^2)$ time.

Case 2: weak external field ($0 < h < 2$)**.** In this case, the ground state of $\boldsymbol{H}$ is uniquely $|0^n\rangle$, as all spins are slightly favored to be in the $|0\rangle$ state instead of the $|1\rangle$ state. When the domain walls are far apart, e.g. $|\cdots 0001111000\cdots\rangle$, the applied external field causes an attraction across the domain of 1's, which energetically favors the domain walls to move closer together. The presence of the field h removes all the suboptimal local minima that were in the previous case. The state $|1^n\rangle$, which was a ground state in the previous case, becomes now the only suboptimal local minimum.

We now more formally characterize the energy landscape of $\boldsymbol{H}$ using the energy gradient operator. Again applying Lemma 4.12.8, we may consider the gradient

induced by a single jump operator by focusing on its neighborhood. The relevant neighborhood Hamiltonian is

$$\boldsymbol{H}_{\ni j} = -\mathbf{Z}_{j-1}\mathbf{Z}_j - \mathbf{Z}_j\mathbf{Z}_{j+1} - h\mathbf{Z}_j.$$

Turning the crank, we see that the negative Bohr-frequencies and the associated jumps are

$$\begin{aligned} \boldsymbol{A}^j_{\nu_1} &= |000\rangle\langle 010|_{j-1,j,j+1}, & \nu_1 &= -4-2h, \\ \boldsymbol{A}^j_{\nu_2} &= |111\rangle\langle 101|_{j-1,j,j+1}, & \nu_2 &= -4+2h, \\ \boldsymbol{A}^j_{\nu_3} &= (|001\rangle\langle 011| + |100\rangle\langle 110|)_{j-1,j,j+1}, & \nu_3 &= -2h. \end{aligned} \tag{4.50}$$

Then the energy gradient operator associated with jump $\boldsymbol{X}_j$ is

$$\begin{aligned} \mathcal{D}_j^{\dagger\beta,\infty,\boldsymbol{H}}[\boldsymbol{H}] &= \textstyle\sum_{\nu\in B(\boldsymbol{H}_{\ni j})} \nu\gamma_\beta(\nu)\boldsymbol{A}^{j\dagger}_\nu \boldsymbol{A}^j_\nu \\ &= (\theta_1|010\rangle\langle 010| + \theta_2|101\rangle\langle 101| + \theta_3|011\rangle\langle 011| + \theta_3|110\rangle\langle 110|)_{j-1,j,j+1} + O(e^{-2\beta h}), \end{aligned}$$

where $\theta_j = \nu_j\gamma_\beta(\nu_j)$. As we can see, any configuration with a domain wall now has a significant gradient from at least one of the jumps. The only configurations without a significant energy gradient are $|0^n\rangle$, the ground state, and $|1^n\rangle$, a metastable local minimum.

Case 3: strong external field ($h > 2$). In this case, the external field is sufficiently strong that the state $|1^n\rangle$ is no longer a local minimum, and $\boldsymbol{H}$ has no suboptimal local minima. To see this, we note that $h > 2$ implies that $\nu_2 > 0$ in Eq. (4.50), which means the energetically favored jump operator is actually $\boldsymbol{A}^j_{-\nu_2} = \boldsymbol{A}^{j\dagger}_{\nu_2}$. This implies the energy gradient operator induced by the jump $\boldsymbol{X}_j$ in this case is

$$\begin{aligned} \mathcal{D}_j^{\dagger\beta,\infty,\boldsymbol{H}}[\boldsymbol{H}] &= (\theta_1|010\rangle\langle 010| + \theta_2'|111\rangle\langle 111| + \theta_3|011\rangle\langle 011| + \theta_3|110\rangle\langle 110|)_{j-1,j,j+1} \\ &\quad + O(e^{-2\beta(h-2)}), \end{aligned}$$

where $\theta_2' = -\nu_2\gamma_\beta(-\nu_2)$. This gives the state $|1^n\rangle$ a significant energy gradient, and thus the ground state $|0^n\rangle$ is the only local minimum of $\boldsymbol{H}$.

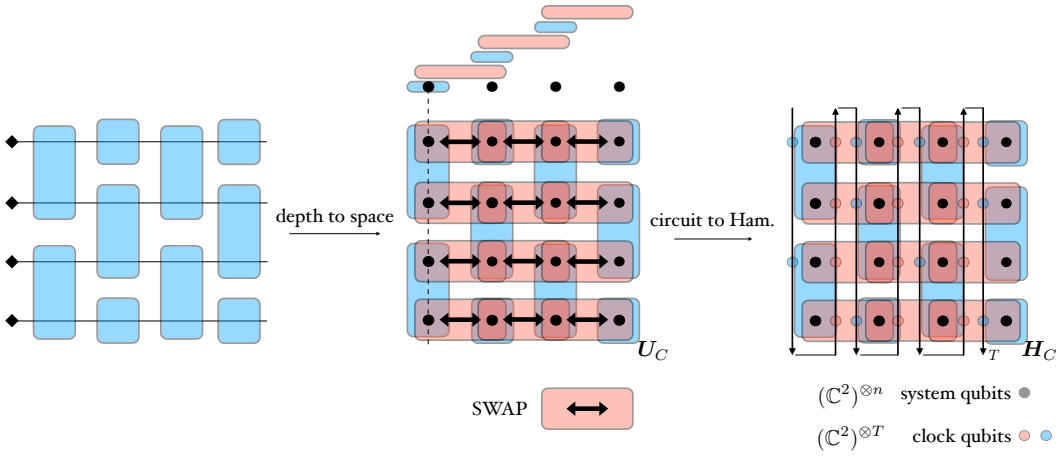


Figure 4.3: The circuit-to-Hamiltonian mapping. For any 1D circuit, we can turn the circuit depth into an additional spatial dimension by introducing layers of swap gates; this particularly structured 2D circuit (now on n qubits) is what we call $\boldsymbol{U}_C$. Then, we map the circuit $\boldsymbol{U}_C$ to a 2D-Hamiltonian by introducing the clock qubits, which weave between the layers of system qubits in a zigzag order. This layout ensures that each clock qubit only controls nearby system qubits (in particular, the odd clock layers control only blue gates, and the even layers control only pink gates.)

4.14 Appendix:All local minima are global in BQP-hard Hamiltonians (Proof of Theorem 4.9.2)

A main result of our work is that the task of finding a local minimum for $\boldsymbol{H}_C$ under thermal perturbation is universal for quantum computation and hence classically hard. As we have seen in the main text and Appendix 4.9, this main result follows from Theorem 4.9.2, which we prove in this appendix.

We start by defining $\boldsymbol{H}_C$ in detail (visualized in Fig. 4.3). Given a 2D n-qubit circuit $\boldsymbol{U}_C = \boldsymbol{U}_T \cdots \boldsymbol{U}_2 \boldsymbol{U}_1$ with $T = 2t_0 + L = \text{poly}(n)$ gates as constructed in Fig. 1 of Ref. [142], where the first and last t_0 gates are identity gates and each gate of the 2D circuit $\boldsymbol{U}_C$ is geometrically adjacent to the subsequent gate. We consider a geometrically local Hamiltonian on a 2D lattice with $n + T$ qubits defined as follows.

Definition 4.14.1 (Modified circuit-to-Hamiltonian construction)**.** *Consider a 2D circuit*

$$\boldsymbol{U}_C = \boldsymbol{U}_T \cdots \boldsymbol{U}_2 \boldsymbol{U}_1$$

on n qubits with $T = 2t_0 + L$ gates, where the first and last $t_0 = cL^2$ gates are identity gates with $c = O(1)$, and each consecutive gates are geometrically adjacent. We define a geometrically-local Hamiltonian $\boldsymbol{H}_C$ on a 2D lattice with $n + T$ qubits as

follows:

$$\boldsymbol{H}_C := \boldsymbol{H}_{\text{clock}} + \boldsymbol{H}_{\text{in}} + \boldsymbol{H}_{\text{prop}} \quad \textit{acting on} \quad (\mathbb{C}^2)^{\otimes n} \otimes (\mathbb{C}^2)^{\otimes T}, \tag{4.51}$$

where each individual term is given by

$$\begin{aligned}
\boldsymbol{H}_{\text{clock}} &:= J_{\text{clock}} \sum_{t=1}^{T-1} f_t \boldsymbol{I} \otimes |01\rangle\langle 01|_{t,t+1}, \\
\boldsymbol{H}_{\text{in}} &:= J_{\text{in}} \sum_{j=1}^{n} g_j |1\rangle\langle 1|_j \otimes |10\rangle\langle 10|_{t_j-1,t_j}, \\
\boldsymbol{H}_{\text{prop}} &:= \frac{1}{2} J_{\text{prop}} \sum_{t=1}^{T} \boldsymbol{H}_{\text{prop}}(t), \\
\boldsymbol{H}_{\text{prop}}(1) &:= \boldsymbol{I} - h_1(\boldsymbol{U}_1 \otimes |10\rangle\langle 00|_{1,2} + \boldsymbol{U}_1^\dagger \otimes |00\rangle\langle 10|_{1,2}), \\
\boldsymbol{H}_{\text{prop}}(t) &:= \boldsymbol{I} - h_t(\boldsymbol{U}_t \otimes |110\rangle\langle 100|_{t-1,t,t+1} + \boldsymbol{U}_t^\dagger \otimes |100\rangle\langle 110|_{t-1,t,t+1}) \\
\boldsymbol{H}_{\text{prop}}(T) &:= \boldsymbol{I} - h_T(\boldsymbol{U}_T \otimes |11\rangle\langle 10|_{T-1,T} + \boldsymbol{U}_T^\dagger \otimes |10\rangle\langle 11|_{T-1,T}) \\
&\quad \textit{for each} \quad 1 < t < T, .
\end{aligned}$$

The T qubits correspond to the T geometrically-local gates and are placed next to each gate to ensure $\boldsymbol{H}_C$ is geometrically local. The couplings are chosen as

$$J_{\text{clock}} = 1, \quad f_t = (T-t)/T, \quad g_j = 1/\xi_{t_j-1}, \quad h_t = \sqrt{t(T-t+1)}.$$

We will set the other parameters $J_{\text{in}}, J_{\text{prop}}$ later. The time t_j is the first time qubit j is acted on.

We will show later in Appendix 4.14 that $\boldsymbol{H}_C$ has a unique ground state given by

$$|\eta_{\mathbf{0}}\rangle = \sum_{t=0}^{T} \sqrt{\xi_t} \left(\boldsymbol{U}_t \cdots \boldsymbol{U}_1 |0^n\rangle\right) \otimes \left|1^t 0^{T-t}\right\rangle \qquad \text{where} \quad \xi_t := \frac{1}{2^T} \binom{T}{t}.$$

Note this state encodes the computational history of the circuit $\boldsymbol{U}_C$. By choosing $t_0 = L^2$, we ensure that each time in the interesting part of the computational history (i.e., the intermediate L gates) can be observed with $\Omega(1/T)$ probability as we will show later in Proposition 4.14.1. This also implies $g_j = O(T)$.

We now state a detailed version of Theorem 4.9.2 based on the definition of $\boldsymbol{H}_C$ in the following.

Theorem 4.14.1 (All local minima are global in $\boldsymbol{H}_C$)**.** *Let $\boldsymbol{P}_G$ be the ground-space projector for the Hamiltonian $\boldsymbol{H}_C$ in Eq. (4.51). For any failure probability $0 < \delta < 1$,*

there is a parameter choice $J_{\text{in}}, J_{\text{prop}} = \text{Poly}(n, T, \delta^{-1})$ *and a choice of* m *two-qubit jump operators*

$$S_0 = \{\boldsymbol{A}^a\}_{a=1}^m := \{\boldsymbol{I} \otimes \boldsymbol{X}_t, \boldsymbol{I} \otimes \boldsymbol{Z}_t\}_{t=1}^T \cup \{\boldsymbol{X}_j \otimes |0\rangle\langle 0|_{t_j}\}_{j=1}^n \tag{4.52}$$

with $m = 2T + n$ *satisfying the following:*

For a sufficiently small $\epsilon = 1/\text{Poly}(n, T, \delta^{-1})$*, any* ϵ*-approximate local minimum* $\boldsymbol{\rho}$ *of* $\boldsymbol{H}_C$ *under thermal perturbations with sufficiently large* $\beta = \text{Poly}(n, T, \delta^{-1})$*,* $\tau = \text{Poly}(n, T, \delta^{-1})$*, and system-bath interactions generated by* S_0 *is an exact global minimum with probability* $\text{tr}(\boldsymbol{P}_G(\boldsymbol{H}_C)\boldsymbol{\rho}) \geq 1 - \delta$.

We remind the reader that the thermal Lindbladians that generate the perturbations are defined in Eq. (4.18). The transition weight $\gamma_\beta(\omega)$ is chosen to be Glauber dynamics as defined in Eq. (4.21), with energy cut-off $\Lambda_0 = 1$ as a convenient choice so that $\|\omega\gamma_\beta(\omega)\|_\infty \leq 1$. We do not expect our result to change with other reasonable choices of $\gamma_\beta(\omega)$.

Remark 4.14.1.1. *Our* $\boldsymbol{H}_C$ *is similar to previous circuit-Hamiltonian constructions (see, e.g., [3, 100, 142]), but there are some significant differences. One key change is that* $\boldsymbol{H}_{\text{prop}}$ *is no longer frustration-free, and its couplings* h_t *are not uniform; consequently, this revised* $\boldsymbol{H}_{\text{prop}}$ *has better spectral properties that enable us to lower bound its Bohr-frequency gap. Furthermore,* $\boldsymbol{H}_{\text{clock}}$ *is given nonuniform couplings* f_t *so that any local excitation has an incentive to move rightwards (e.g.,* $|0011\rangle \to |0001\rangle$*), ensuring* $\boldsymbol{H}_{\text{clock}}$ *has no local minima except its ground states. These modifications allow us to prove that all excited states of* $\boldsymbol{H}_C$ *have significant negative gradients, so that they will all flow to the ground state under thermal perturbations.*

Characterizing low energy states of $\boldsymbol{H}_C$

We will start by characterizing the low energy states of the circuit Hamiltonian. We define the following sequence of Hamiltonians:

$$\begin{aligned}
\boldsymbol{H}_{\text{I}} &= \boldsymbol{H}_{\text{clock}} \\
\boldsymbol{H}_{\text{II}} &= \boldsymbol{H}_{\text{clock}} + \boldsymbol{H}_{\text{prop}} \\
\boldsymbol{H}_{\text{III}} &= \boldsymbol{H}_{\text{clock}} + \boldsymbol{H}_{\text{prop}} + \boldsymbol{H}_{\text{in}} = \boldsymbol{H}_C
\end{aligned}$$

with the ground space projectors $\boldsymbol{P}_j$ such that

$$\boldsymbol{P}_{\text{I}} \supset \boldsymbol{P}_{\text{II}} \supset \boldsymbol{P}_{\text{III}}.$$

Equivalently, we have $\boldsymbol{P}_{\mathrm{I}}\boldsymbol{P}_{\mathrm{II}} = \boldsymbol{P}_{\mathrm{II}}$ and $\boldsymbol{P}_{\mathrm{I}}\boldsymbol{P}_{\mathrm{III}} = \boldsymbol{P}_{\mathrm{II}}\boldsymbol{P}_{\mathrm{III}} = \boldsymbol{P}_{\mathrm{III}}$. Our approach to calculating the gradient for $\boldsymbol{H}_{\mathrm{III}} = \boldsymbol{H}_C$ is *perturbative*: we start with the simple Hamiltonian $\boldsymbol{H}_{\mathrm{I}}$ and gradually add perturbations (which will split the spectrum, Figure 4.4). Remarkably, the gradient is *stable* as long as the perturbation is weak enough. That is, it suffices to analyze the gradient of the simpler, unperturbed Hamiltonians on suitable subspaces.

Now we describe explicitly the ground subspaces $\boldsymbol{P}_{\mathrm{I}}$, $\boldsymbol{P}_{\mathrm{II}}$ and $\boldsymbol{P}_{\mathrm{III}}$. Let

$$|C_t\rangle = \left|1^t 0^{T-t}\right\rangle \quad \text{for each} \quad t = 0, 1, \ldots, T$$

and

$$\left|\eta_{\boldsymbol{x},t}\right\rangle = \left(\boldsymbol{U}_t \cdots \boldsymbol{U}_1 |\boldsymbol{x}\rangle\right) \otimes |C_t\rangle \quad \text{for each} \quad \boldsymbol{x} \in \{0,1\}^n \quad \text{and} \quad 0 \le t \le T.$$

The set of $\left|\eta_{\boldsymbol{x},t}\right\rangle$ forms an orthonormal basis for the ground space of $\boldsymbol{H}_{\mathrm{I}} = \boldsymbol{H}_{\mathrm{clock}}$, with energy 0 and a spectral gap of J_{clock}/T. The ground space projector is

$$\boldsymbol{P}_{\mathrm{I}} = \sum_{\boldsymbol{x}\in\{0,1\}^n} \sum_{t=0}^{T} |\eta_{\boldsymbol{x},t}\rangle\langle\eta_{\boldsymbol{x},t}| \quad \text{and} \quad \langle\eta_{\boldsymbol{y},t'}|\eta_{\boldsymbol{x},t}\rangle = \delta_{\boldsymbol{y}\boldsymbol{x}}\delta_{tt'}.$$

Observe that $[\boldsymbol{H}_{\mathrm{prop}}, \boldsymbol{P}_{\mathrm{I}}] = 0$, so the ground states of $\boldsymbol{H}_{\mathrm{II}} = \boldsymbol{H}_{\mathrm{clock}} + \boldsymbol{H}_{\mathrm{prop}}$ are given as the ground states of

$$\boldsymbol{P}_{\mathrm{I}}\boldsymbol{H}_{\mathrm{prop}}\boldsymbol{P}_{\mathrm{I}} = \frac{J_{\mathrm{prop}}}{2} \sum_{t=1}^{T} [\boldsymbol{I} - h_t(\boldsymbol{U}_t \otimes |C_t\rangle\langle C_{t-1}| + \boldsymbol{U}_t^\dagger \otimes |C_{t-1}\rangle\langle C_t|)].$$

Furthermore, observe the orthogonality relations

$$\langle\eta_{\boldsymbol{x},t}|\boldsymbol{H}_{\mathrm{prop}}|\eta_{\boldsymbol{y},t'}\rangle = 0 \qquad \text{when} \quad \boldsymbol{x} \neq \boldsymbol{y} \quad \text{for each} \quad t, t'.$$

That is, $\boldsymbol{P}_{\mathrm{I}}\boldsymbol{H}_{\mathrm{prop}}\boldsymbol{P}_{\mathrm{I}}$ is block diagonal with blocks labeled by $\boldsymbol{x}$. Moreover, for any $\boldsymbol{x}$, in the basis of $\left|\eta_{\boldsymbol{x},0}\right\rangle, \left|\eta_{\boldsymbol{x},1}\right\rangle, \ldots, \left|\eta_{\boldsymbol{x},T}\right\rangle$, we can explicitly write down the effective $(T+1) \times (T+1)$ Hamiltonian

$$\boldsymbol{P}_{\mathrm{I}}\boldsymbol{H}_{\mathrm{prop}}\boldsymbol{P}_{\mathrm{I}} = \frac{J_{\mathrm{prop}}}{2} \begin{pmatrix} T & -h_1 & & & & \\ -h_1 & T & -h_2 & & & \\ & -h_2 & T & -h_3 & & \\ & & -h_3 & \ddots & \ddots & \\ & & & \ddots & T & -h_T \\ & & & & -h_T & T \end{pmatrix} = J_{\mathrm{prop}}\left(\frac{T}{2}\boldsymbol{I} - \boldsymbol{L}_x\right),$$

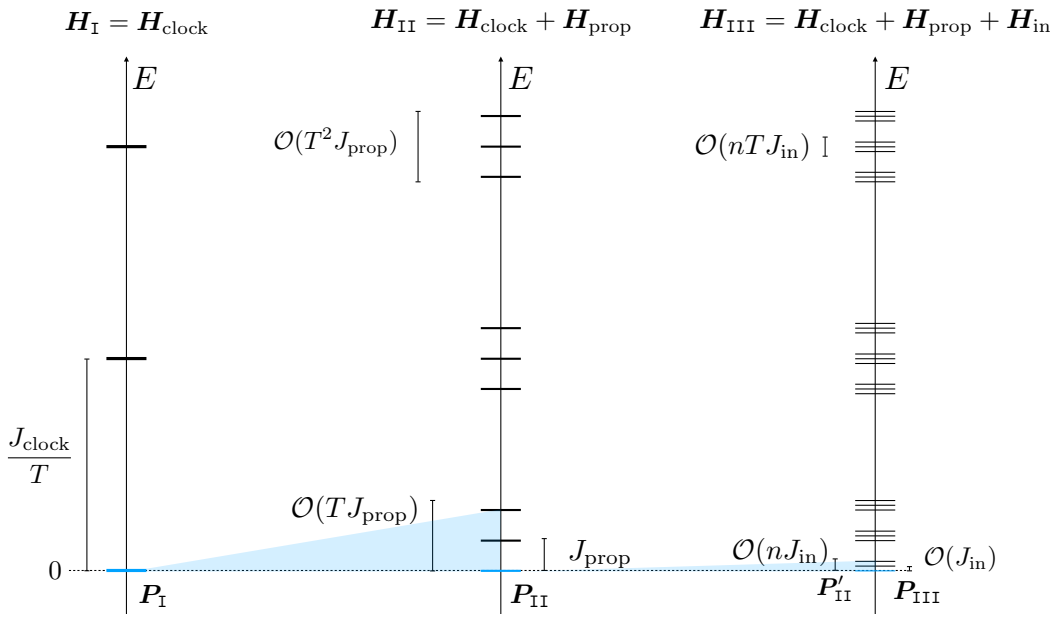


Figure 4.4: The degenerate levels of $\boldsymbol{H}_{\mathrm{clock}}$ split under perturbations $\boldsymbol{H}_{\mathrm{prop}}$ and $\boldsymbol{H}_{\mathrm{in}}$. In particular, the ground state splitting is tracked in blue shades. The careful choice of energy scales ensures that the levels can be identified with the original degenerate blocks.

where $\boldsymbol{L}_x$ is the matrix representation of the spin-$T/2$ angular momentum operator whose spectrum is well known. In particular, the unique ground state is

$$|\eta_{\boldsymbol{x}}\rangle := \sum_{t=0}^{T} \sqrt{\xi_t} |\eta_{\boldsymbol{x},t}\rangle \quad \text{with energy} \quad 0 \quad \text{and spectral gap} \quad J_{\mathrm{prop}}.$$

We will call $|\eta_{\boldsymbol{x}}\rangle$ the *history state* with respect to input $|\boldsymbol{x}\rangle$. The ground space projector of $\boldsymbol{H}_{\mathrm{II}}$ is then given as

$$\boldsymbol{P}_{\mathrm{II}} = \sum_{\boldsymbol{x}\in\{0,1\}^n} |\eta_{\boldsymbol{x}}\rangle\langle\eta_{\boldsymbol{x}}|.$$

Finally, note $|\eta_{\boldsymbol{0}}\rangle$ for $\boldsymbol{0} = (0, 0, \cdots, 0)$ is the unique ground state of $\boldsymbol{H}_C = \boldsymbol{H}_{\mathrm{III}}$ and so

$$\boldsymbol{P}_{\mathrm{III}} = |\eta_{\boldsymbol{0}}\rangle\langle\eta_{\boldsymbol{0}}|.$$

This is because $\boldsymbol{H}_{\mathrm{in}}$ is positive semi-definite, and $|\eta_{\boldsymbol{0}}\rangle$ is the only state in $\boldsymbol{P}_{\mathrm{II}}$ with zero eigenvalue with respect to $\boldsymbol{H}_{\mathrm{in}}$.

Proof of Theorem 4.14.1

To prove Theorem 4.14.1, we show that all excited states in $\boldsymbol{I} - \boldsymbol{P}_{\mathrm{III}}$ have significant gradient relative to $\boldsymbol{H}_{\mathrm{III}}$. Our analysis for the gradient will be carried out in three

subspaces:

$$\boldsymbol{I} - \boldsymbol{P}_{\mathrm{III}} = \underbrace{(\boldsymbol{I} - \boldsymbol{P}_{\mathrm{I}})}_{\text{studying } \boldsymbol{H}_{\mathrm{I}}} + \underbrace{\boldsymbol{P}_{\mathrm{I}}(\boldsymbol{I} - \boldsymbol{P}_{\mathrm{II}})}_{\text{studying } \boldsymbol{H}_{\mathrm{II}}\boldsymbol{P}_{\mathrm{I}}} + \underbrace{\boldsymbol{P}_{\mathrm{II}}(\boldsymbol{I} - \boldsymbol{P}_{\mathrm{III}})}_{\text{studying } \boldsymbol{H}_{\mathrm{III}}\boldsymbol{P}_{\mathrm{II}}}.$$

Let $\mathcal{L}_j := \mathcal{L}^{\beta,\tau,\boldsymbol{H}_j}$ be the thermal Lindbladian with uniform weights as in Eq. (4.30), defined with respect to $\boldsymbol{H}_j$ and the jump operators in Eq. (4.52) .

Case 1: Gradients for $\boldsymbol{I} - \boldsymbol{P}_{\mathrm{I}}$ from $\boldsymbol{H}_{\mathrm{I}}$. We first show excited states of $\boldsymbol{H}_{\mathrm{I}}$ have good energy gradient:

$$-\mathcal{L}_{\mathrm{I}}^{\dagger}[\boldsymbol{H}_{\mathrm{I}}] \geq r_1(\boldsymbol{I} - \boldsymbol{P}_{\mathrm{I}}) - \epsilon_{1a}\boldsymbol{I}. \tag{4.53}$$

Because $\boldsymbol{H}_{\mathrm{I}} = \boldsymbol{H}_{\mathrm{clock}}$ is a commuting Hamiltonian, the global gradient can be lower bounded by checking the local gradient from individual local jumps. We carry out this computation in section 4.14, where we show $r_1 = \Omega(1/T \ln \beta)$ and $\epsilon_{1a} = O(T^{7/4}/\tau^{1/4} + T/\beta + T(1+\beta)\ln\tau/\tau)$ in Lemma 4.14.2.

We then apply Theorem 4.12.1 (with $\boldsymbol{H} = \boldsymbol{H}_{\mathrm{I}}$ and $\boldsymbol{H}' = \boldsymbol{H}_{\mathrm{III}}$) to show excited states of $\boldsymbol{H}_{\mathrm{I}}$ have large gradient with respect to $\boldsymbol{H}_{\mathrm{III}}$. In other words,

$$-\mathcal{L}_{\mathrm{III}}^{\dagger}[\boldsymbol{H}_{\mathrm{III}}] \geq r_1(\boldsymbol{I} - \boldsymbol{P}_{\mathrm{I}}) - \epsilon_1\boldsymbol{I}. \tag{4.54}$$

To do this, we only need to check that the conditions of Theorem 4.12.1 are satisfied. Note $\boldsymbol{P}_{\mathrm{I}}$ also projects onto eigenstates of $\boldsymbol{H}_{\mathrm{III}}$ since $[\boldsymbol{P}_{\mathrm{I}}, \boldsymbol{H}_{\mathrm{III}}] = 0$. Note $\boldsymbol{H}_{\mathrm{I}}$ has a discrete spectrum with a minimum Bohr-frequency gap of at least $\Delta_\nu \geq 1/T$. We can choose sufficiently small $J_{\mathrm{prop}}, J_{\mathrm{in}}$ such that $\|\boldsymbol{V}_{\mathrm{I}}\| := \|\boldsymbol{H}_{\mathrm{prop}} + \boldsymbol{H}_{\mathrm{in}}\| \leq J_{\mathrm{prop}}T^2 + J_{\mathrm{in}}ng_{\max} \ll \Delta_\nu(\boldsymbol{H}_{\mathrm{I}}) = 1/T$, where $g_{max} := \max_{1\leq j\leq n} g_j = O(T)$. And plugging in $\Lambda_0 = 1$ and other parameters into the error bound (4.48)

$$\epsilon_1 = \epsilon_{1a} + |S_0|O\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}_{\mathrm{I}}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau^{1/3}} + \frac{1}{\sqrt{\Delta_\nu\tau}} + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta} + \left(1 + \frac{1+r_1}{\Delta_\nu}\right)\|\boldsymbol{V}_{\mathrm{I}}\|\right).$$

Noting that $|S_0|, \|\boldsymbol{H}_{\mathrm{I}}\|, g_{\max} = O(T)$, we can make $\epsilon_1/r_1 \leq \delta/6$ by choosing appropriate powers

$$\tau \geq \tilde{\Omega}(T^{11}/\delta^4), \quad \beta \geq \tilde{\Omega}(T^2/\delta), \quad J_{\mathrm{prop}} \leq \tilde{O}(\delta/T^5), \quad \text{and} \quad J_{\mathrm{in}} \leq \tilde{O}(\delta/nT^4). \tag{4.55}$$

Case 2: Gradients for $\boldsymbol{P}_{\mathrm{I}}(\boldsymbol{I}-\boldsymbol{P}_{\mathrm{II}})$ from $\boldsymbol{H}_{\mathrm{II}}$. We next restrict our attention to the action of $\mathcal{L}_{\mathrm{II}}^{\dagger}[\boldsymbol{H}_{\mathrm{II}}]$ inside the $\boldsymbol{P}_{\mathrm{I}}$ subspace, which conveniently is also an eigensubspace of both $\boldsymbol{H}_{\mathrm{II}}$ and $\boldsymbol{H}_{\mathrm{III}}$ since $[\boldsymbol{P}_{\mathrm{I}}, \boldsymbol{H}_{\mathrm{II}}] = [\boldsymbol{P}_{\mathrm{I}}, \boldsymbol{H}_{\mathrm{III}}] = 0$. Explicit computation in section 4.14 shows that

$$-\boldsymbol{P}_{\mathrm{I}}\mathcal{L}_{\mathrm{II}}^{\dagger}[\boldsymbol{H}_{\mathrm{II}}]\boldsymbol{P}_{\mathrm{I}} \succeq r_2\boldsymbol{P}_{\mathrm{I}}(\boldsymbol{I}-\boldsymbol{P}_{\mathrm{II}}) - \epsilon_{2a}\boldsymbol{I}, \tag{4.56}$$

with the bounds from Lemma 4.14.4 promising

$$r_2 = \Omega(\frac{J_{\mathrm{prop}}}{T\ln\beta}) \quad \text{and} \quad \epsilon_{2a} = |S_0|\cdot O\Big(\frac{1}{\tau} + \frac{\|\boldsymbol{H}_{\mathrm{II}}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\beta} + \frac{1}{\sqrt{\tau J_{\mathrm{prop}}}}\Big).$$

We then invoke Corollary 4.12.1 with $\boldsymbol{Q} = \boldsymbol{Q}' = \boldsymbol{P}_{\mathrm{I}}$, $\boldsymbol{H} = \boldsymbol{H}_{\mathrm{II}}$ and $\boldsymbol{H}' = \boldsymbol{H}_{\mathrm{III}}$ to show monotonicity of energy gradient on a subspace under perturbation

$$-\boldsymbol{P}_{\mathrm{I}}\mathcal{L}_{\mathrm{III}}^{\dagger}[\boldsymbol{H}_{\mathrm{III}}]\boldsymbol{P}_{\mathrm{I}} \succeq r_2\boldsymbol{P}_{\mathrm{I}}(\boldsymbol{I}-\boldsymbol{P}'_{\mathrm{II}}) - \epsilon_{2b}\boldsymbol{I},$$

where $\boldsymbol{P}'_{\mathrm{II}}$ is the perturbed eigensubspace of $\boldsymbol{H}_{\mathrm{III}}$ that is identified with $\boldsymbol{P}_{\mathrm{II}}$. To justify the application of Corollary 4.12.1, we note in $\boldsymbol{H}_{\mathrm{II}}$, the eigensubspace $\boldsymbol{P}_{\mathrm{I}}$ has an excitation gap of $\Delta_{\boldsymbol{Q}} \geq 1/T - 2\|\boldsymbol{H}_{\mathrm{prop}}\| = \Omega(1/T)$, where extra $2\|\boldsymbol{H}_{\mathrm{prop}}\|$ term is due to shifts in eigenvalues of $\boldsymbol{H}_{\mathrm{I}}$ bounded by Weyl's inequality (see Proposition 4.16.3). The perturbation on $\boldsymbol{H}_{\mathrm{II}}$ has strength $\|\boldsymbol{V}_{\mathrm{II}}\| := \|\boldsymbol{H}_{\mathrm{in}}\| \leq J_{\mathrm{in}} n g_{\max}$, and $\Delta_\nu(\boldsymbol{H}_{\mathrm{II}}|_{\boldsymbol{P}_{\mathrm{I}}}) = J_{\mathrm{prop}}$. Then noting $\Delta_\nu \ll \Delta_{\boldsymbol{Q}}$, we keep the dominant terms in the error bound (4.49) and get

$$\epsilon_{2b} \leq \epsilon_{2a} + |S_0|\cdot O\Big(\frac{\|\boldsymbol{V}_{\mathrm{II}}\|\|\boldsymbol{H}_{\mathrm{II}}\|}{\Delta_\nu\Delta_{\boldsymbol{Q}}} + r_2\frac{\|\boldsymbol{V}_{\mathrm{II}}\|}{\Delta_\nu}\Big).$$

Furthermore, the eigensubspace $\boldsymbol{P}_{\mathrm{I}}$ in $\boldsymbol{H}_{\mathrm{III}}$ is separated by a spectral gap of $1/T - 2\|\boldsymbol{V}_{\mathrm{I}}\| = \Omega(1/T)$ from the other eigenstates, we may apply Lemma 4.12.5 to show

$$-\mathcal{L}_{\mathrm{III}}^{\dagger}[\boldsymbol{H}_{\mathrm{III}}] \succeq r_2\boldsymbol{P}_{\mathrm{I}}(\boldsymbol{I}-\boldsymbol{P}'_{\mathrm{II}}) - \epsilon_2\boldsymbol{I} \quad \text{where} \quad \epsilon_2 = \epsilon_{2b} + |S_0|O(\frac{1}{\beta} + \frac{1}{\tau} + \sqrt{\frac{T}{\tau}}). \tag{4.57}$$

Since $|S_0|,\ g_{\max},\ \|\boldsymbol{H}_{\mathrm{II}}\|,\ \|\boldsymbol{H}_{\mathrm{III}}\| = O(T)$, we can make $\epsilon_2/r_2 \leq \delta/6$ by choosing

$$\tau \geq \tilde{\Omega}\Big(\frac{T^{11}}{J_{\mathrm{prop}}^2\delta^4}\Big), \quad \beta \geq \tilde{\Omega}\Big(\frac{T^2}{J_{\mathrm{prop}}\delta}\Big), \quad \text{and} \quad J_{\mathrm{in}} \leq \tilde{O}\Big(\frac{J_{\mathrm{prop}}^2\delta}{nT^5}\Big). \tag{4.58}$$

Case 3: Gradients for $\boldsymbol{P}_{\mathrm{II}}(\boldsymbol{I}-\boldsymbol{P}_{\mathrm{III}})$ from $\boldsymbol{H}_{\mathrm{III}}$. Now, we restrict our attention to $\boldsymbol{P}'_{\mathrm{II}}$, the perturbed eigensubspace in $\boldsymbol{H}_{\mathrm{III}}$ that corresponds to $\boldsymbol{P}_{\mathrm{II}}$. We can show by explicit computation (deferred to section 4.14) that

$$-\boldsymbol{P}'_{\mathrm{II}}\mathcal{L}^{\dagger}_{\mathrm{III}}[\boldsymbol{H}_{\mathrm{III}}]\boldsymbol{P}'_{\mathrm{II}} \succeq r_3\boldsymbol{P}'_{\mathrm{II}}(\boldsymbol{I}-\boldsymbol{P}_{\mathrm{III}}) - \epsilon_{3a}\boldsymbol{I}. \tag{4.59}$$

This computation shows that all valid history states $|\eta_{\boldsymbol{x}}\rangle$ except for $\boldsymbol{x}=\boldsymbol{0}$ have nonzero gradient with respect to $\mathcal{L}_{\mathrm{III}}$. The derivation uses a more fine-grained version of subspace gradient monotonicity (Corollary 4.12.2) since the standard version yields insufficient bounds. Roughly, we need to capture the fact that off-diagonal perturbations induce only *second-order* perturbation on the eigenvalues. The final calculated bounds in Eqs. (4.14) and (4.14) give us

$$r_3 = \Omega\Big(\frac{J_{\mathrm{in}}}{T^2\ln\beta}\Big) \quad \text{and} \quad \epsilon_{3a} \le TO\Big(\frac{T^{3/4}}{\tau^{1/4}} + \frac{1}{\beta} + \frac{1}{\sqrt{J_{\mathrm{in}}\tau}} + \mathrm{e}^{-\beta J_{\mathrm{in}}} + n\frac{(nTJ_{\mathrm{in}})^2}{J_{\mathrm{prop}}}\Big).$$

Using the fact that $\boldsymbol{P}'_{\mathrm{II}}$ is separated by energy of at least $J_{\mathrm{prop}} - 2\|\boldsymbol{H}_{\mathrm{in}}\|$ from the other eigenstates in $\boldsymbol{H}_{\mathrm{III}}$, we can apply Lemma 4.12.5 to get

$$-\mathcal{L}^{\dagger}_{\mathrm{III}}[\boldsymbol{H}_{\mathrm{III}}] \succeq r_3\boldsymbol{P}'_{\mathrm{II}}(\boldsymbol{I}-\boldsymbol{P}_{\mathrm{III}}) - \epsilon_3\boldsymbol{I} \quad \text{where} \quad \epsilon_3 = \epsilon_{3a} + |S_0|O\Big(\frac{1}{\beta}+\frac{1}{\tau}+\frac{1}{\sqrt{J_{\mathrm{prop}}\tau}}\Big). \tag{4.60}$$

We may ensure $\epsilon_3/r_3 \le \delta/6$ by choosing

$$\tau \ge \tilde{\Omega}\Big(\frac{T^{15}}{J_{\mathrm{in}}^4\delta^4}\Big), \quad \beta \ge \tilde{\Omega}\Big(\frac{T^3}{J_{\mathrm{in}}\delta}\Big), \quad \text{and} \quad J_{\mathrm{in}} \le \tilde{O}\Big(\frac{J_{\mathrm{prop}}\delta}{n^3T^5}\Big). \tag{4.61}$$

Altogether. Based on the conditions in Eqs. (4.55) (4.58) (4.61) and the fact that $T=\Omega(n)$, a consistent choice of parameters that satisfies all the bounds and ensures $\epsilon_j/r_j \le \delta/6$ are

$$\tau = \tilde{\Theta}\Big(\frac{T^{79}}{\delta^{16}}\Big), \quad \beta = \tilde{\Theta}\Big(\frac{T^{19}}{\delta^4}\Big), \quad J_{\mathrm{prop}} = \tilde{\Theta}\Big(\frac{\delta}{T^5}\Big), \quad \text{and} \quad J_{\mathrm{in}} = \tilde{\Theta}\Big(\frac{\delta^3}{T^{16}}\Big).$$

Then combining Eqs. (4.54), (4.57), and (4.60) implies that

$$\begin{aligned}
\boldsymbol{I}-\boldsymbol{P}_{\mathrm{I}} &\preceq \frac{\epsilon_1}{r_1}\boldsymbol{I} - \frac{1}{r_1}\mathcal{L}^{\dagger}_{\mathrm{III}}[\boldsymbol{H}_{\mathrm{III}}],\\
\boldsymbol{P}_{\mathrm{I}}-\boldsymbol{P}'_{\mathrm{II}} &\preceq \frac{\epsilon_2}{r_2}\boldsymbol{I} - \frac{1}{r_2}\mathcal{L}^{\dagger}_{\mathrm{III}}[\boldsymbol{H}_{\mathrm{III}}],\\
\boldsymbol{P}'_{\mathrm{II}}-\boldsymbol{P}_{\mathrm{III}} &\preceq \frac{\epsilon_3}{r_3}\boldsymbol{I} - \frac{1}{r_3}\mathcal{L}^{\dagger}_{\mathrm{III}}[\boldsymbol{H}_{\mathrm{III}}].
\end{aligned}$$

Note we have used $\boldsymbol{P}_{\mathrm{I}}\boldsymbol{P}'_{\mathrm{II}} = \boldsymbol{P}'_{\mathrm{II}}$ and $\boldsymbol{P}'_{\mathrm{II}}\boldsymbol{P}_{\mathrm{III}} = \boldsymbol{P}_{\mathrm{III}}$. Recall that $\mathcal{L}_{\mathrm{III}} = \sum_{a=1}^{m} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}_C}$. Adding all three inequalities together and normalizing suitably by the number of jumps $m = |S_0|$, we have

$$\boldsymbol{I} - \boldsymbol{P}_{\mathrm{III}} \preceq -\left(\sum_{j=1}^{3} \frac{m}{r_j}\right)\left(\frac{1}{m}\sum_{a=1}^{m} \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}_C}[\boldsymbol{H}_{\mathrm{III}}]\right) + \left(\sum_{j=1}^{3} \frac{\epsilon_j}{r_j}\right)\boldsymbol{I}.$$

The above provides the desired negative gradient condition on the full Hamiltonian $\boldsymbol{H}_{\mathrm{III}} = \boldsymbol{H}_C$. From Lemma 4.8.2, any ϵ-approximate local minimum $\boldsymbol{\rho}$ of $\boldsymbol{H}_C$ under thermal perturbation satisfies

$$1 - \mathrm{tr}(\boldsymbol{P}_{\mathrm{III}}\boldsymbol{\rho}) \leq \sum_{j=1}^{3} \frac{\epsilon_j + m\epsilon}{r_j} \leq \frac{\delta}{2} + \epsilon m \sum_{j=1}^{3} \frac{1}{r_j}.$$

By choosing $\epsilon \leq \frac{\delta}{2m}\left(\sum_{j=1}^{3} 1/r_j\right)^{-1} = 1/\mathrm{Poly}(n, T, \delta^{-1})$, we guarantee that $1 - \mathrm{tr}(\boldsymbol{P}_{\mathrm{III}}\boldsymbol{\rho}) \leq \delta$. This concludes our proof of Theorem 4.14.1.

Explicit calculations for energy gradients

In this section of the appendix, we provide the missing calculations supporting the claims that were asserted in Eqs. (4.53), (4.56), and (4.59) in the above proof of Theorem 4.14.1.

Gradient from $\boldsymbol{H}_{\mathrm{clock}}$

Note that $\boldsymbol{H}_{\mathrm{I}} = \boldsymbol{H}_{\mathrm{clock}}$ is a commuting Hamiltonian

$$\boldsymbol{H}_{\mathrm{I}} = \boldsymbol{H}_{\mathrm{clock}} = J_{\mathrm{clock}} \sum_{t=1}^{T-1} f_t \boldsymbol{h}_{t,t+1}, \quad \text{where} \qquad \boldsymbol{h}_{t,t+1} = \boldsymbol{I} \otimes |01\rangle\langle 01|_{t,t+1},$$

where we set $f_t = (T-t)/T$ and $J_{\mathrm{clock}} = 1$. We start by computing the gradient from a single jump operator, using the simplification from Lemma 4.12.8:

Lemma 4.14.1. *Let the jump operator $\boldsymbol{A}^t = \boldsymbol{I} \otimes \boldsymbol{X}_t$ for each $t \in [T]$. For all $t = 2, \ldots, T$, we have*

$$-\mathcal{D}_t^{\dagger\beta,\tau,\boldsymbol{H}_{\mathrm{I}}}[\boldsymbol{H}_{\mathrm{I}}] \succeq r_1 \boldsymbol{h}_{t-1,t} - \epsilon_0 \boldsymbol{I},$$

where $r_1 = \Omega(\frac{1}{T \ln \beta})$ and $\epsilon_0 = O(\sqrt{T/\tau} + [(1+\beta)\ln\tau]/\tau + 1/\beta)$.

Proof. By Lemma 4.12.8, we have $\mathcal{D}_t^{\dagger\beta,\tau,\boldsymbol{H}_{\mathrm{I}}}[\boldsymbol{H}_{\mathrm{I}}] = \mathcal{D}_t^{\dagger\beta,\tau,\boldsymbol{H}_{\ni t}}[\boldsymbol{H}_{\ni t}]$. We then proceed in two cases.

Case 1: $2 \le t \le T-1$**.** In this case, the relevant part of $\boldsymbol{H}_{\mathrm{I}}$ that does not commute with $\boldsymbol{A}^t$ is

$$\boldsymbol{H}_{\ni t} = f_{t-1}|01\rangle\langle 01|_{t-1,t} + f_t|01\rangle\langle 01|_{t,t+1}.$$

Observe that $\boldsymbol{H}_{\ni t}$ has three degenerate eigenspaces with corresponding energies as follows:

$$\begin{aligned}
\boldsymbol{P}_t^0 &= \sum_{s\in\{000,100,110,111\}} |s\rangle\langle s|_{t-1,t,t+1}, && E_0 = 0,\\
\boldsymbol{P}_t^L &= \sum_{s\in\{010,011\}} |s\rangle\langle s|_{t-1,t,t+1}, && E_L = f_{t-1},\\
\boldsymbol{P}_t^R &= \sum_{s\in\{001,101\}} |s\rangle\langle s|_{t-1,t,t+1}, && E_R = f_t.
\end{aligned}$$

The possible negative Bohr frequencies and the associated jumps are

$$\begin{aligned}
\boldsymbol{A}_{\nu_1}^t &= \boldsymbol{P}_t^0 \boldsymbol{X}_t \boldsymbol{P}_t^L = |000\rangle\langle 010|_{t-1,t,t+1}, && \nu_1 = -f_{t-1},\\
\boldsymbol{A}_{\nu_2}^t &= \boldsymbol{P}_t^0 \boldsymbol{X}_t \boldsymbol{P}_t^R = |111\rangle\langle 101|_{t-1,t,t+1}, && \nu_2 = -f_t,\\
\boldsymbol{A}_{\nu_3}^t &= \boldsymbol{P}_t^R \boldsymbol{X}_t \boldsymbol{P}_t^L = |001\rangle\langle 011|_{t-1,t,t+1}, && \nu_3 = f_t - f_{t-1} = -1/T.
\end{aligned}$$

Furthermore, observe that the Bohr frequencies are exactly integer multiples of $1/T$, so we can lower bound the Bohr-frequency gap $\Delta_\nu(\boldsymbol{H}_{\ni t}) \ge 1/T$. Then by Lemma 4.12.3 and 4.12.4, we can replace $\mathcal{D}_t^{\dagger\beta,\tau,\boldsymbol{H}_{\ni t}}$ with $\mathcal{D}_t^{\dagger\beta,\infty,\boldsymbol{H}_{\ni t}}$ up to an $O(\sqrt{T/\tau} + [(1+\beta)\ln\tau]/\tau)$ error.

Letting $\theta_j = \nu_j \gamma_\beta(\nu_j)$ for $j = 1,2,3$ (recall γ_β is given in Eq. (4.21) with $\Lambda_0 = 1$), we have

$$\begin{aligned}
\mathcal{D}_t^{\dagger\beta,\infty,\boldsymbol{H}_{\ni t}}[\boldsymbol{H}_{\ni t}] &= \sum_{\nu\in B(\boldsymbol{H}_{\ni t})} \nu\gamma_\beta(\nu)\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{A}_\nu^a\\
&= \big(\theta_1|010\rangle\langle 010| + \theta_2|101\rangle\langle 101| + \theta_3|011\rangle\langle 011|\big)_{t-1,t,t+1} + O(1/\beta),
\end{aligned}$$

where the last error term is due to heating transitions (positive Bohr frequencies), which incur errors of at most $\|\omega\gamma_\beta(\omega)\mathbb{1}(\omega > 0)\|_\infty = O(1/\beta)$. Note that $\theta_1, \theta_2, \theta_3 < 0$, and furthermore we have

$$\min\{|\theta_1|, |\theta_3|\} \ge \min_{\omega\in[-1,-1/T]} |\omega|\gamma_\beta(\omega) =: r_1 = \Omega\Big(\frac{1}{T\ln\beta}\Big).$$

Hence,

$$-\mathcal{D}_t^{\dagger\beta,\infty,\boldsymbol{H}_{\ni t}}[\boldsymbol{H}_{\ni t}] \ge r_1\big(|010\rangle\langle 010| + |011\rangle\langle 011|\big)_{t-1,t,t+1} - O(1/\beta)\boldsymbol{I}.$$

Note the first term combines to make $\boldsymbol{h}_{t-1,t}$. We then return to the finite-τ Lindbladian up to the aforementioned error:

$$-\mathcal{D}_t^{\dagger\beta,\tau,\boldsymbol{H}_{\ni t}}[\boldsymbol{H}_{\ni t}] \geq r_1 \boldsymbol{h}_{t-1,t} - O\Big(\sqrt{\frac{T}{\tau}} + \frac{(1+\beta)\ln\tau}{\tau} + \frac{1}{\beta}\Big)\boldsymbol{I}$$

Case 2: $t = T$**.** The relevant part of $\boldsymbol{H}_{\mathrm{I}}$ in this case is

$$\boldsymbol{H}_{\ni T} = f_{T-1}|01\rangle\langle 01|_{T-1,T},$$

which has two eigenspaces. There is only one negative Bohr frequency with a corresponding jump operator filtered at ν

$$\boldsymbol{A}_\nu^T = (\boldsymbol{I} - |01\rangle\langle 01|_{T-1,T})\boldsymbol{X}_T|01\rangle\langle 01|_{T-1,T} = |00\rangle\langle 01|_{T-1,T} \quad \text{where} \quad \nu = -f_{T-1} = -\frac{1}{T}.$$

Then,

$$\mathcal{D}_t^{\dagger\beta,\infty,\boldsymbol{H}_{\ni T}}[\boldsymbol{H}_{\ni T}] = -\tfrac{1}{T}\gamma_\beta(-\tfrac{1}{T})|01\rangle\langle 01|_{T-1,T} + O(\tfrac{1}{\beta})\boldsymbol{I}.$$

Note that $\frac{1}{T}\gamma_\beta(-\frac{1}{T}) \geq r_1$. Applying Lemma 4.12.3 and 4.12.4 to return to the finite τ expression,

$$-\mathcal{D}_t^{\dagger\beta,\tau,\boldsymbol{H}_{\ni T}}[\boldsymbol{H}_{\ni T}] \geq r_1 \boldsymbol{h}_{T-1,T} - O\Big(\sqrt{\frac{T}{\tau}} + \frac{(1+\beta)\ln\tau}{\tau} + \frac{1}{\beta}\Big)\boldsymbol{I},$$

which is the advertised result. ∎

We are now ready to prove Eq. (4.53), which we state as the following lemma:

Lemma 4.14.2. *Assume* $1 \leq T \leq \tau$*. We have*

$$-\mathcal{L}_{\mathrm{I}}^\dagger[\boldsymbol{H}_{\mathrm{I}}] \geq r_1(\boldsymbol{I} - \boldsymbol{P}_{\mathrm{I}}) - \epsilon_{1a}\boldsymbol{I},$$

where

$$r_1 = \Omega\Big(\frac{1}{T\ln\beta}\Big) \quad \textit{and} \quad \epsilon_{1a} = O\Big(\frac{T^{7/4}}{\tau^{1/4}} + \frac{T}{\beta} + \frac{T(1+\beta)\ln\tau}{\tau}\Big).$$

Proof. Note by linearity, we have

$$\mathcal{L}_{\mathrm{I}}^\dagger[\boldsymbol{H}_{\mathrm{I}}] = \sum_{a\in S} \mathcal{L}_a^{\dagger\beta,\tau,\boldsymbol{H}_{\mathrm{I}}}[\boldsymbol{H}_{\mathrm{I}}].$$

Let $S_{\mathrm{I}} = \{\boldsymbol{I}\otimes\boldsymbol{X}_t : 2 \leq t \leq T\}$ be a subset of the jump operators. Then

$$-\mathcal{L}_{\mathrm{I}}^\dagger[\boldsymbol{H}_{\mathrm{I}}] \geq -\sum_{a\in S_{\mathrm{I}}} \mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}_{\mathrm{I}}}[\boldsymbol{H}_{\mathrm{I}}] - O\Big(|S_0|\Big(\frac{\|\boldsymbol{H}_{\mathrm{I}}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta}\Big)\Big)\boldsymbol{I},$$

where the error contribution from neglecting the Lamb-shift term and the other jump operators in $S_0 \setminus S_{\mathrm{I}}$ are bounded by Proposition 4.10.3 and Lemma 4.12.1.

Applying Lemma 4.14.1 to the sum on the right hand side above, we get

$$-\sum_{a\in S_{\mathrm{I}}} \mathcal{D}_a^{\dagger\beta,\tau,\boldsymbol{H}_{\mathrm{I}}}[\boldsymbol{H}_{\mathrm{I}}] \geq r_1 \sum_{t=1}^{T-1} \boldsymbol{h}_{t,t+1} - T\epsilon_0 \boldsymbol{I}.$$

It is not difficult to see that

$$\sum_{t=1}^{T-1} \boldsymbol{h}_{t,t+1} \geq \boldsymbol{I} - \boldsymbol{P}_{\mathrm{I}},$$

that is, the smallest excitation has energy 1. Hence,

$$-\mathcal{L}_{\mathrm{I}}^{\dagger}[\boldsymbol{H}_{\mathrm{I}}] \geq r_1(\boldsymbol{I} - \boldsymbol{P}_{\mathrm{I}}) - \epsilon_{1a}\boldsymbol{I},$$

where

$$\begin{aligned}\epsilon_{1a} &= O\Big(|S_0|\Big(\frac{\|\boldsymbol{H}_{\mathrm{I}}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta}\Big)\Big) + O\Big(\frac{T^{3/2}}{\tau^{1/2}} + \frac{T(1+\beta)\ln\tau}{\tau} + \frac{T}{\beta}\Big) \\ &= O\Big(\frac{T^{7/4}}{\tau^{1/4}} + \frac{T}{\tau} + \frac{T}{\beta} + \frac{T^{3/2}}{\tau^{1/2}} + \frac{T(1+\beta)\ln\tau}{\tau}\Big).\end{aligned}$$

The last equality uses that $|S_0|, \|\boldsymbol{H}_{\mathrm{I}}\| = O(T)$. Since $1 \leq T \leq \tau$, we have $T^{7/4}/\tau^{1/4} \geq T/\tau$ and $T^{7/4}/\tau^{1/4} \geq T^{3/2}/\tau^{1/2}$, so we drop the latter two terms for the final error estimate in the lemma statement. ∎

Gradient from $\boldsymbol{H}_{\mathrm{prop}}$

In this subsection, we prove

$$-\boldsymbol{P}_{\mathrm{I}}\mathcal{L}_{\mathrm{II}}^{\dagger}[\boldsymbol{H}_{\mathrm{II}}]\boldsymbol{P}_{\mathrm{I}} \geq r_2\boldsymbol{P}_{\mathrm{I}}(\boldsymbol{I} - \boldsymbol{P}_{\mathrm{II}}) - \epsilon_{2a}\boldsymbol{I}.$$

Denote $|t\rangle := \big|\eta_{\boldsymbol{x},t}\big\rangle$ in what follows. Let $\boldsymbol{L}_+$ be the raising operator whose only nontrivial action is

$$\boldsymbol{L}_+|t\rangle = \sqrt{(t+1)(T-t)}|t+1\rangle \quad \text{for each} \quad 0 \leq t \leq T-1$$
$$\text{and} \quad \boldsymbol{L}_- =: \boldsymbol{L}_+^{\dagger}.$$

Furthermore, let $\boldsymbol{L}_x = \frac{1}{2}(\boldsymbol{L}_+ + \boldsymbol{L}_-)$, $\boldsymbol{L}_y = \frac{1}{2\mathrm{i}}(\boldsymbol{L}_+ - \boldsymbol{L}_-)$, $\boldsymbol{L}_z = \sum_{t=0}^{T}(t - T/2)|t\rangle\langle t|$. These operators form a set of angular momentum operators as $[\boldsymbol{L}_a, \boldsymbol{L}_b] = \mathrm{i}\epsilon_{abc}\boldsymbol{L}_c$ for $a, b, c \in \{x, y, z\}$. As noted earlier, we have

$$\boldsymbol{P}_{\mathrm{I}}\boldsymbol{H}_{\mathrm{prop}}\boldsymbol{P}_{\mathrm{I}} = J_{\mathrm{prop}}(\frac{T}{2} - \boldsymbol{L}_x).$$

The eigenstates are known to be

$$|v_k\rangle = \mathrm{e}^{\mathrm{i}\pi \boldsymbol{L}_y/2}|k\rangle \quad \text{with eigenvalues} \quad \lambda_k = kJ_{\text{prop}} \quad \text{for} \quad k = 0, 1, \ldots, T.$$

This integer spectrum means the minimum Bohr-frequency gap in the subspace $\boldsymbol{P}_{\mathtt{I}}$ is $\Delta_\nu(\boldsymbol{H}_{\text{prop}}|_{\boldsymbol{P}_{\mathtt{I}}}) = J_{\text{prop}}$.

Next, we give jump operators with nontrivial gradient on any excited state of $\boldsymbol{P}_{\mathtt{I}}\boldsymbol{H}_{\text{prop}}\boldsymbol{P}_{\mathtt{I}}$. These will be the 1-local jumps acting on the clock register

$$\boldsymbol{I} \otimes \boldsymbol{Z}_\ell \quad \text{for each} \quad 1 \le \ell \le T$$

which nicely respects the block-diagonal structure of $\boldsymbol{H}_{\text{prop}}$ such that

$$\langle \eta_{\boldsymbol{x},t}|\boldsymbol{I} \otimes \boldsymbol{Z}_\ell|\eta_{\boldsymbol{y},t'}\rangle = 0 \quad \text{if} \quad \boldsymbol{y} \neq \boldsymbol{x}. \tag{4.62}$$

Thus, fixing $\boldsymbol{x}$, we merely need to consider effective jump operators

$$\boldsymbol{P}_{\mathbb{I}}(\boldsymbol{I} \otimes \boldsymbol{Z}_\ell)\boldsymbol{P}_{\mathbb{I}} \equiv \boldsymbol{\sigma}_\ell \quad \text{such that} \quad \boldsymbol{\sigma}_\ell|t\rangle = (-1)^{\mathbb{1}_{t\ge\ell}}|t\rangle. \tag{4.63}$$

Lemma 4.14.3 (Good transition rates)**.** *For the operators $\boldsymbol{\sigma}_\ell$ in Eq. (4.63) and any $0 \le k < T$, we have that*

$$\max_{\ell\in[T]} |\langle v_k|\boldsymbol{\sigma}_\ell|v_{k+1}\rangle| \ge \frac{1}{\sqrt{T}}.$$

Proof. Fix any $0 \le k < T$. Observe that

$$\langle v_k|\boldsymbol{L}_z|v_{k+1}\rangle = \langle k|\mathrm{e}^{-\mathrm{i}\pi \boldsymbol{L}_y/2}\boldsymbol{L}_z\mathrm{e}^{\mathrm{i}\pi \boldsymbol{L}_y/2}|k+1\rangle = \langle k|\boldsymbol{L}_x|k+1\rangle = \frac{1}{2}\sqrt{(k+1)(T-k)}.$$

Then,

$$\begin{aligned}
\max_t |\langle v_k|\boldsymbol{\sigma}_t|v_{k+1}\rangle| &\ge \frac{1}{T}\sum_{t=1}^{T} |\langle v_k|\boldsymbol{\sigma}_t|v_{k+1}\rangle| \\
&\ge \frac{2}{T}|\langle v_k|\boldsymbol{L}_z|v_{k+1}\rangle| \\
&\ge \frac{1}{\sqrt{T}}
\end{aligned}$$

as advertised. ∎

Now that we understand the connectivity between the t labels, we may restore the 2^n many labels $\boldsymbol{x}$

$$|v_k\rangle \to \big|v_{k,\boldsymbol{x}}\big\rangle \quad \text{such that} \quad \langle v_{k',\boldsymbol{y}}|v_{k,\boldsymbol{x}}\rangle = \delta_{kk'}\delta_{\boldsymbol{x}\boldsymbol{y}}.$$

Fortunately, we do not need to address the explicit labels $\boldsymbol{x}$ due to the orthogonality properties in Eq. (4.62). We may now calculate the gradient operator $\boldsymbol{P}_{\mathtt{I}}\mathcal{L}^\dagger_{\mathtt{II}}[\boldsymbol{H}_{\mathtt{II}}]\boldsymbol{P}_{\mathtt{I}}$.

Lemma 4.14.4. *Consider the thermal Lindbladian* $\mathcal{L}_{\mathrm{II}} = \sum_{a\in S_0} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}_{\mathrm{II}}}$. *Then,*

$$-\boldsymbol{P}_{\mathrm{I}}\mathcal{L}_{\mathrm{II}}^{\dagger}[\boldsymbol{H}_{\mathrm{II}}]\boldsymbol{P}_{\mathrm{I}} \succeq r_2\boldsymbol{P}_{\mathrm{I}}(\boldsymbol{I}-\boldsymbol{P}_{\mathrm{II}}) - \epsilon_2\boldsymbol{I}$$

for

$$r_2 = \Omega(\frac{J_{\mathrm{prop}}}{T\ln\beta}) \quad \textit{and} \quad \epsilon_2 = |S_0|\cdot O\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}_{\mathrm{II}}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\beta} + \frac{1}{\sqrt{\tau J_{\mathrm{prop}}}}\right).$$

Proof. Observe that $\boldsymbol{P}_{\mathrm{I}}$ projects onto a low-energy subspace of $\boldsymbol{H}_{\mathrm{II}}$ with an excitation gap of at least $J_{\mathrm{clock}}/T - 2\|\boldsymbol{H}_{\mathrm{prop}}\|$ from Weyl's inequality. Furthermore, $\boldsymbol{H}_{\mathrm{II}}$ restricted to $\boldsymbol{P}_{\mathrm{I}}$ has eigenvalues that are integer multiples of J_{prop}, so the Bohr-frequency gap in the subspace is $\Delta_\nu(\boldsymbol{H}_{\mathrm{II}}|_{\boldsymbol{P}_{\mathrm{I}}}) = J_{\mathrm{prop}}$. Assuming $J_{\mathrm{prop}}/2 < J_{\mathrm{clock}}/T - 2\|\boldsymbol{H}_{\mathrm{prop}}\|$, we may apply Lemma 4.12.6 with $\boldsymbol{H} = \boldsymbol{H}_{\mathrm{II}}$, $\boldsymbol{Q} = \boldsymbol{P}_{\mathrm{I}}$ to get

$$\begin{aligned}\boldsymbol{P}_{\mathrm{I}}\mathcal{L}_{\mathrm{II}}^{\dagger}[\boldsymbol{H}_{\mathrm{II}}]\boldsymbol{P}_{\mathrm{I}} &\overset{E}{\approx} \sum_{a\in S_0}\sum_{\nu\in B(\boldsymbol{H}_{\mathrm{II}}|_{\boldsymbol{P}_{\mathrm{I}}})} \boldsymbol{P}_{\mathrm{I}}\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{P}_{\mathrm{I}}\boldsymbol{A}_\nu^{a}\boldsymbol{P}_{\mathrm{I}} \int_{-\infty}^{0}\gamma_\beta(\omega)\omega\left|\hat{f}_\mu(\omega-\nu)\right|^2\mathrm{d}\omega \\ &\preceq \sum_{a\in S_{\mathrm{II}}}\sum_{\nu\in B(\boldsymbol{H}_{\mathrm{II}}|_{\boldsymbol{P}_{\mathrm{I}}})} \boldsymbol{P}_{\mathrm{I}}\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{P}_{\mathrm{I}}\boldsymbol{A}_\nu^{a}\boldsymbol{P}_{\mathrm{I}} \int_{-\infty}^{0}\gamma_\beta(\omega)\omega\left|\hat{f}_\mu(\omega-\nu)\right|^2\mathrm{d}\omega,\end{aligned} \tag{4.68}$$

where $\mu = \Delta_\nu/2$ and

$$E = |S_0|O\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta} + \frac{1}{\sqrt{\Delta_\nu\tau}}\right).$$

The second line uses the negativity of the half-integral to reduce to the following subset of jump operators from Eq. (4.52)

$$S_{\mathrm{II}} = \{\boldsymbol{I}\otimes\boldsymbol{Z}_\ell : \ell\in[T]\}.$$

Let us now explicitly display the matrix elements of the above jump operators in the $|v_{k,\boldsymbol{x}}\rangle$ basis:

$$\begin{aligned}\boldsymbol{P}_{\mathrm{I}}(\boldsymbol{I}\otimes\boldsymbol{Z}_\ell)\boldsymbol{P}_{\mathrm{I}} &= \sum_{k,k',\boldsymbol{x},\boldsymbol{y}} |v_{k',\boldsymbol{y}}\rangle\langle v_{k',\boldsymbol{y}}|\boldsymbol{I}\otimes\boldsymbol{Z}_\ell|v_{k,\boldsymbol{x}}\rangle\langle v_{k,\boldsymbol{x}}| \\ &= \sum_{k,k',\boldsymbol{x}} |v_{k',\boldsymbol{x}}\rangle\langle v_{k',\boldsymbol{x}}|\boldsymbol{I}\otimes\boldsymbol{Z}_\ell|v_{k,\boldsymbol{x}}\rangle\langle v_{k,\boldsymbol{x}}|,\end{aligned}$$

where we applied Eq. (4.62) to drop the sum on $\boldsymbol{y}$. Thus we can rewrite the RHS of Eq. (4.68) as

$$\begin{aligned}(cont.) &= \sum_{\ell,k',k,\boldsymbol{x}} |v_{k,\boldsymbol{x}}\rangle\langle v_{k,\boldsymbol{x}}|\boldsymbol{\sigma}_\ell|v_{k',\boldsymbol{x}}\rangle\langle v_{k',\boldsymbol{x}}|\boldsymbol{\sigma}_\ell|v_{k,\boldsymbol{x}}\rangle\langle v_{k,\boldsymbol{x}}| \int_{-\infty}^{0} \gamma_\beta(\omega)\omega\left|\hat{f}_\mu(\omega - J_{\text{prop}}(k'-k))\right|^2 \mathrm{d}\omega \\ &\preceq -\Omega\left(\frac{J_{\text{prop}}}{\ln\beta}\right)\sum_{k,\boldsymbol{x}}\left|v_{k,\boldsymbol{x}}\right\rangle\left\langle v_{k,\boldsymbol{x}}\right| \max_{\ell\in[T]}\left|\left\langle v_{k,\boldsymbol{x}}\right|\boldsymbol{\sigma}_\ell\left|v_{k-1,\boldsymbol{x}}\right\rangle\right|^2 \\ &\preceq -\Omega\left(\frac{J_{\text{prop}}}{T\ln\beta}\right)\sum_{k\geq 1,\boldsymbol{x}}\left|v_{k,\boldsymbol{x}}\right\rangle\left\langle v_{k,\boldsymbol{x}}\right| \\ &\preceq -\Omega\left(\frac{J_{\text{prop}}}{T\ln\beta}\right)\left(\boldsymbol{P}_{\mathtt{I}} - \sum_{\boldsymbol{x}}\left|v_{0,\boldsymbol{x}}\right\rangle\left\langle v_{0,\boldsymbol{x}}\right|\right) = -\Omega(\frac{J_{\text{prop}}}{T\ln\beta})\cdot\boldsymbol{P}_{\mathtt{I}}(\boldsymbol{I}-\boldsymbol{P}_{\mathtt{II}}).\end{aligned}$$

The first line uses the orthogonality condition $\langle v_{k',\boldsymbol{y}}|v_{k,\boldsymbol{x}}\rangle = \delta_{kk'}\delta_{\boldsymbol{xy}}$, and the fact that the identical ν labels on $\boldsymbol{A}_\nu^{a\dagger}$ and $\boldsymbol{A}_\nu^a$ enforce that transitions from k' need to be to the same k on both sides. The second line uses the negativity of the half-integral to focus on cooling transitions (which includes $k \to k-1$) and evaluates the integral (which concentrates near $\omega \approx J_{\text{prop}}$). Lastly, we combine the above with the error bound on E to conclude the proof. ∎

Gradient from $\boldsymbol{H}_{\text{in}}$

The goal of this subsection is to prove

$$-\boldsymbol{P}'_{\mathtt{II}}\mathcal{L}^\dagger_{\mathtt{III}}[\boldsymbol{H}_{\mathtt{III}}]\boldsymbol{P}'_{\mathtt{II}} \succeq r_3\boldsymbol{P}'_{\mathtt{II}}(\boldsymbol{I}-\boldsymbol{P}_{\mathtt{III}}) - \epsilon_{3a}\boldsymbol{I}.$$

Here $\boldsymbol{P}'_{\mathtt{II}}$ is the projector onto perturbed low-energy eigenstates of $\boldsymbol{H}_{\mathtt{III}} = \boldsymbol{H}_{\mathtt{II}} + \boldsymbol{H}_{\text{in}}$, corresponding to

$$\boldsymbol{P}_{\mathtt{II}} = \sum_{\boldsymbol{x}\in\{0,1\}^n} |\eta_{\boldsymbol{x}}\rangle\langle\eta_{\boldsymbol{x}}|,$$

where

$$|\eta_{\boldsymbol{x}}\rangle = \sum_{t=0}^{T}\sqrt{\xi_t}\left|\eta_{\boldsymbol{x},t}\right\rangle = \sum_{t=0}^{T}\sqrt{\xi_t}U_t\cdots U_1|\boldsymbol{x}\rangle\otimes|C_t\rangle \quad \text{and} \qquad \xi_t = \frac{1}{2^T}\binom{T}{t}.$$

Recall our Definition 4.14.1 where given a circuit with L computational gates, we pad it in the beginning and the end with t_0 identity gates to make a total $T = 2t_0 + L$ gates. We can understand ξ_t as the probability from a symmetric binomial distribution $\text{Binom}(T, \frac{1}{2})$, which has substantial weight near the center where the interesting computation takes place.

Proposition 4.14.1 (Lower bound ξ_t in the center)**.** *Suppose $T = 2t_0 + L$ and $t_0 = cL^2$ are positive integers. Then we have*

$$\xi_t \geq \frac{\mathrm{e}^{-c/4}}{T+1} \qquad \text{for each} \quad t \in [t_0, T - t_0].$$

Proof. As a property of the binomial distribution $\mathrm{Binom}(T, \frac{1}{2})$, we have $\xi_t \geq \xi_{t_0}$ for all $t \in [t_0, T - t_0]$. Observe that

$$\binom{T}{t}^{-1} = (T+1)\int_0^1 x^t(1-x)^{T-t}\mathrm{d}x \leq (T+1)\Big(\frac{t}{T}\Big)^t\Big(1-\frac{t}{T}\Big)^{T-t},$$

where the inequality comes from the fact that $\arg\max_{x\in[0,1]} x^t(1-x)^{T-t} = t/T$. Then

$$\xi_{t_0} = \frac{1}{2^T}\binom{T}{t_0} \geq \frac{f(L)}{T+1},$$
$$\text{where} \quad f(L) = \frac{1}{2^T}\Big(\frac{T}{t_0}\Big)^{t_0}\Big(\frac{T}{T-t_0}\Big)^{T-t_0} = \Big(1+\frac{1}{2cL}\Big)^{cL^2}\Big(1-\frac{1}{2(cL+1)}\Big)^{cL^2+L}.$$

The last equality is obtained after plugging in $T = 2t_0 + L$, $t_0 = cL^2$ and simplifying. We can use the first-derivative test to check that $f(L)$ is monotonically decreasing, and so $f(L) \geq \lim_{L\to\infty} f(L) = \mathrm{e}^{-c/4}$. Hence, $\xi_t \geq \xi_{t_0} \geq f(L)/(T+1) \geq \mathrm{e}^{-c/4}/(T+1)$. ∎

Using the fact that $U_{t_j-1}\cdots U_1$ acts trivially on the j-th qubit (by definition of t_j), we see that $\boldsymbol{P}_{\mathtt{II}}\boldsymbol{H}_{\mathrm{in}}\boldsymbol{P}_{\mathtt{II}}$ is diagonal in the $|\eta_{\boldsymbol{x}}\rangle$ basis:

$$\begin{aligned}\langle\eta_{\boldsymbol{x}}|\boldsymbol{H}_{\mathrm{in}}|\eta_{\boldsymbol{y}}\rangle &= J_{\mathrm{in}}\sum_{t,t'=0}^{T}\sqrt{\xi_t\xi_{t'}}\langle\eta_{\boldsymbol{x},t}|\Big(\sum_{j=1}^{n} g_j|1\rangle\langle 1|_j \otimes |C_{t_j-1}\rangle\langle C_{t_j-1}|\Big)|\eta_{\boldsymbol{y},t'}\rangle \\ &= \delta_{\boldsymbol{x},\boldsymbol{y}}\cdot J_{\mathrm{in}}\sum_{j=1}^{n} x_j g_j \xi_{t_j-1}.\end{aligned}$$

Since $g_j = 1/\xi_{t_j-1}$ (see Definition 4.14.1), then the above implies that that $|\eta_{\boldsymbol{x}}\rangle$ are eigenstates of $\boldsymbol{P}_{\mathtt{II}}\boldsymbol{H}_{\mathrm{in}}\boldsymbol{P}_{\mathtt{II}}$ with eigenvalue $J_{\mathrm{in}}\cdot\mathrm{wt}(\boldsymbol{x})$, where $\mathrm{wt}(\boldsymbol{x})$ is the Hamming weight of bit string $\boldsymbol{x}$. While $|\eta_{\boldsymbol{x}}\rangle$ are eigenstates of $\boldsymbol{P}_{\mathtt{II}}\boldsymbol{H}_{\mathrm{in}}\boldsymbol{P}_{\mathtt{II}}$, unfortunately, only $|\eta_{\boldsymbol{0}}\rangle$ is an eigenstate of $\boldsymbol{H}_{\mathrm{in}}$; this will require an additional perturbation step to handle this off-block-diagonal effect.

Let $\boldsymbol{H}_{\mathtt{III}} = \tilde{\boldsymbol{H}}_{\mathtt{III}} + \boldsymbol{V}_{\mathtt{III}}$, where

$$\tilde{\boldsymbol{H}}_{\mathtt{III}} = \boldsymbol{H}_{\mathtt{II}} + \boldsymbol{P}_{\mathtt{II}}\boldsymbol{H}_{\mathrm{in}}\boldsymbol{P}_{\mathtt{II}} + \boldsymbol{P}_{\mathtt{II}}^{\perp}\boldsymbol{H}_{\mathrm{in}}\boldsymbol{P}_{\mathtt{II}}^{\perp} \quad \text{and} \quad \boldsymbol{V}_{\mathtt{III}} = \boldsymbol{P}_{\mathtt{II}}\boldsymbol{H}_{\mathrm{in}}\boldsymbol{P}_{\mathtt{II}}^{\perp} + \boldsymbol{P}_{\mathtt{II}}^{\perp}\boldsymbol{H}_{\mathrm{in}}\boldsymbol{P}_{\mathtt{II}}.$$

We will also denote $\tilde{\mathcal{L}}_{\mathtt{III}}$ as the thermal Lindbladian with respect to $\tilde{\boldsymbol{H}}_{\mathtt{III}}$.

We start by studying the gradient on $\tilde{\boldsymbol{H}}_{\mathtt{III}}$ and showing all its excited states have negative energy gradients. It suffices to focus on the states in $\boldsymbol{P}_{\mathtt{II}}$, which is a low-energy subspace of eigenstates of $\tilde{\boldsymbol{H}}_{\mathtt{III}}$ with excitation gap of at least $J_{\text{prop}} - 2\|\boldsymbol{H}_{\text{in}}\|$ from Weyl's inequality. The effective Hamiltonian in this subspace has a simple form of decoupled qubits, which we write as

$$\tilde{\boldsymbol{H}}_{\text{eff}} := \tilde{\boldsymbol{H}}_{\mathtt{III}}|_{\boldsymbol{P}_{\mathtt{II}}} = \boldsymbol{P}_{\mathtt{II}}\boldsymbol{H}_{\text{in}}\boldsymbol{P}_{\mathtt{II}} = J_{\text{in}} \sum_{\boldsymbol{x}} \text{wt}(\boldsymbol{x})|\eta_{\boldsymbol{x}}\rangle\langle\eta_{\boldsymbol{x}}| \equiv J_{\text{in}} \sum_{j=1}^{n} (\boldsymbol{I} - \boldsymbol{Z}_j^{\text{eff}})/2,$$

where $\boldsymbol{Z}_j^{\text{eff}}$ is the Pauli Z operator of a virtual qubit defined as $\boldsymbol{Z}_j^{\text{eff}}|\eta_{\boldsymbol{x}}\rangle = (-1)^{x_j}|\eta_{\boldsymbol{x}}\rangle$.

Observe that $\tilde{\boldsymbol{H}}_{\text{eff}}$ has eigenvalues that are integer multiples of J_{in}, and thus its Bohr-frequency gap is $\Delta_\nu(\tilde{\boldsymbol{H}}_{\text{eff}}) = J_{\text{in}}$. Hence, assuming $J_{\text{in}}/2 < J_{\text{prop}} - 2\|\boldsymbol{H}_{\text{in}}\|$, we can apply Lemma 4.12.6 to see that the gradient operator sandwiched by $\boldsymbol{P}_{\mathtt{II}}$ can be understood by fully restricting to the subspace:

$$\boldsymbol{P}_{\mathtt{II}}\tilde{\mathcal{L}}_{\mathtt{III}}^{\dagger}[\tilde{\boldsymbol{H}}_{\mathtt{III}}]\boldsymbol{P}_{\mathtt{II}} \overset{\tilde{E}_a}{\approx} \sum_{a\in S_0}\sum_{\nu\in B(\tilde{\boldsymbol{H}}_{\text{eff}})} \boldsymbol{P}_{\mathtt{II}}\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{P}_{\mathtt{II}}\boldsymbol{A}_\nu^{a}\boldsymbol{P}_{\mathtt{II}} \int_{-\infty}^{0} \gamma_\beta(\omega)\omega|\hat{f}_\mu(\omega-\nu)|^2 \mathrm{d}\omega, \tag{4.77}$$

where $\mu = J_{\text{in}}/2$ and

$$\tilde{E}_a = |S_0|O\left(\frac{\|\tilde{\boldsymbol{H}}_{\mathtt{III}}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta} + \frac{1}{\sqrt{J_{\text{in}}\tau}}\right).$$

To show that this has good gradients for all states in $\boldsymbol{P}_{\mathtt{II}}(\boldsymbol{I} - \boldsymbol{P}_{\mathtt{III}})$, it is sufficient to consider the following subset of the jump operators from Eq. (4.52):

$$S_{\mathtt{III}} = \left\{\boldsymbol{X}_j \otimes |0\rangle\langle 0|_{t_j}\right\}_{j=1}^{n}.$$

These jump operators from $S_{\mathtt{III}}$ effectively flip the individual virtual qubits by

$$\langle\eta_{\boldsymbol{y}}|(\boldsymbol{X}_j \otimes |0\rangle\langle 0|_{t_j})|\eta_{\boldsymbol{x}}\rangle = \langle \boldsymbol{y}|\boldsymbol{X}_j|\boldsymbol{x}\rangle \textstyle\sum_{t=0}^{t_j-1} \xi_t =: \langle \boldsymbol{y}|\boldsymbol{X}_j|\boldsymbol{x}\rangle\sqrt{\alpha_j},$$

where we have denoted $\alpha_j = (\sum_{t<t_j} \xi_t)^2$. Let

$$\boldsymbol{X}_j^{\text{eff}} = \boldsymbol{P}_{\mathtt{II}}(\boldsymbol{X}_j \otimes |0\rangle\langle 0|_{t_j})\boldsymbol{P}_{\mathtt{II}}/\sqrt{\alpha_j} \quad \text{such that} \quad \|\boldsymbol{X}_j^{\text{eff}}\| = 1 \quad \text{for each} \quad j = 1, \ldots, n.$$

Note $\boldsymbol{X}_j^{\text{eff}}$ is effectively the Pauli X operator for the j-th virtual qubit. Furthermore, note since $t_j \in [t_0, T - t_0]$ by our circuit construction in Definition 4.14.1, we have $\alpha_j \geq \xi_{t_0}^2 \geq \Omega(1/T^2)$ by Proposition 4.14.1.

Next, we replace $\tilde{\boldsymbol{H}}_{\mathtt{III}}$ by $\tilde{\boldsymbol{H}}_{\text{eff}}$ so we only need to talk about the virtual qubits. We first restore the RHS of Eq. (4.77) to the thermal Lindbladian form by undoing the approximation in Lemma 4.12.6, which incurs another error bounded by $\tilde{E}_a$:

$$\boldsymbol{P}_{\mathtt{II}}\tilde{\mathcal{L}}^{\dagger}_{\mathtt{III}}[\tilde{\boldsymbol{H}}_{\mathtt{III}}]\boldsymbol{P}_{\mathtt{II}} \overset{2\tilde{E}_a}{\approx} \mathcal{L}^{\dagger\beta,\tau,\boldsymbol{H}_{\text{eff}}}[\boldsymbol{H}_{\text{eff}}].$$

We then focus on the subset $S_{\mathtt{III}}$ of jump operators and write

$$\mathcal{L}^{\dagger\beta,\tau,\boldsymbol{H}_{\text{eff}}}[\boldsymbol{H}_{\text{eff}}] \overset{\tilde{E}_b}{\approx} \textstyle\sum_{j=1}^{n} \alpha_j \tilde{\mathcal{L}}_j^{\dagger}[\tilde{\boldsymbol{H}}_{\text{eff}}], \tag{4.78}$$

where we denoted $\tilde{\mathcal{L}}_j := \mathcal{L}_j^{\beta,\tau,\tilde{\boldsymbol{H}}_{\text{eff}}}$ to be the thermal Lindbladian associated with effective jump operator $\boldsymbol{X}_j^{\text{eff}}$ and pulled out the normalization factor α_j. The error

$$\tilde{E}_b = |S_0| O\left(\frac{\|\tilde{\boldsymbol{H}}_{\text{eff}}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta}\right)$$

comes from neglecting the other jump operators $S_0 \setminus S_{\mathtt{III}}$, and is bounded using Lemma 4.12.1.

Since the effective operators on different virtual qubits commute, we can treat them independently. More formally, by Lemma 4.12.8 we have $\tilde{\mathcal{L}}_j^{\dagger}[\tilde{\boldsymbol{H}}_{\text{eff}}] = \mathcal{L}_j^{\dagger\beta,\tau,\boldsymbol{h}_j}[\boldsymbol{h}_j]$, where $\boldsymbol{h}_j = J_{\text{in}}(\boldsymbol{I} - \boldsymbol{Z}_{\text{eff}}^j)/2$. To bound the global gradient in Eq. (4.78), we first consider cooling a single qubit.

Lemma 4.14.5 (Cooling a qubit)**.** *On a qubit, consider the thermal Lindbladian $\mathcal{L} = \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$ with the Hamiltonian $\boldsymbol{H} = J_{\text{in}}(\boldsymbol{I} - \boldsymbol{Z})/2$ and one jump operator $\boldsymbol{A} = \boldsymbol{X}$. Then,*

$$-\mathcal{L}^{\dagger}[\boldsymbol{H}] \succeq r_{\text{in}}(\boldsymbol{I} - \boldsymbol{Z}) - \epsilon_{\text{in}}\boldsymbol{I},$$

where

$$r_{\text{in}} = \Omega\left(\frac{J_{\text{in}}}{\ln\beta}\right) \quad \textit{and} \quad \epsilon_{\text{in}} = O\left(\frac{J_{\text{in}}^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\sqrt{\tau J_{\text{in}}}} + \mathrm{e}^{-\beta J_{\text{in}}}\right).$$

Proof. Again, we invoke a series of approximations:

$$\begin{aligned}
\mathcal{L}^{\dagger}[\boldsymbol{H}] &\approx \mathcal{D}^{\dagger}[\boldsymbol{H}] \\
&\approx \int_{-\infty}^{\infty} \gamma_\beta(\omega)\omega \boldsymbol{A}(\omega)^{\dagger}\boldsymbol{A}(\omega)\mathrm{d}\omega \\
&\approx \int_{-\infty}^{\infty} \gamma_\beta(\omega)\omega \boldsymbol{S}(\omega)^{\dagger}\boldsymbol{S}(\omega)\mathrm{d}\omega \\
&= \sum_{\nu\in B(\boldsymbol{H})} \boldsymbol{A}_\nu^{\dagger}\boldsymbol{A}_\nu \int_{-\infty}^{\infty} \gamma_\beta(\omega)\omega \left|\hat{f}_\mu(\omega-\nu)\right|^2 \mathrm{d}\omega \\
&\preceq -\Omega(J_{\text{in}}/\ln\beta)|1\rangle\langle 1| + O(\mathrm{e}^{-\beta J_{\text{in}}})|0\rangle\langle 0|.
\end{aligned}$$

The last line uses the transition matrix elements for the two Bohr frequencies $\nu = \pm J_{\text{in}}$:

$$\boldsymbol{A}_{+1} = |1\rangle\langle 0|, \qquad \text{and} \qquad \boldsymbol{A}_{-1} = |0\rangle\langle 1|.$$

Combine the error bound to conclude the proof. ∎

Summing up over the contributions from the individual qubits, the global gradient satisfies

$$\begin{aligned} -\boldsymbol{P}_{\text{II}}\tilde{\mathcal{L}}^{\dagger}_{\text{III}}[\tilde{\boldsymbol{H}}_{\text{III}}]\boldsymbol{P}_{\text{II}} &\geq \sum_{j=1}^{n} \alpha_j \boldsymbol{P}_{\text{II}}\Big[r_{\text{in}}(\boldsymbol{I} - \boldsymbol{Z}_j^{\text{eff}}) - \epsilon_{\text{in}}\boldsymbol{I}\Big]\boldsymbol{P}_{\text{II}} - (2\tilde{E}_a + \tilde{E}_b)\boldsymbol{I} \\ &\geq r_3 \boldsymbol{P}_{\text{II}}(\boldsymbol{I} - \boldsymbol{P}_{\text{III}}) - \tilde{\epsilon}_3 \boldsymbol{I}, \end{aligned}$$

where we used the fact that $\sum_{j=1}^{n}(\boldsymbol{I} - \boldsymbol{Z}_j^{\text{eff}}) \geq \boldsymbol{P}_{\text{III}}$, and denoted $r_3 = r_{\text{in}} \min_j \alpha_j$, and $\tilde{\epsilon}_3 = 2\tilde{E}_a + \tilde{E}_b + \epsilon_{\text{in}} \sum_{j=1}^{n} \alpha_j$. Since $\alpha_j = \Omega(1/T^2)$ and $\alpha_j \leq 1$, we have that

$$r_3 = \Omega\Big(\frac{J_{\text{in}}}{T^2 \ln \beta}\Big) \quad \text{and} \quad \tilde{\epsilon}_3 \leq 2\tilde{E}_a + \tilde{E}_b + n\epsilon_{\text{in}} = |S_0| O\Big(\frac{\|\tilde{\boldsymbol{H}}_{\text{III}}\|^{3/4}}{\tau^{1/4}} + \frac{1}{\tau} + \frac{1}{\beta} + \frac{1}{\sqrt{J_{\text{in}}\tau}} + \mathrm{e}^{-\beta J_{\text{in}}}\Big). \tag{4.85}$$

Lastly, to obtain the gradient for the final Hamiltonian $\boldsymbol{H}_{\text{III}}$, we need to add the off-block-diagonal perturbation and show that the gradient persists on the subspace. Directly applying subspace monotonicity (Corollary 4.12.1) yields loose bounds; we will need to invoke a finer-grained subspace monotonicity (Corollary 4.12.2) that exploits the fact that $\boldsymbol{V}$ is off-block-diagonal and contribute to eigenvalue change at *second order* $O(\|\boldsymbol{V}\|^2)$. We apply Corollary 4.12.2 with $\boldsymbol{H} = \tilde{\boldsymbol{H}}_{\text{III}}$ and $\boldsymbol{H}' = \boldsymbol{H}_{\text{III}}$, and parameters

$$\begin{aligned} \boldsymbol{Q} &= \boldsymbol{P}_{\text{II}} \\ \Delta_{\boldsymbol{Q}} &= J_{\text{prop}} - 2\|\boldsymbol{H}_{\text{in}}\| \\ \boldsymbol{V} &= \boldsymbol{P}_{\text{II}}\boldsymbol{H}_{\text{in}}(\boldsymbol{I} - \boldsymbol{P}_{\text{II}}) + (\boldsymbol{I} - \boldsymbol{P}_{\text{II}})\boldsymbol{H}_{\text{in}}\boldsymbol{P}_{\text{II}} \\ \Delta_{\nu} &= J_{\text{in}} \end{aligned}$$

Therefore, by Corollary 4.12.2,

$$-\boldsymbol{P}'_{\text{II}}\mathcal{L}^{\dagger}_{\text{III}}[\boldsymbol{H}_{\text{III}}]\boldsymbol{P}'_{\text{II}} \geq r_3 \boldsymbol{P}'_{\text{II}}(\boldsymbol{I} - \boldsymbol{P}_{\text{III}}) - \epsilon_{3a}\boldsymbol{I},$$

where

$$\begin{aligned} \epsilon_{3a} \leq \tilde{\epsilon}_3 + |S_0| \cdot O\Big(&\frac{1}{\tau} + \frac{\|\tilde{\boldsymbol{H}}_{\text{III}}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu \tau}} + \frac{\Lambda_0}{\sqrt{\Delta_{\boldsymbol{Q}}\tau}} + \frac{\mathrm{e}^{-\beta\Delta_{\boldsymbol{Q}}/4}}{\beta} \\ &+ \frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}} + \|\boldsymbol{H}_{\boldsymbol{Q}}\| \cdot \Big(\frac{\|\boldsymbol{H}_{\boldsymbol{Q}}\|\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}\Delta_\nu} + \frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}\Delta_\nu}\Big) + r_3\Big(\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}} + \frac{\|\boldsymbol{V}\|^2}{\Delta_\nu\Delta_{\boldsymbol{Q}}}\Big)\Big). \end{aligned}$$

Noting that $|S_0|, \|\tilde{\boldsymbol{H}}_{\mathtt{III}}\| = O(T)$, $\|\boldsymbol{V}\| = O(nTJ_{\text{in}})$, $\|\boldsymbol{H}_{\boldsymbol{Q}}\| = nJ_{\text{in}}$, and $J_{\text{in}} \ll J_{\text{prop}}$, we simplify the error bound above by keeping the dominant term to get

$$\epsilon_{3a} \leq TO\left(\frac{T^{3/4}}{\tau^{1/4}} + \frac{1}{\beta} + \frac{1}{\sqrt{J_{\text{in}}\tau}} + \mathrm{e}^{-\beta J_{\text{in}}} + n\frac{(nTJ_{\text{in}})^2}{J_{\text{prop}}}\right) \tag{4.91}$$

as advertised earlier in Eq. (4.59).

4.15 Appendix:Operator Fourier Transform

Recall that the exact form of thermal Lindbladians in Appendix 4.10 involves the operator Fourier transform (OFT) [39] for a set of jump operators $\boldsymbol{A}^a$. In this appendix, we provide the key properties of OFT which are used in the proofs of many statements in Appendices 4.12 and 4.16. For any operator $\boldsymbol{A}$ (the jump operator), and Hermitian operator $\boldsymbol{H}$ (the Hamiltonian), and a weight function f, the *operator Fourier Transform* (OFT) is an integral over time-evolution of the operator $\boldsymbol{A}$,

$$\hat{\boldsymbol{A}}_{\hat{f}}(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A} \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} \mathrm{e}^{-\mathrm{i}\omega t} f(t) \mathrm{d}t.$$

Often, we will also write $\hat{\boldsymbol{A}}(\omega)$ when we choose f to be the default *normalized window function*

$$f_\tau(t) = \frac{1}{\sqrt{\tau}} \cdot \begin{cases} 1 & \text{if} \quad t \in [-\tau/2, \tau/2] \\ 0 & \text{else.} \end{cases} \tag{4.92}$$

It is usually helpful to consider the energy eigenspaces $\boldsymbol{H} = \sum_i E_i |\psi_i\rangle\langle\psi_i| = \sum_{E \in \text{Spec}(\boldsymbol{H})} E \boldsymbol{P}_E$ and and write $\boldsymbol{A}$ as the following decomposition

$$\boldsymbol{A} = \sum_{E_2, E_1 \in \text{Spec}(\boldsymbol{H})} \boldsymbol{P}_{E_2} \boldsymbol{A} \boldsymbol{P}_{E_1} = \sum_{\nu \in B(\boldsymbol{H})} \boldsymbol{A}_\nu \quad \text{where} \quad \boldsymbol{A}_\nu := \sum_{E_2 - E_1 = \nu} \boldsymbol{P}_{E_2} \boldsymbol{A} \boldsymbol{P}_{E_1}.$$

Formally, these energy differences $\nu \in B(\boldsymbol{H}) := \{E_i - E_j \,|\, E_i, E_j \in \text{Spec}(\boldsymbol{H})\}$ are called the *Bohr frequencies* (Figure 4.2), and $\boldsymbol{A}_\nu$ collects the matrix elements $\langle\psi_i|\boldsymbol{A}|\psi_j\rangle$ that changes the energy by $\nu = E_i - E_j$. The Bohr frequencies are natural for the Heisenberg-picture evolution of $\boldsymbol{A}$ since

$$\mathrm{e}^{\mathrm{i}\boldsymbol{H}t} \boldsymbol{A} \mathrm{e}^{-\mathrm{i}\boldsymbol{H}t} = \sum_{\nu \in B(\boldsymbol{H})} \mathrm{e}^{\mathrm{i}\nu t} \boldsymbol{A}_\nu.$$

Then, executing the Fourier integral yields the frequency domain representation

$$\hat{\boldsymbol{A}}_f(\omega) = \sum_{\nu \in B} \boldsymbol{A}_\nu \hat{f}(\omega - \nu) \quad \text{where} \quad \hat{f}(\omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t,$$

which contains a collection of Bohr frequencies ν near ω. Conceptually, we can think of the operator Fourier transform $\hat{\boldsymbol{A}}_f(\omega)$ as the *smooth* probe of $\boldsymbol{A}_\nu$ with exact Bohr frequency ν, which generally requires resolving arbitrarily close eigenvalues.

Useful properties

We instantiate some useful properties of the Operator Fourier Transform.

Proposition 4.15.1 (Operator Parseval's identity [39])**.** *Consider a set of operators* $\{\boldsymbol{A}^a\}_a$ *and its operator Fourier transform with weight* $f \in \mathcal{L}_2(\mathbb{R})$ *and Hamiltonian* $\boldsymbol{H}$*. Then, we have a certain symmetry* $(\boldsymbol{A}_f^a(\omega))^\dagger = \boldsymbol{A}_{f^*}^{a\dagger}(-\omega)$ *and certain Parseval-type identity*

$$\begin{aligned}\sum_{a\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}_f^a(\omega)^\dagger\hat{\boldsymbol{A}}_f^a(\omega)\mathrm{d}\omega &= \sum_{a\in S}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}|f(t)|^2\mathrm{d}t\\ &\leq \|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\|f\|_2^2\cdot\boldsymbol{I}\\ \sum_{a\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}_f^a(\omega)\hat{\boldsymbol{A}}_f^a(\omega)^\dagger\mathrm{d}\omega &= \sum_{a\in S}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^{a}\boldsymbol{A}^{a\dagger}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}|f(t)|^2\mathrm{d}t\\ &\leq \|\sum_{a\in S}\boldsymbol{A}^{a}\boldsymbol{A}^{a\dagger}\|\|f\|_2^2\cdot\boldsymbol{I}.\end{aligned}$$

Intuitively, the above tells us that the Fourier Transforms $\hat{\boldsymbol{A}}_f^a(\omega)$ with different frequencies ω are "orthogonal" to each other and that the average of squares of strengths is bounded (reminiscent of a probability). Without using this norm sum constraint from Fourier transforms, one easily gets loose bounds. An alternative view of the above (as a natural purification) will prove useful for manipulating norms of expression involving $\boldsymbol{A}_f^a(\omega)$.

Corollary 4.15.1 (Purification of Operator Fourier Transform)**.** *In the prevailing notation, the abstract operator Fourier Transform has a norm bound,*

$$\|\sum_{a\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}_f^a(\omega)\otimes|a\rangle\otimes|\omega\rangle\mathrm{d}\omega\| \leq \|f\|_2\sqrt{\|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|},$$

where the continuous basis vectors satisfy the normalization

$$\langle\omega'|\omega\rangle = \delta(\omega'-\omega).$$

Proof. We multiply the conjugate

$$\begin{aligned}&\left(\sum_{a'\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}_f^{a'}(\omega')^\dagger\otimes\langle a'|\otimes\langle\omega'|\mathrm{d}\omega'\right)\cdot\left(\sum_{a\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}_f^a(\omega)\otimes|a\rangle\otimes|\omega\rangle\mathrm{d}\omega\right)\\ &=\sum_{a\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}_f^a(\omega)^\dagger\hat{\boldsymbol{A}}_f^a(\omega)\mathrm{d}\omega.\end{aligned}$$

Take the operator norm and use Proposition 4.15.1 to conclude the proof. ∎

The above "purification" trick applies to other quantities with a liberal choice of summation indices, whether they are a, ω, or ν.

Lemma 4.15.1 (Norm inequalities from operator purification)**.** *For any operator $\boldsymbol{O}$ and any set of operators $\boldsymbol{A}_i, \boldsymbol{A}'_j$ acting on the same Hilbert space, we have that*

$$\left\|\sum_{i,j} \boldsymbol{A}_i^\dagger \boldsymbol{O} \boldsymbol{A}'_j G_{ij}\right\| \leq \|\boldsymbol{G}\|\|\boldsymbol{O}\| \sqrt{\left\|\sum_i \boldsymbol{A}_i^\dagger \boldsymbol{A}_i\right\| \left\|\sum_j \boldsymbol{A}'^\dagger_j \boldsymbol{A}'_j\right\|}$$

$$\left\|\sum_{i,j} \boldsymbol{A}_i^\dagger \boldsymbol{A}'_j G_{ij}\right\| \leq \|\boldsymbol{G}\| \sqrt{\left\|\sum_i \boldsymbol{A}_i^\dagger \boldsymbol{A}_i\right\| \left\|\sum_j \boldsymbol{A}'^\dagger_j \boldsymbol{A}'_j\right\|}$$

Proof. By homogeneity, it suffices to set normalization to be $\|\sum_j \boldsymbol{A}'^\dagger_j \boldsymbol{A}'_j\| = \|\sum_i \boldsymbol{A}_i^\dagger \boldsymbol{A}_i\| = \|\boldsymbol{O}\| = \|\boldsymbol{G}\| = 1$. Introduce purifications

$$\boldsymbol{G}' := \boldsymbol{I} \otimes \sum_{i,j} G_{ij} |i\rangle\langle j|, \qquad \boldsymbol{O}' := \boldsymbol{O} \otimes \boldsymbol{I},$$
$$\boldsymbol{V}' := \sum_j \boldsymbol{A}'_j \otimes |j\rangle, \qquad \boldsymbol{V} := \sum_i \boldsymbol{A}_i \otimes |i\rangle,$$

which are all bounded by $\|\boldsymbol{G}'\|, \|\boldsymbol{O}'\|, \|\boldsymbol{V}'\|, \|\boldsymbol{V}\| \leq 1$. Then,

$$\left\| \sum_{i,j} \boldsymbol{A}_i^\dagger \boldsymbol{O} \boldsymbol{A}'_j G_{ij} \right\| = \|\boldsymbol{V}' \boldsymbol{O}' \boldsymbol{G}' \boldsymbol{V}\| \leq 1$$
$$\left\| \sum_{i,j} \boldsymbol{A}_i^\dagger \boldsymbol{A}'_j G_{ij} \right\| = \|\boldsymbol{V}' \boldsymbol{G}' \boldsymbol{V}\| \leq 1.$$

Rescale to obtain the advertised result. ∎

Secular approximation

Due to the energy-time uncertainty principle, the energies ω that are accessed by finite-time quantum algorithms always inherit an uncertainty. Indeed, when we choose our weight function $f_\tau(t)$ in Eq. (4.92) for our operator Fourier transform in the frequency domain, we have

$$\hat{\boldsymbol{A}}_f(\omega) = \sum_{\nu \in B} \boldsymbol{A}_\nu \hat{f}(\omega - \nu) \quad \text{where} \quad \hat{f}(\omega) = \frac{e^{i\omega\tau/2} - e^{-i\omega\tau/2}}{i\omega\sqrt{2\pi\tau}} \quad \text{when} \quad f(t) = f_\tau(t).$$

Note with this choice, $\hat{f}(\omega)$ has a heavy tail $\sim 1/\omega$, which is reminiscent of unamplified phase estimation. Therefore, even when restricting to jumps with $\omega < 0$,

there is a decent chance that $\hat{\boldsymbol{A}}_f(\omega)$ mistakenly activates a heating transition ($\nu > 0$) instead of a cooling transition ($\nu < 0$), unintentionally heating up the system instead of cooling it.

To control the resulting error, in this section, we introduce the *secular approximation* [39] of the Fourier transformed operators $\boldsymbol{A}_f(\omega)$. The secular approximation applies truncation to the Fourier-transformed operators in the frequency domain by truncating Bohr frequencies $E \in B$ that deviate substantially from the frequency label ω. Truncation at energy difference μ can be achieved by setting a step function and defining the following secular-approximated operators as follows:

$$\hat{\boldsymbol{S}}_{f,\mu}(\omega) := \sum_{\nu \in B} \boldsymbol{A}_\nu \hat{f}(\omega - \nu) \cdot \hat{s}_\mu(\omega - \nu) \quad \text{where} \quad \hat{s}_\mu(\omega) := \mathbb{1}(|\omega| < \mu).$$

We often drop subscript f, μ for simplicity. The truncation error is an operator whose norm can be bounded by the following (as a variant of Corollary 4.15.1).

Corollary 4.15.2 (Secular approximation)**.** *In the prevailing notation,*

$$\|\sum_{a \in S} \int_{-\infty}^{\infty} (\hat{\boldsymbol{S}}^a_f(\omega) - \hat{\boldsymbol{A}}^a_f(\omega)) \otimes |a\rangle \otimes |\omega\rangle \mathrm{d}\omega\| \leq \|\hat{f} \cdot (1 - \hat{s}_\mu)\|_2 \sqrt{\|\sum_{a \in S} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a\|}.$$

The error is controlled by the 2-norm for the truncated tail

$$\|\hat{f} \cdot (1 - \hat{s}_\mu)\|_2^2 = \int_{|\omega| \geq \mu} |\hat{f}(\omega)|^2 \mathrm{d}\omega$$

noting that conveniently, the Fourier transform preserves the 2-norm of functions. For our bump function (4.92) in particular, we can integrate the tail-bound

$$\|\hat{f}_\tau \cdot (1 - \hat{s}_\mu)\|_2^2 \leq \frac{4}{\pi \mu \tau}. \tag{4.93}$$

This conveniently leads to bounds on other quantities involving operator Fourier Transforms $\hat{\boldsymbol{A}}^a(\omega)$.

Corollary 4.15.3 (Error from secular approximation)**.** *For any real function $\theta : \mathbb{R} \to \mathbb{R}$,*

$$\left\| \sum_{a \in S} \int_{-\infty}^{\infty} \theta(\omega') \hat{\boldsymbol{A}}^a_f(\omega')^\dagger \hat{\boldsymbol{A}}^a_f(\omega') \mathrm{d}\omega' - \sum_{a \in S} \int_{-\infty}^{\infty} \theta(\omega') \hat{\boldsymbol{S}}^a_{f,\mu}(\omega')^\dagger \hat{\boldsymbol{S}}^a_{f,\mu}(\omega') \mathrm{d}\omega' \right\|$$
$$\leq 2\|\theta\|_\infty \|\hat{f} \cdot (1 - \hat{s}_\mu)\|_2 \|f\|_2 \sqrt{\|\sum_{a \in S} \boldsymbol{A}^{a\dagger} \boldsymbol{A}^a\|}.$$

Proof. It suffices to set normalization $\|\sum_{a\in S} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| = 1$. Introduce purifications

$$\begin{aligned}
\boldsymbol{F} &:= \boldsymbol{I} \otimes \boldsymbol{I} \otimes \int_{-\infty}^{\infty} \theta(\omega)|\omega\rangle\langle\omega|\mathrm{d}\omega, \\
\boldsymbol{V} &:= \left(\sum_{a\in S} \int_{-\infty}^{\infty} \hat{\boldsymbol{A}}^a_f(\omega) \otimes |a\rangle \otimes |\omega\rangle \mathrm{d}\omega\right), \\
\boldsymbol{V}' &:= \left(\sum_{a\in S} \int_{-\infty}^{\infty} \hat{\boldsymbol{S}}^a_f(\omega) \otimes |a\rangle \otimes |\omega\rangle \mathrm{d}\omega\right).
\end{aligned}$$

Then, by a telescoping sum,

$$(LHS) = \|\boldsymbol{V}^\dagger \boldsymbol{F}\boldsymbol{V} - \boldsymbol{V}'^\dagger \boldsymbol{F}\boldsymbol{V}'\| \leq \|(\boldsymbol{V}^\dagger - \boldsymbol{V}'^\dagger)\boldsymbol{F}\boldsymbol{V}\| + \|\boldsymbol{V}'^\dagger \boldsymbol{F}(\boldsymbol{V} - \boldsymbol{V}')\|.$$

We conclude the proof using the individual bounds on the operator norm

$$\begin{aligned}
\|\boldsymbol{V}\| &\leq \|f\|_2 \\
\|\boldsymbol{V}'\| &\leq \|\hat{f} \cdot \hat{s}_\mu\|_2 \leq \|f\|_2 \\
\|\boldsymbol{F}\| &\leq \|\theta\|_\infty \\
\|\boldsymbol{V}' - \boldsymbol{V}\| &\leq \|\hat{f} \cdot (1 - \hat{s}_\mu)\|_2 \quad .
\end{aligned}$$

This concludes the proof. ∎

4.16 Appendix:Proving monotonicity of energy gradient under level splitting

It will often be helpful to understand how the energy gradients of a Hamiltonian $\boldsymbol{H}$ change when a perturbation $\boldsymbol{V}$ is added to yield $\boldsymbol{H}' = \boldsymbol{H} + \boldsymbol{V}$. This allows us to characterize the energy gradient of $\boldsymbol{H}'$ by analyzing the unperturbed spectrum of $\boldsymbol{H}$, which is usually much simpler than that of $\boldsymbol{H}'$. Indeed, this is an important part of our proof strategy for showing our key result that $\boldsymbol{H}_C$ has no suboptimal local minima (Theorem 4.14.1 in section 4.14).

The relationship we can prove, which was previously stated in Theorem 4.12.1 in section 4.12, takes the form of *monotonicity*. As the name implies, the result only holds in one direction; it fails when the $\boldsymbol{H}'$ and $\boldsymbol{H}$ are switched. It is imperative that $\boldsymbol{H}$ have a highly degenerate spectrum with the *Bohr-frequency gap*

$$\Delta_\nu(\boldsymbol{H}) := \min_{\nu_1 \neq \nu_2 \in B(\boldsymbol{H})} |\nu_1 - \nu_2|,$$

which sets an energy scale for which $\boldsymbol{V}$ is perturbative. Note that the Bohr-frequency gap is upper bounded by the gap in the spectrum[10]

$$\Delta_\nu \le \Delta_E \quad \text{where} \quad \Delta_E = \min_{E_1 \neq E_2 \in \mathrm{Spec}(\boldsymbol{H})} |E_1 - E_2|.$$

We now state a more general version of Theorem 4.12.1, which we prove in the remainder of this appendix.

Theorem 4.16.1 (Monotonicity of gradient under level splitting, expanded version)**.** *Consider a Hamiltonian $\boldsymbol{H}$ with a highly degenerate spectrum and Bohr-frequency gap $\Delta_\nu := \min_{\nu_1 \neq \nu_2 \in B(\boldsymbol{H})} |\nu_1 - \nu_2|$, and a perturbed Hamiltonian $\boldsymbol{H}' = \boldsymbol{H} + \boldsymbol{V}$. Suppose the perturbation is weaker than the Bohr-frequency gap, $\|\boldsymbol{V}\| \le \frac{1}{8}\Delta_\nu$. For any $\beta, \tau > 0$, let $\mathcal{L} = \sum_{a \in S} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$, $\mathcal{L}' = \sum_{a \in S} \mathcal{L}_a^{\beta,\tau,\boldsymbol{H}'}$ be thermal Lindbladians with jumps $\{\boldsymbol{A}^a\}_{a \in S}$, where $\|\boldsymbol{A}^a\| \le 1$ and the transition weight $\gamma_\beta(\omega)$ is given by Eq. (4.21). Let $\delta_\lambda = \max_j |\lambda_j(\boldsymbol{H}) - \lambda_j(\boldsymbol{H}')|$, where $\lambda_j(\boldsymbol{X})$ is the j-th largest eigenvalue of $\boldsymbol{X}$, and let $\theta_{\max} = \max_{\nu \in B(\boldsymbol{H})} |\nu \gamma_\beta(\nu) \mathbb{1}(\nu \le \Delta_\nu/2)|$. For any two operators $\boldsymbol{O}$ and $\boldsymbol{O}'$, where $[\boldsymbol{O}', \boldsymbol{H}'] = 0$, we have the monotone property that*

$$-\mathcal{L}^\dagger[\boldsymbol{H}] \ge r\boldsymbol{O} - \epsilon \boldsymbol{I} \quad \textit{implies} \quad -\mathcal{L}'^\dagger[\boldsymbol{H}'] \ge r\boldsymbol{O}' - \epsilon' \boldsymbol{I},$$

where

$$\epsilon' \le \epsilon + |S| \cdot O\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu \tau}} + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta} + \delta_\lambda + \theta_{\max}\frac{\|\boldsymbol{V}\|}{\Delta_\nu} + r\|\boldsymbol{O} - \boldsymbol{O}'\|\right).$$

For the special case of $\boldsymbol{O} = \boldsymbol{I} - \boldsymbol{P}$ and $\boldsymbol{O}' = \boldsymbol{I} - \boldsymbol{P}'$, where $\boldsymbol{P}$ projects onto an eigensubspace of $\boldsymbol{H}$, and $\boldsymbol{P}'$ projects onto the corresponding perturbed eigensubspace in $\boldsymbol{H}'$, then we have the following simpler error bound:

$$\epsilon' \le \epsilon + |S| \cdot O\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu \tau}} + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta} + \big(1 + \frac{\Lambda_0 + r}{\Delta_\nu}\big)\|\boldsymbol{V}\|\right).$$

The above result is nontrivial because naïve perturbation theory fails: the Lindbladian depends sensitively on the perturbation $\boldsymbol{V}$ (as it uses a long Hamiltonian simulation time $\tau\|\boldsymbol{V}\| \gg 1$). In fact, it drastically fails if the energy spectrum of $\boldsymbol{H}$ has a (nearly) continuous spectrum (as the opposite of the premise of gapped degenerate subspaces).

[10] The Bohr-frequency gap can be much smaller than the eigenvalue gap. For example, consider the energies $\{-0.99, 0, 1\}$, which has an eigenvalue gap of 0.99 and a Bohr-frequency gap of 0.01.

We can understand the energy scale associated with the minimum Bohr-frequency gap Δ_ν as the meaningful quantity for which $\boldsymbol{V}$ is a perturbation

$$\frac{1}{\tau} \ll \|\boldsymbol{V}\| \ll \Delta_\nu.$$

Otherwise, the $1/\tau$ energy resolution is too small compared to the intended perturbation $\|\boldsymbol{V}\|$.

The proof of Theorem 4.16.1 will be quite involved. Technically, we heavily utilize the manipulations using the operator Fourier Transform (Appendix 4.15). The key subroutines of the proof are discussed separately as follows. First, in Appendix 4.16, we will simplify the intimidating expression for energy gradient $\mathcal{L}^\dagger[\boldsymbol{H}]$. Secondly, in Appendix 4.16, we isolate the key nonperturbative argument, which roughly says level splitting only improves the gradient. We then provide some results from perturbation theory in section 4.16. The altogether proof is presented in Appendix 4.16, with minor supporting calculations in section 4.16. We also prove two corollaries of Theorem 4.16.1 that apply to subspace gradients in Appendix 4.16.

Since this appendix only consider thermal Lindbladians, in what follows we will drop the superscripts $\beta, \tau, \boldsymbol{H}$, i.e. $\mathcal{L} \equiv \mathcal{L}^{\beta,\tau,\boldsymbol{H}}$, $\mathcal{L}' \equiv \mathcal{L}^{\beta,\tau,\boldsymbol{H}'}$, $\mathcal{D} \equiv \mathcal{D}^{\beta,\tau,\boldsymbol{H}}$, etc.

Expressing the energy gradient

The thermal Lindbladian is quite cumbersome to manipulate. Nicely, the energy gradient operator associated with the dissipative part $\mathcal{D}^\dagger[\boldsymbol{H}]$ permits a much simpler approximate form up to a controllable error. Combining with error bounds on the Lamb-shift term $[\boldsymbol{H}_{LS}, \boldsymbol{H}]$ (Proposition 4.10.3) allows us to approximate the full gradient operator $\mathcal{L}^\dagger[\boldsymbol{H}]$.

Lemma 4.16.1 (Expression for energy gradient)**.** *Consider the operator Fourier Transforms $\hat{\boldsymbol{A}}^a(\omega)$ weighted by the bump function f_τ in Eq. (4.92) with Hamiltonian $\boldsymbol{H}$. Then, for any Fourier transform pairs $\gamma(\omega)$ and $c(t)$, the energy gradient associated with the purely dissipative Lindbladian*

$$\mathcal{D}^\dagger[\boldsymbol{H}] = \sum_{a\in S} \int_{-\infty}^{\infty} \gamma(\omega)\left(\hat{\boldsymbol{A}}^a(\omega)^\dagger \boldsymbol{H} \hat{\boldsymbol{A}}^a(\omega) - \frac{1}{2}\{\hat{\boldsymbol{A}}^a(\omega)^\dagger \hat{\boldsymbol{A}}^a(\omega), \boldsymbol{H}\}\right)\mathrm{d}\omega$$

can be well approximated as a simpler form using

$$\|\mathcal{D}^\dagger[\boldsymbol{H}] - \sum_{a\in S} \int_{-\infty}^{\infty} \gamma(\omega)\omega \hat{\boldsymbol{A}}^a(\omega)^\dagger \hat{\boldsymbol{A}}^a(\omega)\mathrm{d}\omega\| \leq \frac{2}{\sqrt{2\pi}\tau} \cdot \|c\|_1 \cdot \|\sum_{a\in S} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|.$$

Intuitively, the following expression is the simplest proxy one can write down to capture the rate of energy change

$$\sum_{a\in S}\int_{-\infty}^{\infty}\omega\times\gamma(\omega)\hat{\boldsymbol{A}}^{a}(\omega)^{\dagger}\hat{\boldsymbol{A}}^{a}(\omega)\mathrm{d}\omega\sim(\text{energy difference})\times(\text{rate}).$$

Indeed, the Bohr frequency ω is essentially the energy difference after jump operator $\boldsymbol{A}^{a}(\omega)$; but because of the energy uncertainty in the operator Fourier transform (i.e., because $\hat{f}(\omega)$ is not a delta function), this interpretation must be corrected by an error scaling as the energy resolution $\sim 1/\tau$. The starting point of the calculation is a certain integration-by-part trick that relates the Hamiltonian operator $\boldsymbol{H}$ to the scalar ω.

Proposition 4.16.1 (Integration by parts)**.** *In the setting of Lemma 4.16.1,*

$$[\boldsymbol{H},\hat{\boldsymbol{A}}(\omega)]=\omega\hat{\boldsymbol{A}}(\omega)+\frac{1}{\mathrm{i}\sqrt{2\pi\tau}}\Big(\boldsymbol{A}(\tau/2)\mathrm{e}^{-\mathrm{i}\omega\tau/2}-\boldsymbol{A}(-\tau/2)\mathrm{e}^{\mathrm{i}\omega\tau/2}\Big).$$

Proof. Integration of the derivative can be expanded by the product rule

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi\tau}}\Big(\boldsymbol{A}(\tau/2)\mathrm{e}^{-\mathrm{i}\omega\tau/2}-\boldsymbol{A}(-\tau/2)\mathrm{e}^{\mathrm{i}\omega\tau/2}\Big)\\
&=\frac{1}{\sqrt{2\pi\tau}}\int_{-\tau/2}^{\tau/2}\frac{\mathrm{d}}{\mathrm{d}t}\Big(\boldsymbol{A}(t)\mathrm{e}^{-\mathrm{i}\omega t}\Big)\mathrm{d}t\\
&=\frac{1}{\sqrt{2\pi\tau}}\int_{-\tau/2}^{\tau/2}\Big(\mathrm{i}[\boldsymbol{H},\boldsymbol{A}(t)]\mathrm{e}^{-\mathrm{i}\omega t}-\mathrm{i}\omega\boldsymbol{A}(t)\mathrm{e}^{-\mathrm{i}\omega t}\Big)\mathrm{d}t\\
&=\mathrm{i}[\boldsymbol{H},\hat{\boldsymbol{A}}(\omega)]-\mathrm{i}\omega\hat{\boldsymbol{A}}(\omega).
\end{aligned}$$

Rearrange to conclude the proof. ∎

Observe that taking the infinite time limit $\tau\to\infty$ (i.e., perfect energy resolution) in the above proposition recovers the relation for the true Bohr frequencies ν

$$[\boldsymbol{H},\boldsymbol{A}_{\nu}]=\nu\boldsymbol{A}_{\nu}\quad\text{for each}\quad\nu\in B(\boldsymbol{H}).$$

At finite τ, the above leads to simple bounds on the correction term. We now present the proof of Lemma 4.16.1.

Proof of Lemma 4.16.1. We calculate

$$\begin{aligned}\mathcal{D}^\dagger[\boldsymbol{H}] &= \sum_{a\in S}\int_{-\infty}^{\infty}\gamma(\omega)\Big(\hat{\boldsymbol{A}}^a(\omega)^\dagger \boldsymbol{H}\hat{\boldsymbol{A}}^a(\omega) - \frac{1}{2}\{\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega), \boldsymbol{H}\}\Big)\mathrm{d}\omega \\ &= \sum_{a\in S}\int_{-\infty}^{\infty}\gamma(\omega)\frac{1}{2}\Big(\hat{\boldsymbol{A}}^a(\omega)^\dagger[\boldsymbol{H}, \hat{\boldsymbol{A}}^a(\omega)] - [\boldsymbol{H}, \hat{\boldsymbol{A}}^a(\omega)^\dagger]\hat{\boldsymbol{A}}^a(\omega)\Big)\mathrm{d}\omega \\ &= \sum_{a\in S}\int_{-\infty}^{\infty}\gamma(\omega)\omega\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega)\mathrm{d}\omega + \boldsymbol{E},\end{aligned}$$

where the error term $\boldsymbol{E}$ is given by Proposition 4.16.1 as

$$\begin{aligned}\boldsymbol{E} := \frac{-\mathrm{i}}{2\sqrt{2\pi\tau}}\sum_{a\in S}\int_{-\infty}^{\infty}\gamma(\omega)\hat{\boldsymbol{A}}^a(\omega)^\dagger\Big(\boldsymbol{A}^a(\tau/2)\mathrm{e}^{-\mathrm{i}\omega\tau/2} - \boldsymbol{A}^a(-\tau/2)\mathrm{e}^{\mathrm{i}\omega\tau/2}\Big)\mathrm{d}\omega \\ + \frac{\mathrm{i}}{2\sqrt{2\pi\tau}}\sum_{a\in S}\int_{-\infty}^{\infty}\gamma(\omega)\Big(\boldsymbol{A}^a(\tau/2)^\dagger\mathrm{e}^{\mathrm{i}\omega\tau/2} - \boldsymbol{A}^a(-\tau/2)^\dagger\mathrm{e}^{-\mathrm{i}\omega\tau/2}\Big)\hat{\boldsymbol{A}}^a(\omega)\mathrm{d}\omega.\end{aligned}$$

To bound this error term, let us calculate one of the four individual terms as an example

$$\begin{aligned}&\sum_{a\in S}\int_{-\infty}^{\infty}\gamma(\omega)\boldsymbol{A}^a(\tau/2)^\dagger\hat{\boldsymbol{A}}^a(\omega)\mathrm{e}^{\mathrm{i}\omega\tau/2}\mathrm{d}\omega \\ &= \sum_{a\in S}\frac{1}{2\pi\sqrt{\tau}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c(t_1)\mathrm{e}^{-\mathrm{i}\omega t_1}\mathrm{d}t_1\int_{-\tau/2}^{\tau/2}\mathrm{e}^{\mathrm{i}\omega\tau/2}\boldsymbol{A}^a(\tau/2)^\dagger\boldsymbol{A}^a(t_2)\mathrm{e}^{-\mathrm{i}\omega t_2}\mathrm{d}t_2\mathrm{d}\omega \\ &= \frac{1}{\sqrt{\tau}}\sum_{a\in S}\int_{-\tau/2}^{\tau/2}c(\tau/2-t_2)\boldsymbol{A}^a(\tau/2)^\dagger\boldsymbol{A}^a(t_2)\mathrm{d}t_2.\end{aligned}$$

We can bound the operator norm of this term by $\|c\|_1/\sqrt{\tau}$ after applying the triangle inequality and using the fact that

$$\begin{aligned}&\|\sum_{a\in S}\boldsymbol{A}^a(\tau/2)^\dagger\boldsymbol{A}^a(t_2)\| = \|\sum_{a\in S}\boldsymbol{A}^{a\dagger}(\tau/2)\boldsymbol{A}^a(t_2)\| \\ &\leq \|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\otimes\langle a|\|\cdot\|\sum_{a\in S}\boldsymbol{A}^a\otimes|a\rangle\| = \|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|.\end{aligned}$$

Repeat a similar argument for the other three terms to conclude the proof. ∎

Monotonicty of rates

Thermal Lindbladians generally depend sensitively on the Hamiltonian as it uses Hamiltonian simulation for a long time τ

$$\mathrm{e}^{\mathrm{i}\boldsymbol{H}\tau} \quad \text{for} \quad \tau \gg 1.$$

Therefore, even adding a small perturbation to the Hamiltonian $\boldsymbol{H}' = \boldsymbol{H} + \boldsymbol{V}$ may have a nonperturbative effect on $\mathcal{L}$ since

$$\tau\|\boldsymbol{V}\| \gg 1 \quad \text{implies} \quad \|\mathcal{L} - \mathcal{L}'\|_{1-1} \gg 0.$$

In other words, at a large τ, it is not obvious at all why the Lindbladians $\mathcal{L}, \mathcal{L}'$ are related. Indeed, the original Davies' generator ($\tau \to \infty$) is unstable against arbitrarily small perturbations to the Hamiltonian; whenever energy degeneracy is broken, the Lindbladian can change substantially.

Nevertheless, what we can show as a compromise is that the rate only *increases* if the perturbation only introduces *level splitting*; this amounts to the assumption that the original Hamiltonian has highly degenerate subspaces with a certain Bohr frequencies gap Δ_ν as another large energy scale

$$\frac{1}{\tau} \ll \|\boldsymbol{V}\| \ll \Delta_\nu.$$

Intuitively, level splitting causes *decoherence* (and only decoherence) in the Bohr frequencies; for large τ, the Lindblidan can indeed tell the transitions ω, ω' apart if the Bohr frequencies are sufficiently different. Fortunately, even though decoherence can change the Lindbladian by a lot, we establish certain *monotonicity* of transition rates. That is, $\mathcal{L}'$ must have as good transition rates as $\mathcal{L}$. A good example of $\boldsymbol{O}$ would be an energy subspace projector. However, the argument works for general $\boldsymbol{O}$, which makes it more flexible to use.

Lemma 4.16.2 (Decoherence increases the rates)**.** *For any set of operators $\{\boldsymbol{A}^a\}_{a\in S}$, suppose there exists an operator $\boldsymbol{O}$ such that*

$$\sum_{a\in S} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a \succeq \boldsymbol{O} \quad \textit{where} \quad [\boldsymbol{O}, \boldsymbol{H}] = 0.$$

Then, the operator Fourier Transforms $\hat{\boldsymbol{A}}^a(\omega)$ (subscript f omitted) for some normalized weight $\int_{-\infty}^{\infty} |f(t)|^2 \mathrm{d}t = 1$ and Hamiltonian $\boldsymbol{H}$ satisfies

$$\sum_{a\in S} \int_{-\infty}^{\infty} \hat{\boldsymbol{A}}^a(\omega)^\dagger \hat{\boldsymbol{A}}^a(\omega) \mathrm{d}\omega \succeq \boldsymbol{O}.$$

Proof. By Proposition 4.15.1,

$$\begin{aligned}\sum_{a\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}^a(\omega)^\dagger\hat{\boldsymbol{A}}^a(\omega)\mathrm{d}\omega-\boldsymbol{O}&=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\left(\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right)\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}|f(t)|^2\mathrm{d}t-\boldsymbol{O}\\&=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\left(\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a-\boldsymbol{O}\right)\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}|f(t)|^2\mathrm{d}t\\&=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{X}^\dagger\cdot\boldsymbol{X}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}|f(t)|^2\mathrm{d}t\\&\geq 0.\end{aligned}$$

The second equality uses that $\int_{-\infty}^{\infty}|f(t)|^2\mathrm{d}t=1$ and that $\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{O}\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}=\boldsymbol{O}$. The last line establishes PSD order using the assumption that there exists operator $\boldsymbol{X}$ such that $\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a-\boldsymbol{O}=\boldsymbol{X}^\dagger\boldsymbol{X}$. Together, we establish the desired statement. ∎

Perturbation theory of eigenstates and eigenvalues

We state a few useful facts about perturbed eigenspace and eigenvalues that would be useful in the proofs.

Proposition 4.16.2 (Davis-Kahan $\sin\Theta$ theorem (see also Theorem VII.3.1 of [22]))**.** *Let $\boldsymbol{H}$ and $\tilde{\boldsymbol{H}}$ be two equal-sized Hermitian matrices. Let $\boldsymbol{P}$ be the projector onto eigenstates of $\boldsymbol{H}$ with eigenvalue in an interval $[a,b]$. Let $\tilde{\boldsymbol{P}}^\perp$ be the projector onto eigenstates of $\tilde{\boldsymbol{H}}$ with eigenvalues outside the interval $[a-\delta,b+\delta]$. Then*

$$\|\boldsymbol{P}\tilde{\boldsymbol{P}}^\perp\|\leq\|\boldsymbol{H}-\tilde{\boldsymbol{H}}\|/\delta.$$

Here $\|\cdot\|$ is the spectral norm (or any unitarily invariant norm).

Furthermore, the following fact bounds errors on perturbed eigenvalues:

Proposition 4.16.3 (Weyl's inequality)**.** *For any two equal-sized Hermitian matrices $\boldsymbol{H}$ and $\tilde{\boldsymbol{H}}$, we have $|\lambda_j(\boldsymbol{H})-\lambda_j(\tilde{\boldsymbol{H}})|\leq\|\boldsymbol{H}-\tilde{\boldsymbol{H}}\|$ for all j, where $\lambda_j(\boldsymbol{X})$ is the j-th largest eigenvalue of matrix $\boldsymbol{X}$.*

Together, these facts imply that

Lemma 4.16.3. *Let $\boldsymbol{H}$ and $\tilde{\boldsymbol{H}}=\boldsymbol{H}+\boldsymbol{V}$ be Hamiltonians. Let $\boldsymbol{P}$ be the projector onto eigenstates of $\boldsymbol{H}$ with eigenvalues in some interval $[a,b]$, which are separated from the other eigenvalues by a gap of at least Δ. If $\|\boldsymbol{V}\|\leq\Delta/4$, then*

1. *There exists a spectral projector $\tilde{\boldsymbol{P}}$ onto eigenstates of $\tilde{\boldsymbol{H}}$ with eigenvalues in $[a - \Delta/4, b + \Delta/4]$, which are separated from the other eigenvalues by a gap of at least $\Delta/2$.*

2. $\|\boldsymbol{P} - \tilde{\boldsymbol{P}}\| \leq 8\|\boldsymbol{V}\|/\Delta$.

Proof. The existence of $\tilde{\boldsymbol{P}}$ (item 1) holds because of Proposition 4.16.3 above. Then observe that

$$\|\boldsymbol{P} - \tilde{\boldsymbol{P}}\| = \|\boldsymbol{P} - \boldsymbol{P}\tilde{\boldsymbol{P}} + \boldsymbol{P}\tilde{\boldsymbol{P}} - \tilde{\boldsymbol{P}}\| \leq \|\boldsymbol{P}\tilde{\boldsymbol{P}}^{\perp}\| + \|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{P}}\| \leq 8\|\boldsymbol{V}\|/\Delta,$$

where the last inequality is obtained by applying Proposition 4.16.2 with $\delta = \Delta/4$ to bound $\|\boldsymbol{P}\tilde{\boldsymbol{P}}^{\perp}\|$ and $\|\boldsymbol{P}^{\perp}\tilde{\boldsymbol{P}}\|$. ∎

Lemma 4.16.4 (Off-block-diagonal perturbation)**.** *Consider a block diagonal Hermitian matrix $\boldsymbol{D} = \boldsymbol{D}_1 + \boldsymbol{D}_2$, where the two blocks correspond to orthogonal subspace projectors $\boldsymbol{P}_1$ and $\boldsymbol{I} - \boldsymbol{P}_1 = \boldsymbol{P}_2$ and are separated by eigenvalue gap at least Δ. Add an off-block-diagonal Hermitian perturbation $\boldsymbol{V} = \boldsymbol{V}_{12} + \boldsymbol{V}_{21}$ such that $\|\boldsymbol{V}\| \leq \Delta/4$. Then, there is an anti-Hermitian operator $\boldsymbol{B}$ and an absolute constant C_0 such that*

$$\boldsymbol{D} + \boldsymbol{V} = \mathrm{e}^{-\boldsymbol{B}}\boldsymbol{D}\mathrm{e}^{\boldsymbol{B}} + (\boldsymbol{D} + \boldsymbol{V} - \mathrm{e}^{-\boldsymbol{B}}\boldsymbol{D}\mathrm{e}^{\boldsymbol{B}})$$

$$\textit{where} \qquad \|\boldsymbol{B}\| \leq C_0\frac{\|\boldsymbol{V}\|}{\Delta}, \qquad \textit{and} \qquad \|\mathrm{e}^{\boldsymbol{B}}(\boldsymbol{D} + \boldsymbol{V})\mathrm{e}^{-\boldsymbol{B}} - \boldsymbol{D}\| \leq C_0\frac{\|\boldsymbol{V}\|^2}{\Delta}.$$

This implies the sorted eigenvalues of $\boldsymbol{D}$ are perturbed by $C_0\|\boldsymbol{V}\|^2/\Delta$.

We remark that the scaling with respect to $\|\boldsymbol{V}\|$ is consistent with perturbation theory: the *angle* change is first-order $\sim \frac{\|\boldsymbol{V}\|}{\Delta}$, and the *eigenvalue* change is second-order $\sim \frac{\|\boldsymbol{V}\|^2}{\Delta}$. Note that for diagonal perturbation, the eigenvalue change is only bounded by $\sim \|\boldsymbol{V}\|$.

Proof. Observe that

$$\mathrm{e}^{\boldsymbol{B}}(\boldsymbol{D} + \boldsymbol{V})\mathrm{e}^{-\boldsymbol{B}} = \boldsymbol{D} + (\boldsymbol{V} + [\boldsymbol{B}, \boldsymbol{D}]) + [\boldsymbol{B}, \boldsymbol{V}] + \sum_{k=2}^{\infty}\frac{1}{k!}\operatorname{ad}_{\boldsymbol{B}}^{k}(\boldsymbol{D} + \boldsymbol{V}).$$

Let us choose $\boldsymbol{B}$ to cancel the first order term in $\boldsymbol{V}$, i.e.,

$$\boldsymbol{V} = -[\boldsymbol{B}, \boldsymbol{D}]. \tag{4.113}$$

We can solve for $\boldsymbol{B}$ by working in the eigenbasis of $\boldsymbol{D} = \sum_i D_i|\psi_i\rangle\langle\psi_i|$. Then denoting $O_{ij} = \langle\psi_i|\boldsymbol{O}|\psi_j\rangle$, we can rewrite Eq. (4.16) as

$$V_{ij} = B_{ij}(D_i - D_j) \qquad \text{or} \qquad B_{ij} = \frac{V_{ij}}{D_i - D_j}.$$

Note $B_{ij} = V_{ij} = 0$ whenever $|D_i - D_j| \geq \Delta$ by assumption. Hence, we can solve for $\boldsymbol{B}$ using the Heisenberg picture Fourier transform:

$$B_{ij} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \mathrm{e}^{\mathrm{i}(D_i - D_j)t} V_{ij} \mathrm{d}t = V_{ij} \hat{f}(D_j - D_i) \quad \text{for each} \quad i, j$$
$$\text{or} \quad \boldsymbol{B} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \underbrace{\mathrm{e}^{\mathrm{i}\boldsymbol{D}t} \boldsymbol{V} \mathrm{e}^{-\mathrm{i}\boldsymbol{D}t}}_{=:\boldsymbol{V}(t)} \mathrm{d}t,$$

where we choose the function $f(t)$ whose Fourier transform matches the reciprocal at sufficiently large values,

$$\hat{f}(\omega) = \frac{1}{-\omega} \quad \text{when} \quad |\omega| \geq \Delta,$$

but remain "nice" near $\omega = 0$. One example is to use a smooth bump function

$$-\frac{1}{\omega} \cdot b(\frac{\omega}{\Delta}) \quad \text{where} \quad b(x) = \begin{cases} 1 & \text{if} \quad |x| \geq 1 \\ O(x^2) & \text{if} \quad |x| \approx 0. \end{cases}$$

For concreteness, we take

$$b(x) = \begin{cases} 1 - \exp(\frac{1}{1 - \frac{1}{x^2}}) & \text{if} \quad |x| < 1 \\ 1 & \text{else} \end{cases}.$$

Then, taking triangle inequality and using the unitary invariance of the operator norm,

$$\|\boldsymbol{B}\| \leq \frac{1}{\sqrt{2\pi}} \|f\|_1 \cdot \|\boldsymbol{V}\| \leq C_0 \frac{\|\boldsymbol{V}\|}{\Delta}.$$

The last inequality bounds the Fourier transform by change-of-variable $x = \omega/\Delta$ and leaves a constant C_0 that depends on the inverse Fourier transform of the "dimensionless function" $b(x)/x$. This bound on $\boldsymbol{B}$ then allows us to control the higher-order errors

$$\mathrm{e}^{\boldsymbol{B}}(\boldsymbol{D}+\boldsymbol{V})\mathrm{e}^{-\boldsymbol{B}} = \boldsymbol{D}+\boldsymbol{V}+\int_0^1 \mathrm{e}^{\boldsymbol{B}s}[\boldsymbol{B},\boldsymbol{V}]\mathrm{e}^{-\boldsymbol{B}s}\mathrm{d}s+[\boldsymbol{B},\boldsymbol{D}]+\int_0^1 \mathrm{e}^{\boldsymbol{B}s}[\boldsymbol{B},[\boldsymbol{B},\boldsymbol{D}]]\mathrm{e}^{-\boldsymbol{B}s}(1-s)\mathrm{d}s.$$

Substitute $[\boldsymbol{B}, \boldsymbol{D}] = -\boldsymbol{V}$ and rearranging, we get

$$e^{\boldsymbol{B}}(\boldsymbol{D}+\boldsymbol{V})e^{-\boldsymbol{B}} - \boldsymbol{D} = \int_0^1 e^{\boldsymbol{B}s}[\boldsymbol{B}, \boldsymbol{V}]e^{-\boldsymbol{B}s} s ds.$$

Apply the triangle inequality, we get

$$\|e^{\boldsymbol{B}}(\boldsymbol{D}+\boldsymbol{V})e^{-\boldsymbol{B}} - \boldsymbol{D}\| \leq \|[\boldsymbol{B}, \boldsymbol{V}]\| \int_0^1 s ds \leq C_0 \frac{\|\boldsymbol{V}\|^2}{\Delta}.$$

Finally, to obtain sorted eigenvalues of $\boldsymbol{D}$, use the fact that $e^{-\boldsymbol{B}}$ is unitary and apply Weyl's inequality for $e^{-\boldsymbol{B}}\boldsymbol{D}e^{\boldsymbol{B}}$. ■

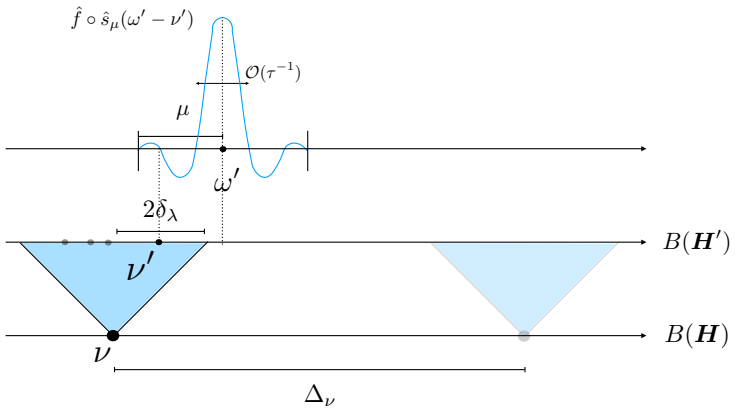


Figure 4.5: The energy scales $\frac{1}{\tau} \ll \|\boldsymbol{V}\| \ll \Delta_\nu$ in one plot. The Hamiltonian perturbation $\boldsymbol{V}$ causes eigenvalues of $\boldsymbol{H}$ to change by at most $\delta_\lambda \leq \|\boldsymbol{V}\|$, which splits the Bohr frequencies such that $|\nu - \nu'| \leq 2\delta_\lambda$. Here μ is the cut-off frequency for the secular approximation $\hat{\boldsymbol{S}}^a_{f,\mu}(\omega') = \sum_{\nu'} \boldsymbol{A}^a_{\nu'} \hat{f}(\omega' - \nu')\hat{s}_\mu(\omega' - \nu')$. As long as we choose $\mu < \frac{\Delta_\nu - 4\delta_\lambda}{2}$ small enough, the secular approximation $\hat{\boldsymbol{S}}^a(\omega')$ can only contain Bohr frequencies $\boldsymbol{A}_{\nu'}$ from at most one block ν, i.e., different blocks decohere.

Proof of Theorem 4.16.1

We combine the above ingredients for the proof of Theorem 4.16.1. In what follows, let E_j, $\boldsymbol{P}_{E_j}$, ν be the eigenvalues, eigenspace projectors, and Bohr frequencies of the unperturbed Hamiltonian $\boldsymbol{H}$; furthermore let E'_j, $\boldsymbol{P}_{E'_j}$ ν', be their counterpart for the perturbed Hamiltonian $\boldsymbol{H}'$.

It will be helpful to display the structure of the Bohr frequencies and the energy eigenspaces under perturbation by

$$\begin{aligned}\boldsymbol{A}^a &= \sum_{\nu \in B(\boldsymbol{H})} \sum_{E_1 - E_2 = \nu} \boldsymbol{P}_{E_1} \boldsymbol{A}^a \boldsymbol{P}_{E_2} = \sum_{\nu \in B(\boldsymbol{H})} \boldsymbol{A}^a_\nu \\ &= \sum_{\nu' \in B(\boldsymbol{H}')} \sum_{E'_1 - E'_2 = \nu'} \boldsymbol{P}_{E'_1} \boldsymbol{A}^a \boldsymbol{P}_{E'_2} = \sum_{\nu' \in B(\boldsymbol{H}')} \boldsymbol{A}^a_{\nu'} = \sum_{\nu \in B(\boldsymbol{H})} \boldsymbol{A}^a_{\approx \nu},\end{aligned}$$

where we defined

$$\boldsymbol{A}^a_{\approx\nu} := \sum_{\nu' \in B(\boldsymbol{H}'),\ \nu' \approx \nu} \boldsymbol{A}^a_{\nu'} \qquad \text{and} \quad \nu' \approx \nu \iff |\nu' - \nu| \le 2\delta_\lambda.$$

In other words, the perturbed set of Bohr frequencies can be identified with the original degenerate blocks according to eigenvalue perturbation[11] bounded by δ_λ, under the assumption that the perturbation is weaker than the Bohr frequency differences $\Delta_\nu > 4\delta_\lambda$. This structure is crucial for proving the monotonicity of gradients; see Figure 4.5. For later use, we also define $\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')$ to be the operator Fourier transform of $\boldsymbol{A}^a_{\approx\nu}$ with respect to the perturbed Hamiltonian $\boldsymbol{H}'$, and consider its secular approximation $\hat{\boldsymbol{S}}^a_{\approx\nu}(\omega')$ at truncation scale μ, i.e.,

$$\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega') = \sum_{\nu' \approx \nu} \boldsymbol{A}^a_{\nu'} \hat{f}_\tau(\omega' - \nu'), \qquad \hat{\boldsymbol{S}}^a_{\approx\nu}(\omega') = \sum_{\nu' \approx \nu} \boldsymbol{A}^a_{\nu'} \hat{f}_\tau(\omega' - \nu') \mathbb{1}(|\omega' - \nu'| < \mu). \tag{4.114}$$

In what follows, we will denote $\theta(\omega) = \gamma(\omega)\omega$. It is worth recalling the following bounds:

$$\|\boldsymbol{A}^a\|,\ \|f_\tau\|_2,\ \frac{\|c_\beta\|_1}{\sqrt{2\pi}} \le 1, \qquad \|\theta\|_\infty = O(\Lambda_0), \qquad \Lambda_0 = \Theta(1).$$

Our strategy for proving Theorem 4.12.1 is to rewrite the energy gradients $\mathcal{L}^\dagger[\boldsymbol{H}]$ and $\mathcal{L}'^\dagger[\boldsymbol{H}']$ in a form amenable to Lemma 4.16.2 between a set of operators and their Fourier Transforms.

Step 1. For the perturbed Hamiltonian $\boldsymbol{H}'$, we apply a sequence of approximations to establish

$$\left\| \sum_{a \in S} \sum_{\nu \in B(\boldsymbol{H})} \theta_-(\nu) \int_{-\infty}^{\infty} \hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')^\dagger \hat{\boldsymbol{A}}^a_{\approx\nu}(\omega') \mathrm{d}\omega' - \mathcal{L}'^\dagger[\boldsymbol{H}'] \right\| \le \epsilon_A \tag{4.115}$$

for some $\epsilon_A > 0$ and a function θ_- to be soon specified. Recall we write $\boldsymbol{A} \overset{E}{\approx} \boldsymbol{B}$ if

[11] Note that we always have $\delta_\lambda \le \|\boldsymbol{V}\|$ by Proposition 4.16.3, but we keep δ_λ as an separate parameter which helps yield better bounds when $\delta_\lambda \ll \|\boldsymbol{V}\|$, such as the case when the $\boldsymbol{V}$ is an off-diagonal perturbation (see Lemma 4.16.4).

$\|\boldsymbol{A} - \boldsymbol{B}\| \leq E$.

$$
\begin{aligned}
\mathcal{L}'^{\dagger}[\boldsymbol{H}'] &\overset{E_1}{\approx} \mathcal{D}'^{\dagger}[\boldsymbol{H}'] \\
&\overset{E_2}{\approx} \sum_{a\in S}\int_{-\infty}^{\infty}\theta(\omega')\hat{\boldsymbol{A}}^a(\omega')^\dagger\hat{\boldsymbol{A}}^a(\omega')\mathrm{d}\omega' \\
&\overset{E_3}{\approx} \sum_{a\in S}\int_{-\infty}^{\infty}\theta(\omega')\hat{\boldsymbol{S}}^a(\omega')^\dagger\hat{\boldsymbol{S}}^a(\omega')\mathrm{d}\omega' \\
&= \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\int_{-\infty}^{\infty}\theta(\omega')\hat{\boldsymbol{S}}^a_{\approx\nu}(\omega')^\dagger\hat{\boldsymbol{S}}^a_{\approx\nu}(\omega')\mathrm{d}\omega' \\
&\overset{E_4}{\approx} \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\int_{-\infty}^{\infty}\theta(\omega')\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')^\dagger\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')\mathrm{d}\omega' \\
&\overset{E_5}{\approx} \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\theta(\nu)\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')^\dagger\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')\mathrm{d}\omega' \\
&\overset{E_6}{\approx} \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\theta_-(\nu)\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')^\dagger\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')\mathrm{d}\omega'.
\end{aligned}
$$

The approximations E_1, E_2 are bounded by

$$
\begin{aligned}
E_1 &\leq O\left(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}}\|c_\beta\|_1\|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\right) = |S|O(\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}}) \\
E_2 &\leq \frac{2\|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|}{\sqrt{2\pi}\tau}\|c_\beta\|_1 = |S|O(\frac{1}{\tau}).
\end{aligned}
$$

In the approximations E_3 and E_4, we choose the secular approximation parameter μ such that

$$
\mu < (\Delta_\nu - 4\delta_\lambda)/2 \tag{4.124}
$$

with associated errors given by Corollary 4.15.3 as

$$
\begin{aligned}
E_3 &\leq 2\|\theta\|_\infty\|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\|\hat{f}_\tau\cdot(1-\hat{s}_\mu)\|_2\|f_\tau\|_2 = |S|O(\Lambda_0\sqrt{\frac{1}{\mu\tau}}), \\
E_4 &\leq 2\|\theta\|_\infty\|\sum_{a,\nu}\boldsymbol{A}^{a\dagger}_{\approx\nu}\boldsymbol{A}^a_{\approx\nu}\|\|\hat{f}_\tau\cdot(1-\hat{s}_\mu)\|_2\|f_\tau\|_2 = |S|O(\Lambda_0\sqrt{\frac{1}{\mu\tau}}),
\end{aligned}
$$

where we applied Eq. (4.93) to bound $\|\hat{f}_\tau\cdot(1-\hat{s}_\mu)\|_2 \leq \sqrt{4/\pi\mu\tau}$, and used Proposition 4.16.5 to bound the spectral norm of the sum of jump operators in the second line.

To justify the equality on the fourth line (different blocks νs decohere), observe the choice of the parameter μ in Eq. (4.16) implies

$$\hat{\boldsymbol{S}}^a(\omega') = \sum_{\nu \in B(\boldsymbol{H})} \hat{\boldsymbol{S}}^a_{\approx\nu}(\omega') \mathbb{1}(|\omega' - \nu| < \mu + 2\delta_\lambda) = \sum_{\nu \in B(\boldsymbol{H})} \hat{\boldsymbol{S}}^a_{\approx\nu}(\omega') \mathbb{1}(|\omega' - \nu| < \Delta_\nu/2), \tag{4.125}$$

where $\hat{\boldsymbol{S}}^a_{\approx\nu}(\omega')$ is given in Eq. (4.114). This ensures that for any given ω', $\hat{\boldsymbol{S}}^a(\omega')$ can activate at most *one* block of transitions with Bohr frequencies closest to ν (see Figure 4.5). Consequently,

$$\hat{\boldsymbol{S}}^{a\dagger}(\omega')\hat{\boldsymbol{S}}^a(\omega') = \sum_{\nu \in B(\boldsymbol{H})} \hat{\boldsymbol{S}}^{a\dagger}_{\approx\nu}(\omega')\hat{\boldsymbol{S}}^a_{\approx\nu}(\omega').$$

Next, for the approximation E_5, we define the following "rounded" function $\bar{\theta}(\omega')$ where an input ω' close to $\nu \in B(\boldsymbol{H})$ is assigned the same value $\theta(\nu)$, with uniqueness of ν guaranteed by Eq. (4.125),

$$\bar{\theta}(\omega') := \begin{cases} \theta(\nu) & \text{if} \quad |\omega' - \nu| \leq \mu + 2\delta_\lambda \quad \text{for} \quad \nu \in B(\boldsymbol{H}) \\ \theta(\omega') & \text{else} \end{cases}.$$

This lets us formally pull $\theta(\omega')$ out of the integral. Of course, this rounding introduces an error scaling with the energy spread multiplied with the derivative

$$\|\bar{\theta} - \theta\|_\infty \leq (2\mu + 4\delta_\lambda) \cdot \|\mathrm{d}\theta/\mathrm{d}\omega\|_\infty.$$

Roughly, this error quantifies how the energy gradient (i.e., $\theta(\omega) = \gamma_\beta(\omega)\omega$) changes due to perturbation in Bohr frequency. Thus,

$$E_5 \leq \|\theta - \bar{\theta}\|_\infty \left\| \sum_{a,\nu} \boldsymbol{A}^{a\dagger}_{\approx\nu} \boldsymbol{A}^a_{\approx\nu} \right\| \|\hat{f}_\tau \cdot \hat{s}_\mu\|_2^2 = |S| O(2\mu + 4\delta_\lambda),$$

where we applied Propositions 4.16.4 and 4.16.5 (deferred to section 4.16), and used the fact that $\|\hat{f}_\tau \hat{s}_\mu\|_2 \leq \|\hat{f}_\tau\|_2 \leq 1$.

Finally, in the last approximation E_6, we define the truncated weight

$$\theta_-(\nu) = \theta(\nu)\mathbb{1}(\nu \leq \Delta_\nu/2).$$

This truncation has the property that $\theta_-(\nu) \leq 0$ for each $\nu \in B(\boldsymbol{H})$, which ensures the last line is negative semidefinite. Thus,

$$E_6 \leq \|\theta - \theta_-\|_\infty \| \sum_{a,\nu} \boldsymbol{A}^{a\dagger}_{\approx\nu} \boldsymbol{A}^a_{\approx\nu} \| \|f_\tau\|_2^2 \leq |S| \|\theta - \theta_-\|_\infty,$$

$$\text{where} \qquad \|\theta - \theta_-\|_\infty \leq \max_{\omega \geq \Delta_\nu/2} \omega\gamma_\beta(\omega) \leq \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta},$$

using the tail bound in Eq. (4.20). Altogether,

$$\begin{aligned}\epsilon_A &= E_1 + E_2 + E_3 + E_4 + E_5 + E_6 \\ &\le O\left(|S|\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \mu + \frac{\Lambda_0}{\sqrt{\mu\tau}} + \delta_\lambda + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta}\right)\right).\end{aligned}$$

We then choose $\mu = \min(\Lambda_0^{2/3}/\tau^{1/3}, (\Delta_\nu - 4\delta_\lambda)/4)$ so as to optimize the error $O(\mu + \Lambda_0/\sqrt{\mu\tau})$ while subject to the constraint that $\mu < (\Delta_\nu - 4\delta_\lambda)/2$. This choice implies $\Lambda_0/\sqrt{\mu\tau} \le \Lambda_0^{2/3}/\tau^{1/3} + 2\Lambda_0/\sqrt{(\Delta_\nu - 4\delta_\lambda)\tau} \le O(\Lambda_0^{2/3}/\tau^{1/3} + \Lambda_0/\sqrt{\Delta_\nu\tau})$, where we used $\delta_\lambda \le \|\boldsymbol{V}\| \le \Delta_\nu/8$ which is a combination of Proposition 4.16.3 and the assumption in the theorem statement. This yields the following error bound

$$\epsilon_A \le O\left(|S|\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu\tau}} + \delta_\lambda + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta}\right)\right). \tag{4.126}$$

Step 2. For the original Hamiltonian $\boldsymbol{H}$, we may repeat the above argument with trivial perturbation ($\boldsymbol{V} = 0$) to get

$$\|\sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})} \theta_-(\nu)(\boldsymbol{A}^a_{\approx\nu})^\dagger \boldsymbol{A}_{\approx\nu} - \mathcal{L}^\dagger[\boldsymbol{H}]\| \le \epsilon_B \tag{4.127}$$

for some $\epsilon_B > 0$. More detailedly, we have

$$\begin{aligned}\mathcal{L}^\dagger[\boldsymbol{H}] &\overset{E_7}{\approx} \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})} \int_{-\infty}^{\infty} \theta_-(\nu)\hat{\boldsymbol{A}}^a_\nu(\omega)^\dagger \hat{\boldsymbol{A}}^a_\nu(\omega)\mathrm{d}\omega \\ &= \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})} \theta_-(\nu)(\boldsymbol{A}^a_\nu)^\dagger \boldsymbol{A}^a_\nu \\ &\overset{E_8}{\approx} \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})} \theta_-(\nu)(\boldsymbol{A}^a_{\approx\nu})^\dagger \boldsymbol{A}^a_{\approx\nu}.\end{aligned}$$

The second line is operator Parseval's identity, where the time evolution simplifies due to commutativity $[\boldsymbol{A}^{a\dagger}_\nu \boldsymbol{A}^a_\nu, \boldsymbol{H}] = 0$. The error E_7 can be bounded by the same bounds for ϵ_A in Eq. (4.126) by setting $\boldsymbol{V} = 0$ (i.e., $\boldsymbol{A}^a_{\approx\nu} \to \boldsymbol{A}^a_\nu$).

The last line is a brute-force rewriting of $\boldsymbol{A}^a_\nu$ into $\boldsymbol{A}^a_{\approx\nu}$, which acts on eigenstates of $\boldsymbol{H}'$ instead of $\boldsymbol{H}$, with error E_8 bounded by perturbation theory. This rewriting allows us to prove Theorem 4.16.1 by directly applying Lemma 4.16.2 between the following set of operators and their Fourier transforms:

$$\{\sqrt{|\theta_-(\nu)|}\boldsymbol{A}^a_{\approx\nu}\}_{a,\nu} \quad \text{and} \quad \{\sqrt{|\theta_-(\nu)|}\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')\}_{a,\nu,\omega'} \quad \text{for the perturbed Hamiltonian} \quad \boldsymbol{H}'.$$

We give an explicit error bound on E_8 in Proposition 4.16.6 (deferred to section 4.16), which yields

$$E_8 = O\left(\left\|\sum_{a\in S} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right\|\theta_{\max}\frac{\|\boldsymbol{V}\|}{\Delta_\nu}\right) = O\left(|S|\theta_{\max}\frac{\|\boldsymbol{V}\|}{\Delta_\nu}\right),$$

where we used the bound $\max_{\nu\in B(\boldsymbol{H})}|\theta_-(\nu)| = \theta_{\max}$ provided in the theorem statement. Collect the errors to bound $\epsilon_B = E_7 + E_8 \le \epsilon_A + E_8$.

Step 3. Now we may finish the proof of Theorem 4.16.1 by applying Lemma 4.16.2. First, Eq. (4.127) implies

$$\begin{aligned}-\sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\theta_-(\nu)(\boldsymbol{A}^a_{\approx\nu})^\dagger\boldsymbol{A}_{\approx\nu} + \epsilon_B\boldsymbol{I} &\succeq -\mathcal{L}^\dagger[\boldsymbol{H}]\\ &\succeq r\boldsymbol{O} - \epsilon\boldsymbol{I}\\ &\succeq r\boldsymbol{O}' - (r\|\boldsymbol{O}'-\boldsymbol{O}\|+\epsilon)\boldsymbol{I}.\end{aligned}$$

Similarly, Eq. (4.115) implies

$$\begin{aligned}-\mathcal{L}'^\dagger[\boldsymbol{H}'] + \epsilon_A\boldsymbol{I} &\succeq -\sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\theta_-(\nu)\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')^\dagger\hat{\boldsymbol{A}}^a_{\approx\nu}(\omega')\mathrm{d}\omega'\\ &\succeq r\boldsymbol{O}' - (\epsilon+\epsilon_B + r\|\boldsymbol{O}-\boldsymbol{O}'\|)\boldsymbol{I}.\end{aligned}$$

Note in the last step we used the assumption that $[\boldsymbol{O}', \boldsymbol{H}'] = 0$ and applied Lemma 4.16.2. Hence, we have shown that $-\mathcal{L}'^\dagger[\boldsymbol{H}'] \succeq r\boldsymbol{O}' - \epsilon'\boldsymbol{I}$, where $\epsilon' = \epsilon + \epsilon_A + \epsilon_B + r\|\boldsymbol{O}-\boldsymbol{O}'\|$ can be bounded by

$$\epsilon' \le \epsilon + O\left(|S|\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu\tau}} + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta} + \delta_\lambda + \theta_{\max}\frac{\|\boldsymbol{V}\|}{\Delta_\nu} + r\|\boldsymbol{O}-\boldsymbol{O}'\|\right)\right).$$

A simpler bound. We now consider the special case of $\boldsymbol{O} = \boldsymbol{I} - \boldsymbol{P}$, $\boldsymbol{O} = \boldsymbol{I} - \boldsymbol{P}'$ to derive a simpler bound as in the theorem statement. Note we have $\|\boldsymbol{O}-\boldsymbol{O}'\| = \|\boldsymbol{P}-\boldsymbol{P}'\| \le 8\|\boldsymbol{V}\|/\Delta_E \le 8\|\boldsymbol{V}\|/\Delta_\nu$, using Lemma 4.16.3 and the fact the spectral gap is lower bounded by the Bohr-frequency gap, $\Delta_E \ge \Delta_\nu$. Furthermore generally $\theta_{\max} = \|\theta_-\|_\infty = O(\Lambda_0)$ for our choice of $\gamma_\beta(\omega)$ in Equation 4.21. And we always have $\delta_\lambda \le \|\boldsymbol{V}\|$ by Proposition 4.16.3. Plugging these into Equation 4.16, we have

$$\epsilon' \le \epsilon + O\left(|S|\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu\tau}} + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta} + \|\boldsymbol{V}\| + (\Lambda_0 + r)\frac{\|\boldsymbol{V}\|}{\Delta_\nu}\right)\right).$$

This concludes our proof of Theorem 4.16.1.

Supplementary calculations

In this section, we provide some missing calculations that prove some propositions used in the proof in the previous section.

Proposition 4.16.4 (Bounds on the derivative)**.** *There exists an absolute constant C such that for any β, Λ_0,*

$$\|\frac{\mathrm{d}}{\mathrm{d}\omega}\theta(\omega)\|_\infty = O\left(\|\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{\mathrm{e}^{-\omega^2/2\Lambda_0^2}}{1+\mathrm{e}^{\beta\omega}}\omega\right)\|_\infty\right) \leq C.$$

Proof. By the product rule,

$$\left|\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{\omega\mathrm{e}^{-\omega^2/2\Lambda_0^2}}{1+\mathrm{e}^{\beta\omega}}\right)\right| = \left|\frac{\mathrm{e}^{-\omega^2/2\Lambda_0^2} - \mathrm{e}^{-\omega^2/2\Lambda_0^2}\omega^2/\Lambda_0^2}{1+\mathrm{e}^{\beta\omega}} - \frac{\mathrm{e}^{-\omega^2/2\Lambda_0^2}\beta\omega\mathrm{e}^{\beta\omega}}{(1+\mathrm{e}^{\beta\omega})^2}\right| \leq (const.)$$

using change of variable $x = \beta\omega$ and $y = \omega/\Lambda_0$ to obtain the absolute constant bound. ∎

Proposition 4.16.5. *In the prevailing notation,*

$$\left\|\sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\boldsymbol{A}_\nu^{a\dagger}\boldsymbol{A}_\nu^a\right\|, \quad \left\|\sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\boldsymbol{A}_{\approx\nu}^{a\dagger}\boldsymbol{A}_{\approx\nu}^a\right\| \leq \left\|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right\|.$$

Proof. We focus on $\boldsymbol{A}_{\approx\nu}^{a\dagger}$ since the case of $\boldsymbol{A}_\nu^{a\dagger}$ is a special case. Resolve the identity by nearby energy subspaces

$$\boldsymbol{I} = \sum_E \boldsymbol{P}_E = \sum_{\bar{E}} \underbrace{\sum_{E\approx\bar{E}} \boldsymbol{P}_E}_{=:\boldsymbol{P}_{\approx\bar{E}}} \quad \text{such that} \quad \boldsymbol{A}_{\approx\nu}^a = \sum_{\bar{E}_2-\bar{E}_1=\nu} \boldsymbol{P}_{\approx\bar{E}_2}\boldsymbol{A}^a\boldsymbol{P}_{\approx\bar{E}_1}.$$

Now, we calculate

$$\begin{aligned}\left\|\sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})}\boldsymbol{A}_{\approx\nu}^{a\dagger}\boldsymbol{A}_{\approx\nu}^a\right\| &= \left\|\sum_{a\in S,\bar{E}_2,\bar{E}_1}\boldsymbol{P}_{\approx\bar{E}_1}\boldsymbol{A}^{a\dagger}\boldsymbol{P}_{\approx\bar{E}_2}\boldsymbol{A}^a\boldsymbol{P}_{\approx\bar{E}_1}\right\| = \left\|\sum_{a\in S,\bar{E}_1}\boldsymbol{P}_{\approx\bar{E}_1}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\boldsymbol{P}_{\approx\bar{E}_1}\right\| \\ &= \max_{\bar{E}}\left\|\boldsymbol{P}_{\approx\bar{E}_1}\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\boldsymbol{P}_{\approx\bar{E}_1}\right\| \leq \left\|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right\|.\end{aligned}$$

The last line uses that the operator norm of block-diagonal matrices equals the maximum among the blocks. ∎

Proposition 4.16.6 (Jumps with perturbed Hamiltonian)**.** *In the prevailing notation, and for any function $h(\nu)$, we have that*

$$\|\sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})} h(\nu)(\boldsymbol{A}^a_\nu)^\dagger \boldsymbol{A}^a_\nu - \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})} h(\nu)(\boldsymbol{A}^a_{\approx\nu})^\dagger \boldsymbol{A}^a_{\approx\nu}\| \\ \leq O\left(\left\|\sum_{a\in S} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\right\| \cdot \max_{\nu\in B(\boldsymbol{H})}|h(\nu)| \cdot \frac{\|\boldsymbol{V}\|}{\Delta_\nu}\right).$$

Proof. It suffices to set $\|\sum_{a\in S} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| = 1$. Consider the operator Fourier transform with a smooth bump weight $\|g\|_2 = 1$,

$$\boldsymbol{A}^a_{g,\boldsymbol{H}}(\omega) \quad \text{for} \quad \hat{g}(\omega) \propto \begin{cases} 0 & \text{if} \quad |\omega| \geq \frac{\Delta_\nu}{2} \\ O(1) & \text{else.} \end{cases}$$

which automatically decohere different Bohr frequency blocks (i.e., no need to apply secular approximation). Also, extend the function locally:

$$h(\omega) := \Big\{ h(\nu) \quad \text{if} \quad |\omega - \nu| \leq 2\|\boldsymbol{V}\|\,.$$

Then,

$$\begin{aligned} &\sum_{a\in S}\int_{-\infty}^{\infty} h(\omega)\boldsymbol{A}^a_{g,\boldsymbol{H}}(\omega)^\dagger \boldsymbol{A}^a_{g,\boldsymbol{H}}(\omega)\mathrm{d}\omega \\ &= \sum_{a\in S}\sum_{\nu,\nu'\in B(\boldsymbol{H})}\int_{-\infty}^{\infty} h(\omega)(\boldsymbol{A}^a_{\nu'})^\dagger \boldsymbol{A}^a_\nu \hat{g}^*(\omega-\nu')\hat{g}(\omega-\nu)\mathrm{d}\omega \\ &= \sum_{a\in S}\sum_{\nu,\in B(\boldsymbol{H})} h(\nu)(\boldsymbol{A}^a_\nu)^\dagger \boldsymbol{A}^a_\nu \int_{-\infty}^{\infty} |\hat{g}(\omega-\nu)|^2\mathrm{d}\omega \\ &= \sum_{a\in S}\sum_{\nu\in B(\boldsymbol{H})} h(\nu)(\boldsymbol{A}^a_\nu)^\dagger \boldsymbol{A}^a_\nu. \end{aligned}$$

Now, we add the perturbation $\boldsymbol{H}+\boldsymbol{V}$. The insight is that we can introduce an artificial Hamiltonian

$$\bar{\boldsymbol{H}} := \sum_{E\in\mathrm{spec}(\boldsymbol{H})} E \sum_{E'\approx E} \boldsymbol{P}_{E'} \quad \text{such that} \quad \|\bar{\boldsymbol{H}} - \boldsymbol{H}\| \leq 2\|\boldsymbol{V}\|$$

with exactly the same spectrum of the original Hamiltonian $\boldsymbol{H}$, but with the basis according to the perturbed Hamiltonian $\boldsymbol{H}'$. Then, the same argument with the artificial Hamiltonian implies

$$\sum_{a\in S}\int_{-\infty}^{\infty} h(\omega)\boldsymbol{A}^a_{g,\bar{\boldsymbol{H}}}(\omega)^\dagger \boldsymbol{A}^a_{g,\bar{\boldsymbol{H}}}(\omega)\mathrm{d}\omega = \sum_{a\in S}\sum_{\nu\in B(\bar{\boldsymbol{H}})} h(\nu)(\boldsymbol{A}^a_{\approx\nu})^\dagger \boldsymbol{A}^a_{\approx\nu}.$$

Lastly, we may bound the difference by the purification

$$\begin{aligned}
&\|\sum_{a\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}^a_{g,\boldsymbol{H}}(\omega)\otimes|a\rangle\otimes|\omega\rangle\mathrm{d}\omega-\sum_{a\in S}\int_{-\infty}^{\infty}\hat{\boldsymbol{A}}^a_{g,\bar{\boldsymbol{H}}}(\omega)\otimes|a\rangle\otimes|\omega\rangle\mathrm{d}\omega\| \\
&=\|\sum_{a\in S}\int_{-\infty}^{\infty}\boldsymbol{A}^a_{\boldsymbol{H}}(t)\otimes|a\rangle\otimes g(t)|t\rangle\mathrm{d}t-\sum_{a\in S}\int_{-\infty}^{\infty}\boldsymbol{A}^a_{\bar{\boldsymbol{H}}}(t)\otimes|a\rangle\otimes g(t)|t\rangle\mathrm{d}t\| \\
&\leq\sqrt{\int_{-\infty}^{\infty}\left|2\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}-\mathrm{e}^{\mathrm{i}\bar{\boldsymbol{H}}t}\|g(t)\right|^2\mathrm{d}t}=O(\frac{\|\boldsymbol{V}\|}{\Delta_\nu}).
\end{aligned}$$

The factor of 2 is due to left and right Hamiltonian evolution

$$\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}-\mathrm{e}^{\mathrm{i}\bar{\boldsymbol{H}}t}\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\bar{\boldsymbol{H}}t}=\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}\boldsymbol{A}^a(\mathrm{e}^{-\mathrm{i}\boldsymbol{H}t}-\mathrm{e}^{-\mathrm{i}\bar{\boldsymbol{H}}t})+(\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}-\mathrm{e}^{\mathrm{i}\bar{\boldsymbol{H}}t})\boldsymbol{A}^a\mathrm{e}^{-\mathrm{i}\bar{\boldsymbol{H}}t}.$$

To evaluate the integral, we use that $\|\mathrm{e}^{\mathrm{i}\boldsymbol{H}t}-\mathrm{e}^{\mathrm{i}\bar{\boldsymbol{H}}t}\|\leq\|\boldsymbol{H}-\bar{\boldsymbol{H}}\|t\leq 2\|\boldsymbol{V}\|t$ and that $g(t)$ is rapidly decaying for large $|t|\geq\frac{1}{\Delta_\nu}$. To conclude the proof, use the purification tricks (Lemma 4.15.1). ∎

Monotonicity of gradient on a subspace

For our proof that BQP-hard Hamiltonians has no suboptimal local minima in section 4.14, we will need the following refinements of Theorem 4.16.1 where the gradient operator acts on a low-energy subspace with an excitation gap. Intuitively, we care only about the Bohr-frequency gap restricted to the low-energy subspace $\boldsymbol{HQ}$ instead of the full Hilbert space; the gradient on that subspace should not be sensitive to the excited states above the excitation gap.

Corollary 4.16.1 (Monotonicity of gradient on a subspace; Corollary 4.12.1 restated)**.** *Consider a Hamiltonian $\boldsymbol{H}=\sum_{\bar{E}}\bar{E}\boldsymbol{P}_{\bar{E}}$ and its perturbation $\boldsymbol{H}':=\boldsymbol{H}+\boldsymbol{V}$. Let $\boldsymbol{P}$ be the ground space projector for $\boldsymbol{H}$ and $\boldsymbol{P}'$ be the corresponding perturbed eigensubspace of $\boldsymbol{H}'$. Let $\boldsymbol{Q}$ be a low-energy eigensubspace projector of $\boldsymbol{H}$ (i.e., $\boldsymbol{Q}=\sum_{E\leq E_{\boldsymbol{Q}}}\boldsymbol{P}_E$ for $E_{\boldsymbol{Q}}\in Spec(\boldsymbol{H})$) with excitation gap $\Delta_{\boldsymbol{Q}}$. Assume $\frac{\|\boldsymbol{V}\|\|\boldsymbol{H}\|}{\Delta_{\boldsymbol{Q}}}\leq\frac{1}{144}\Delta_\nu$ where $\Delta_\nu:=\min_{\nu_1\neq\nu_2\in B(\boldsymbol{H}|_{\boldsymbol{Q}})}|\nu_1-\nu_2|$ is the Bohr-frequency gap of $\boldsymbol{H}$ within the subspace $\boldsymbol{Q}$. For any $\beta,\tau>0$, let $\mathcal{L}=\sum_{a\in S}\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}}$, $\mathcal{L}'=\sum_{a\in S}\mathcal{L}_a^{\beta,\tau,\boldsymbol{H}'}$ be thermal Lindbladians with jumps $\{\boldsymbol{A}^a\}_{a\in S}$, where $\|\boldsymbol{A}^a\|\leq 1$ and the transition weight $\gamma_\beta(\omega)$ is given by Eq. (4.21). Then we have the monotone property that*

$$-\boldsymbol{Q}\mathcal{L}^\dagger[\boldsymbol{H}]\boldsymbol{Q}\geq r\boldsymbol{Q}(\boldsymbol{I}-\boldsymbol{P})-\epsilon\boldsymbol{I}\quad\textit{implies}\quad-\boldsymbol{Q}'\mathcal{L}'^\dagger[\boldsymbol{H}']\boldsymbol{Q}'\geq r\boldsymbol{Q}'(\boldsymbol{I}-\boldsymbol{P}')-\epsilon'\boldsymbol{I}$$

where $\boldsymbol{Q}'$ projects onto the perturbed eigensubspace of $\boldsymbol{H}'$ identified with $\boldsymbol{Q}$, and

$$\epsilon' \leq \epsilon + |S| \cdot O\bigg(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0^{2/3}}{\tau^{1/3}} + \frac{\Lambda_0}{\sqrt{\Delta_\nu \tau}} + \frac{\Lambda_0}{\sqrt{\Delta_{\boldsymbol{Q}}\tau}} + \frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta} + \frac{\mathrm{e}^{-\beta\Delta_{\boldsymbol{Q}}/4}}{\beta} \\ + \bigg(1 + \frac{\Lambda_0}{\Delta_\nu}\bigg)\frac{\|\boldsymbol{V}\|\|\boldsymbol{H}\|}{\Delta_{\boldsymbol{Q}}} + r\bigg(\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}} + \frac{\|\boldsymbol{V}\|}{\Delta_\nu}\bigg)\bigg).$$

Proof. The idea is that $\boldsymbol{Q}'\mathcal{L}'^{\dagger}[\boldsymbol{H}']\boldsymbol{Q}'$ essentially depends only on the low energy subspace $\boldsymbol{Q}'$ and the corresponding restricted transition $\boldsymbol{Q}'\boldsymbol{A}^a\boldsymbol{Q}'$.

$$\begin{aligned}
\boldsymbol{Q}'\mathcal{L}'^{\dagger}[\boldsymbol{H}']\boldsymbol{Q}' &\overset{E_1}{\approx} \sum_{a\in S}\int_{-\infty}^{\infty} \theta(\omega')\boldsymbol{Q}'\hat{\boldsymbol{A}}^a(\omega')^\dagger\hat{\boldsymbol{A}}^a(\omega')\boldsymbol{Q}'\mathrm{d}\omega' \\
&\overset{E_2}{\approx} \sum_{a\in S}\int_{-\infty}^{\infty} \theta(\omega')\boldsymbol{Q}'\hat{\boldsymbol{A}}^a(\omega')^\dagger\boldsymbol{Q}'\hat{\boldsymbol{A}}^a(\omega')\boldsymbol{Q}'\mathrm{d}\omega' \\
&= \sum_{a\in S}\int_{-\infty}^{\infty} \theta(\omega')\hat{\boldsymbol{R}}'^a(\omega')^\dagger\hat{\boldsymbol{R}}'^a(\omega')\mathrm{d}\omega' \\
&\overset{E_3}{\approx} \mathcal{L}^{\dagger\beta,\tau,\boldsymbol{H}'_{\boldsymbol{Q}'}}_{\{\boldsymbol{R}'^a\}}[\boldsymbol{H}'_{\boldsymbol{Q}'}] \\
&\overset{E_4}{\approx} \mathcal{L}^{\dagger\beta,\tau,\boldsymbol{H}'_{\boldsymbol{Q}'}}_{\{\boldsymbol{R}^a\}}[\boldsymbol{H}'_{\boldsymbol{Q}'}].
\end{aligned}$$

The approximation E_2 inserts the low-energy projector $\boldsymbol{Q}'$. To do so, we resolves the identity by $\boldsymbol{I} = \boldsymbol{Q}' + (\boldsymbol{I} - \boldsymbol{Q}')$ and uses that

$$\begin{aligned}
&\sum_{a\in S}\int_{-\infty}^{\infty} \theta(\omega')\boldsymbol{Q}'\hat{\boldsymbol{A}}^a(\omega')^\dagger(\boldsymbol{I}-\boldsymbol{Q}')\hat{\boldsymbol{A}}^a(\omega')\boldsymbol{Q}'\mathrm{d}\omega' \\
&\overset{E_{21}}{\approx} \sum_{a\in S}\int_{-\infty}^{\infty} \theta(\omega')\hat{\boldsymbol{S}}^a(\omega')^\dagger\hat{\boldsymbol{S}}^a(\omega')\mathrm{d}\omega' \\
&\overset{E_{22}}{\approx} 0.
\end{aligned}$$

That is, we need the excitation gap to be large so that $(\boldsymbol{I}-\boldsymbol{Q}')\boldsymbol{A}^a\boldsymbol{Q}'$ have a vanishing contribution to the gradient. These error combines $E_2 = E_{21} + E_{22}$, where

$$\begin{aligned}
E_{21} &\leq 2\|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|\|\theta\|_\infty\|f_\tau\cdot(1-\hat{s}_\mu)\|_2\|f_\tau\|_2 = |S|O(\frac{\Lambda_0}{\sqrt{\mu\tau}}) \\
E_{22} &\leq \|\sum_{a\in S}\boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\| \max_{\omega'\geq\Delta_{\boldsymbol{Q}'}-\mu}|\theta(\omega')|\|f_\tau\cdot\hat{s}_\mu\|_2^2 = |S|\frac{\mathrm{e}^{-\beta(\Delta_{\boldsymbol{Q}'}-\mu)/2}}{\beta}.
\end{aligned}$$

Thus, we consider a safe choice of $\mu = \Delta_{\boldsymbol{Q}'}/2$. Also, since the perturbation is small by assumption, $\|\boldsymbol{V}\| \leq \frac{1}{144}\Delta_\nu\frac{\Delta_{\boldsymbol{Q}}}{\|\boldsymbol{H}\|} \leq \frac{1}{144}\Delta_{\boldsymbol{Q}}$, the excitation gap remains large $\Delta_{\boldsymbol{Q}'} \geq \Delta_{\boldsymbol{Q}} - 2\|\boldsymbol{V}\| \geq \Delta_{\boldsymbol{Q}} - \frac{1}{72}\Delta_{\boldsymbol{Q}} \geq \Delta_{\boldsymbol{Q}}/2$. The third line (after approximation

E_2) formally disposes of the excited state of $\boldsymbol{H}'$ above $E'_{\boldsymbol{Q}'}$ by defining a modified Hamiltonian

$$\boldsymbol{H}'_{\boldsymbol{Q}'} := \boldsymbol{H}'\boldsymbol{Q}' + E'_{\boldsymbol{Q}'}(\boldsymbol{I} - \boldsymbol{Q}')$$

for the Fourier transform $\hat{\boldsymbol{R}'^a}(\omega')$. This Hamiltonian is merely a proof artifact, and one can also set the energy of the excited subspace $\boldsymbol{I} - \boldsymbol{Q}'$ to be infinity. The error E_3 is merely the error to put it back to the form of an energy gradient.

The approximation E_4 changes the jumps with norm bounded by

$$\begin{aligned}&\|\sum_{a\in S} \boldsymbol{R}'^a \otimes |a\rangle - \sum_{a\in S} \boldsymbol{R}^a \otimes |a\rangle\| \\ &\leq 2\sqrt{\|\sum_{a\in S} \boldsymbol{A}^{a\dagger}\boldsymbol{A}^a\|}\|\boldsymbol{Q} - \boldsymbol{Q}'\| = |S|O(\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}})\end{aligned}$$

using subspace perturbation bounds $\|\boldsymbol{Q} - \boldsymbol{Q}'\| \leq 8\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}$ (Lemma 4.16.3). Since the (suitably normalized) gradient operator $\frac{1}{2|S|\|\boldsymbol{O}\|} \cdot \mathcal{L}^\dagger[\boldsymbol{O}]$ can be block-encoded using $O(1)$ block-encodings of the jumps (Theorem 4.10.1, Proposition 4.10.7, Proposition 4.10.8), perturbation to the jumps propagates to the gradient operator by

$$E_4 = O\left(\left\|\sum_a \boldsymbol{R}'^a|a\rangle - \sum_a \boldsymbol{R}^a|a\rangle\right\| \cdot \|\boldsymbol{H}'_{\boldsymbol{Q}'}\|\right) = |S|O\left(\|\boldsymbol{H}'_{\boldsymbol{Q}'}\|\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}\right).$$

To summarize, the above gives the bound

$$\|\boldsymbol{Q}'\mathcal{L}'^\dagger[\boldsymbol{H}']\boldsymbol{Q}' - \mathcal{L}^{\dagger\beta,\tau,\boldsymbol{H}'_{\boldsymbol{Q}'}}_{\{\boldsymbol{R}^a\}}[\boldsymbol{H}'_{\boldsymbol{Q}'}]\| \leq |S| \cdot O\left(\frac{1}{\tau} + \frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}} + \frac{\Lambda_0}{\sqrt{\Delta_{\boldsymbol{Q}}\tau}} + \frac{e^{-\beta\Delta_{\boldsymbol{Q}}/4}}{\beta}\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}\|\boldsymbol{H}'_{\boldsymbol{Q}'}\|\right) \quad (4.148)$$

using $\|\boldsymbol{H}'\| \leq \|\boldsymbol{H}\| + \|\boldsymbol{V}\| \leq 2\|\boldsymbol{H}\|$ and $\Delta_{\boldsymbol{Q}'} \geq \Delta_{\boldsymbol{Q}}/2$. Similarly, the

$$\boldsymbol{Q}\mathcal{L}^\dagger[\boldsymbol{H}]\boldsymbol{Q} \overset{E_5}{\approx} \mathcal{L}^{\dagger\beta,\tau,\boldsymbol{H}_{\boldsymbol{Q}}}_{\{\boldsymbol{R}^a\}}[\boldsymbol{H}_{\boldsymbol{Q}}] \quad \text{for} \quad \boldsymbol{H}_{\boldsymbol{Q}} := \boldsymbol{H}\boldsymbol{Q} + E_{\boldsymbol{Q}}(\boldsymbol{I} - \boldsymbol{Q})$$

with E_5 also bounded by the RHS of Eq. (4.16) but with $\|\boldsymbol{V}\| \to 0$.

Now, we may use the monotonicity of gradient (Theorem 4.16.1) for Hamiltonian pairs $\boldsymbol{H}_{\boldsymbol{Q}}$ and $\boldsymbol{H}'_{\boldsymbol{Q}'}$, jumps $\{\boldsymbol{R}^a\}_{a\in S}$, with the characteristic Bohr-frequency gap $\Delta_\nu = \min_{\nu_1\neq\nu_2\in B(\boldsymbol{H}_{\boldsymbol{Q}})}|\nu_1 - \nu_2|$. The modified Hamiltonian perturbation $\|\boldsymbol{H}_{\boldsymbol{Q}} - \boldsymbol{H}'_{\boldsymbol{Q}'}\|$ is bounded by the sum of the following two errors:

$$\|\boldsymbol{H}\boldsymbol{Q} - \boldsymbol{H}'\boldsymbol{Q}'\| = \|\boldsymbol{H}\boldsymbol{Q} - \boldsymbol{H}\boldsymbol{Q}' + \boldsymbol{H}\boldsymbol{Q}' - \boldsymbol{H}'\boldsymbol{Q}'\| \leq 8\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}\|\boldsymbol{H}\| + \|\boldsymbol{V}\| \leq 9\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}\|\boldsymbol{H}\|, \quad (4.149)$$

$$\|E_{\boldsymbol{Q}}(\boldsymbol{I}-\boldsymbol{Q})-E'_{\boldsymbol{Q}'}(\boldsymbol{I}-\boldsymbol{Q}')\| = \|E_{\boldsymbol{Q}}(\boldsymbol{I}-\boldsymbol{Q})-E_{\boldsymbol{Q}}(\boldsymbol{I}-\boldsymbol{Q}')+E_{\boldsymbol{Q}}(\boldsymbol{I}-\boldsymbol{Q}')-E'_{\boldsymbol{Q}'}(\boldsymbol{I}-\boldsymbol{Q}')\|$$
$$\leq 8E_{\boldsymbol{Q}}\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}+\|\boldsymbol{V}\| \leq 9\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}\|\boldsymbol{H}\|. \tag{4.150}$$

So $\|\boldsymbol{H}_{\boldsymbol{Q}}-\boldsymbol{H}'_{\boldsymbol{Q}'}\| \leq 18\|\boldsymbol{V}\|\|\boldsymbol{H}\|/\Delta_{\boldsymbol{Q}}$, which is less than $\Delta_\nu/8$ by the assumption in the corollary statement so we may apply Theorem 4.16.1. Note that error due to perturbing the $\boldsymbol{Q}(\boldsymbol{I}-\boldsymbol{P})$ can be bounded directly by

$$\|\boldsymbol{Q}'(\boldsymbol{I}-\boldsymbol{P}')-\boldsymbol{Q}(\boldsymbol{I}-\boldsymbol{P})\| \leq \|\boldsymbol{Q}'-\boldsymbol{Q}\|+\|\boldsymbol{P}-\boldsymbol{P}'\| = O\Big(\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}+\frac{\|\boldsymbol{V}\|}{\Delta_\nu}\Big). \tag{4.151}$$

Apply Theorem 4.16.1 with $\boldsymbol{H}=\boldsymbol{H}_{\boldsymbol{Q}}, \boldsymbol{H}'=\boldsymbol{H}'_{\boldsymbol{Q}'}, \boldsymbol{O}=\boldsymbol{Q}(\boldsymbol{I}-\boldsymbol{P})$, and $\boldsymbol{O}'=\boldsymbol{Q}'(\boldsymbol{I}-\boldsymbol{P}')$, $\delta_\lambda \leq \|\boldsymbol{H}_{\boldsymbol{Q}}-\boldsymbol{H}_{\boldsymbol{Q}'}\|$, and $\theta_{\max}=O(\Lambda_0)$, along with the additional approximation error in Eq. (4.16), we obtain the result as advertised in the corollary statement. ■

Corollary 4.16.2 (Monotonicity of gradient on a subspace under off-block-diagonal perturbation; Corollary 4.12.2 restated)**.** *In the setting of Corollary 4.16.1, instead assume* $\frac{\|\boldsymbol{V}\|}{\Delta_\nu}, \frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}} \leq (const.)$*, and that the perturbation is off-block-diagonal, i.e.,* $\boldsymbol{Q}\boldsymbol{V}\boldsymbol{Q}=(\boldsymbol{I}-\boldsymbol{Q})\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{Q})=0$*. Then,*

$$-\boldsymbol{Q}\mathcal{L}^\dagger[\boldsymbol{H}]\boldsymbol{Q} \succeq r\boldsymbol{Q}(\boldsymbol{I}-\boldsymbol{P})-\epsilon\boldsymbol{I} \quad \textit{implies} \quad -\boldsymbol{Q}'\mathcal{L}'^\dagger[\boldsymbol{H}']\boldsymbol{Q}' \succeq r\boldsymbol{Q}'(\boldsymbol{I}-\boldsymbol{P}')-\epsilon'\boldsymbol{I},$$

where

$$\epsilon' \leq \epsilon+|S|\cdot O\bigg(\frac{1}{\tau}+\frac{\|\boldsymbol{H}\|^{3/4}}{\tau^{1/4}}+\frac{\Lambda_0^{2/3}}{\tau^{1/3}}+\frac{\Lambda_0}{\sqrt{\Delta_\nu\tau}}+\frac{\Lambda_0}{\sqrt{\Delta_{\boldsymbol{Q}}\tau}}+\frac{\mathrm{e}^{-\beta\Delta_\nu/4}}{\beta}+\frac{\mathrm{e}^{-\beta\Delta_{\boldsymbol{Q}}/4}}{\beta}$$
$$+\frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}}+\|\boldsymbol{H}_{\boldsymbol{Q}}\|\cdot\Big(\frac{\|\boldsymbol{H}_{\boldsymbol{Q}}\|\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}\Delta_\nu}+\frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}\Delta_\nu}\Big)+r\Big(\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}+\frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}\Delta_\nu}\Big)\bigg).$$

Proof. When the perturbation $\boldsymbol{V}$ is off-diagonal in the energy eigenbasis of $\boldsymbol{H}$, we can use tighter bounds on the changes in eigenvalues and eigensubspaces of $\boldsymbol{H}'$ from Lemma 4.16.4, which implies

$$\delta_\lambda = \max_j |E_j-E'_j| = O\Big(\frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}}\Big)$$
$$\|\boldsymbol{H}_{\boldsymbol{Q}}-\boldsymbol{H}'_{\boldsymbol{Q}'}\| = O\bigg(\|\boldsymbol{H}_{\boldsymbol{Q}}\|\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}+\frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}}\bigg).$$
$$\|\boldsymbol{P}-\boldsymbol{P}'\| \leq O\Big(\frac{\|\boldsymbol{V}\|}{\Delta_{\boldsymbol{Q}}}+\frac{\|\boldsymbol{V}\|^2}{\Delta_{\boldsymbol{Q}}\Delta_\nu}\Big).$$

The second line and third line are due to rotation and then subspace perturbation (Lemma 4.16.3). Essentially, this is because (1) all subspace rotations are small ($\frac{\|V\|}{\Delta_Q} \ll 1$) and (2) the energy perturbation is *smaller* than the level spacing ($\frac{\|V\|^2}{\Delta_Q} \ll \Delta_\nu$). We then follow the same argument in the proof of Corollary 4.16.1 above with the improved bounds for Eqs. (4.149), (4.150), and (4.151). We can also use an improve bound of $\theta_{\max} \leq \|H_Q\|$. Together these improvements yield the better error bound on ϵ' as advertised. ■

Example where perturbation kills energy gradient

We present an example where despite $\|V\| \ll \Delta_\nu$, the gradient is lost due to the perturbation. Therefore, the resulting change in gradient is not multiplicative ($1 - \frac{\|V\|}{\Delta_\nu}$), but merely additive. If the gradient is polynomially small, we need $\frac{\|V\|}{\Delta_\nu}$ to be also polynomially small to secure the gradient.

Proposition 4.16.7 (Perturbation kills the gradient)**.** *Let* $H = Z = |1\rangle\langle 1| - |-1\rangle\langle -1|$, $A = Z + \epsilon X$, *and* $V = \epsilon X$. *Then, for the* $\beta, \tau \to \infty$ *heat bath Lindbladian, we have that*

$$\mathcal{L}^\dagger[H] \leq -2\epsilon^2(I - P) \quad \textit{but} \quad \mathcal{L}'^\dagger[H'] = 0.$$

Proof.

$$\mathcal{L}^\dagger[H] = \epsilon^2\left(|1\rangle\langle -1|H|-1\rangle\langle 1| - \frac{1}{2}|1\rangle\langle 1|H - \frac{1}{2}H|1\rangle\langle 1|\right) = -2\epsilon^2|1\rangle\langle 1|.$$

But $\mathcal{L}'^\dagger[H'] = 0$ because $[H', A] = 0$, i.e., A is diagonal in the new energy basis. ■

Bibliography

[1] Nilin Abrahamsen. A polynomial-time algorithm for ground states of spin trees. *arXiv preprint arXiv:1907.04862*, 2019. 204

[2] Naman Agarwal, Zeyuan Allen-Zhu, Brian Bullins, Elad Hazan, and Tengyu Ma. Finding approximate local minima faster than gradient descent. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1195–1199, 2017. doi: 10.1145/3055399.3055464. 199

[3] Dorit Aharonov, Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM Journal on Computing*, 37(1):166–194, April 2007. ISSN 0097-5397. doi: 10.1137/S0097539705447323. URL http://dx.doi.org/10.1137/S0097539705447323. 201, 203, 269

[4] Dorit Aharonov, Daniel Gottesman, Sandy Irani, and Julia Kempe. The power of quantum systems on a line. *Communications in mathematical physics*, 287 (1):41–65, 2009. 9

[5] Amir Ali Ahmadi and Jeffrey Zhang. On the complexity of finding a local minimizer of a quadratic function over a polytope. *Mathematical Programming*, 195(1-2):783–792, 2022. doi: 10.1007/s10107-021-01714-2. 199, 216

[6] Maxwell Aifer, Kaelan Donatella, Max Hunter Gordon, Thomas Ahle, Daniel Simpson, Gavin E Crooks, and Patrick J Coles. Thermodynamic linear algebra. *arXiv preprint arXiv:2308.05660*, 2023. 206

[7] Tameem Albash and Daniel A. Lidar. Adiabatic quantum computation. *Reviews of Modern Physics*, 90:015002, 2016. 9

[8] Robert Alicki. On the detailed balance condition for non-hamiltonian systems. *Reports on Mathematical Physics*, 10:249–258, 1976. URL https://api.semanticscholar.org/CorpusID:120482500. 191

[9] Mohammad H. Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko. Quantum Boltzmann machine. *Physical Review X*, 8(2): 021050, 2018. doi: 10.1103/PhysRevX.8.021050. arXiv: 1601.02036 10

[10] Anurag Anshu, Itai Arad, and David Gosset. An area law for 2d frustration-free spin systems. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 12–18, 2022. 60

[11] Joran van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf. Quantum SDP-solvers: Better upper and lower bounds. *Quantum*, 4:230, 2020. doi: 10.22331/q-2020-02-14-230. Earlier version in FOCS'17. arXiv: 1705.01843 10, 169

[12] Joran van Apeldoorn, Arjan Cornelissen, András Gilyén, and Giacomo Nannicini. Quantum tomography using state-preparation unitaries. In *Proceedings of the 36th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1265–1318, 2023. doi: 10.1137/1.9781611977554.ch47. arXiv: 2207.08800 43

[13] Itai Arad, Zeph Landau, Umesh Vazirani, and Thomas Vidick. Rigorous RG algorithms and area laws for low energy eigenstates in 1D. *Commun. Math. Phys.*, 356(1):65–105, 2017. doi: 10.1007/s00220-017-2973-z. 204

[14] Ryan Babbush, Nathan Wiebe, Jarrod McClean, James McClain, Hartmut Neven, and Garnet Kin-Lic Chan. Low-depth quantum simulation of materials. *Phys. Rev. X*, 8:011044, 2018. doi: 10.1103/PhysRevX.8.011044. URL https://link.aps.org/doi/10.1103/PhysRevX.8.011044. 1, 9, 138

[15] Sven Bachmann, Spyridon Michalakis, Bruno Nachtergaele, and Robert Sims. Automorphic equivalence within gapped phases of quantum lattice systems. *Communications in Mathematical Physics*, 309(3):835–871, 2012. 185, 186

[16] Ivan Bardet, Ángela Capel, Li Gao, Angelo Lucia, David Pérez-García, and Cambyse Rouzé. Entropy decay for Davies semigroups of a one dimensional quantum lattice, 2021. arXiv: 2112.00601 22, 60

[17] Federico Becca and Sandro Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 2017. doi: 10.1017/9781316417041. 197

[18] Dominic W. Berry, Richard Cleve, and Sevag Gharibian. Gate-efficient discrete simulations of continuous-time quantum query algorithms. *Quantum Information and Computation*, 14(1&2):1–30, 2014. doi: 10.26421/QIC14.1-2-1. arXiv: 1211.4637 124, 126, 127, 128

[19] Dominic W Berry, Andrew M Childs, and Robin Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 792–809. IEEE, 2015. 247, 249

[20] Dominic W. Berry, Yuan Su, Casper Gyurik, Robbie King, Joao Basso, Alexander Del Toro Barba, Abhishek Rajput, Nathan Wiebe, Vedran Dunjko, and Ryan Babbush. Quantifying quantum advantage in topological data analysis, 2022. arXiv: 2209.13581 27, 49

[21] Rajendra Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, 1997. doi: 10.1007/978-1-4612-0653-8. 108, 111, 112

[22] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013. 296

[23] Kurt Binder and A Peter Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.*, 58(4):801, 1986. doi: 10.1103/RevModPhys.58.801. 197

[24] Sergio Boixo, Emanuel Knill, and Rolando D. Somma. Fast quantum algorithms for traversing paths of eigenstates, 2010. arXiv: 1005.3034 46, 131

[25] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. 199, 201

[26] Fernando G. S. L. Brandao and Michael J. Kastoryano. Finite correlation length implies efficient preparation of quantum thermal states, 2019. arXiv: 1609.07877 137

[27] Fernando G. S. L. Brandão and Krysta M. Svore. Quantum speed-ups for solving semidefinite programs. In *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 415–426, 2017. doi: 10.1109/FOCS.2017.45. URL http://ieee-focs.org/FOCS-2017-Papers/3464a415.pdf. arXiv: 1609.05537 10, 169

[28] Heinz-Peter Breuer and Francesco Petruccione. *The theory of open quantum systems*. Oxford University Press, USA, 2002. 197, 199

[29] Heinz-Peter Breuer and Francesco Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, 2007. ISBN 9780199213900. doi: 10.1093/acprof:oso/9780199213900.001.0001. 3, 30

[30] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. Quantum computing enhanced computational catalysis. *Physical Review Research*, 3(3), Jul 2021. ISSN 2643-1564. doi: 10.1103/physrevresearch.3.033055. URL http://dx.doi.org/10.1103/PhysRevResearch.3.033055. 1, 9, 138, 168

[31] Ángela Capel, Cambyse Rouzé, and Daniel Stilck França. The modified logarithmic Sobolev inequality for quantum spin systems: classical and commuting nearest neighbour interactions, 2021. arXiv: 2009.11817 5, 6, 16, 60, 168, 188, 261

[32] Eric A Carlen and Jan Maas. Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. *Journal of Functional Analysis*, 273(5):1810–1869, 2017. 191, 196

[33] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017. ISSN 0036-8075. doi: 10.1126/science.aag2302. 197, 204

[34] David Ceperley and Berni Alder. Quantum Monte Carlo. *Science*, 231(4738): 555–560, 1986. ISSN 0036-8075. doi: 10.1126/science.231.4738.555. URL https://science.sciencemag.org/content/231/4738/555. 1, 197

[35] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9): 625–644, 2021. doi: 10.1038/s42254-021-00348-9. 199, 202, 221

[36] Christopher Chamberland, Kyungjoo Noh, Patricio Arrangoiz-Arriola, Earl T. Campbell, Connor T. Hann, Joseph Iverson, Harald Putterman, Thomas C. Bohdanowicz, Steven T. Flammia, Andrew Keller, Gil Refael, John Preskill, Liang Jiang, Amir H. Safavi-Naeini, Oskar Painter, and Fernando G. S. L. Brandão. Building a fault-tolerant quantum computer using concatenated cat codes, 2020. 1

[37] Christopher Chamberland, Kyungjoo Noh, Patricio Arrangoiz-Arriola, Earl T. Campbell, Connor T. Hann, Joseph K. Iverson, Harald Putterman, Thomas C. Bohdanowicz, Steven T. Flammia, A. J. Keller, Gil Refael, John Preskill, Liang Jiang, Amir H. Safavi-Naeini, Oskar J. Painter, and Fernando G. S. L. Brandão. Building a fault-tolerant quantum computer using concatenated cat codes. *PRX Quantum*, 2020. 9, 138, 168

[38] Chi-Fang Chen and Fernando G. S. L. Brandão. Fast thermalization from the eigenstate thermalization hypothesis, 2021. arXiv: 2112.07646 12, 15, 16, 17, 18, 19, 22, 26, 79, 137, 138, 329

[39] Chi-Fang Chen, Michael J Kastoryano, Fernando GSL Brandão, and András Gilyén. Quantum thermal state preparation. *arXiv preprint arXiv:2303.18224*, 2023. 199, 204, 207, 208, 243, 246, 247, 248, 249, 250, 253, 286, 287, 289

[40] Chi-Fang Chen, Michael J. Kastoryano, Fernando G. S. L. Brandão, and András Gilyén. Quantum thermal state preparation. arXiv: 2303.18224, 2023. 105, 138, 139, 141, 143, 144, 145, 148, 150, 159, 160, 164, 165, 167, 169, 186, 193, 194

[41] Chi-Fang Chen, Michael J Kastoryano, and András Gilyén. An efficient and exact noncommutative quantum gibbs sampler. *arXiv preprint arXiv:2311.09207*, 2023. 201, 204

[42] Chi-Fang Chen, Michael J. Kastoryano, and András Gilyén. Gibbs sampling via exactly detailed balanced lindbladians, in preparation. 33, 60

[43] Andrew M. Childs and Tongyang Li. Efficient simulation of sparse Markovian quantum dynamics. *Quantum Information and Computation*, 17(11&12): 901–947, 2017. doi: 10.26421/QIC17.11-12. arXiv: 1611.05543 39, 159

[44] Andrew M. Childs and Nathan Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Information and Computation*, 12(11&12):901–924, 2012. doi: 10.26421/QIC12.11-12. arXiv: 1202.5822 26

[45] Andrew M Childs, Robin Kothari, and Rolando D Somma. Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. *SIAM Journal on Computing*, 46(6):1920–1950, 2017. 247, 249

[46] Kenny Choo, Giuseppe Carleo, Nicolas Regnault, and Titus Neupert. Symmetries and many-body excitations with neural-network quantum states. *Phys. Rev. Lett.*, 121:167204, 2018. doi: 10.1103/PhysRevLett.121.167204. URL https://link.aps.org/doi/10.1103/PhysRevLett.121.167204. 204

[47] Kenny Choo, Antonio Mezzacapo, and Giuseppe Carleo. Fermionic neural-network states for ab-initio electronic structure. *Nature Communications*, 11 (1):2368, May 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-15724-9. URL https://doi.org/10.1038/s41467-020-15724-9. 204

[48] Pete L. Clark. Honors calculus. 2014. URL http://alpha.math.uga.edu/~pete/2400full.pdf. 95

[49] Richard Cleve and Chunhao Wang. Efficient quantum algorithms for simulating Lindblad evolution. In *Proceedings of the 44th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 17:1–17:14, 2017. doi: 10.4230/LIPIcs.ICALP.2017.17. arXiv: 1612.09512 19, 24, 37, 38, 39, 41, 42, 122, 124, 127, 128, 159

[50] Patrick J. Coles, Collin Szczepanski, Denis Melanson, Kaelan Donatella, Antonio J. Martinez, and Faris Sbahi. Thermodynamic ai and the fluctuation frontier. 2023. 206

[51] Mirko Consiglio, Jacopo Settino, Andrea Giordano, Carlo Mastroianni, Francesco Plastina, Salvatore Lorenzo, Sabrina Maniscalco, John Goold, and Tony J. G. Apollaro. Variational Gibbs state preparation on NISQ devices. arXiv: 2303.11276, 2023. 19

[52] Philippe Corboz. Variational optimization with infinite projected entangled-pair states. *Phys. Rev. B*, 94:035133, 2016. doi: 10.1103/PhysRevB.94.035133. URL https://link.aps.org/doi/10.1103/PhysRevB.94.035133. 204

[53] Toby S. Cubitt. Dissipative ground state preparation and the dissipative quantum eigensolver, 2023. arXiv: 2303.11962 38

[54] Alexander M. Dalzell, Sam McArdle, Mario Berta, Przemyslaw Bienias, Chi-Fang Chen, András Gilyén, Connor T. Hann, Michael J. Kastoryano, Emil T. Khabiboulline, Aleksander Kubica, Grant Salton, Samson Wang, and Fernando G. S. L. Brandão. Quantum algorithms: A survey of applications and end-to-end complexities. arXiv: 2310.03011, 2023. 138

[55] E Brian Davies. Generators of dynamical semigroups. *Journal of Functional Analysis*, 34(3):421–432, 1979. 197, 199

[56] Edward Brian Davies. Markovian master equations. *Communications in Mathematical Physics*, 39(2):91–110, 1974. doi: 10.1007/BF01608389. 3, 13, 101, 151, 196

[57] Edward Brian Davies. Quantum theory of open systems. *IMA*, 1976. 256

[58] Edward Brian Davies. Markovian master equations. II. *Mathematische Annalen*, 219(2):147–158, 1976. doi: 10.1007/BF01351898. 3, 13, 101, 151

[59] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Quantum entanglement in neural network states. *Phys. Rev. X*, 7:021021, 2017. doi: 10.1103/PhysRevX.7.021021. URL https://link.aps.org/doi/10.1103/PhysRevX.7.021021. 197, 204

[60] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Machine learning topological states. *Phys. Rev. B*, 96:195145, 2017. doi: 10.1103/PhysRevB.96.195145. URL https://link.aps.org/doi/10.1103/PhysRevB.96.195145. 204

[61] Zhiyan Ding, Chi-Fang Chen, and Lin Lin. Single-ancilla ground state preparation via lindbladians. *arXiv preprint arXiv:2308.15676*, 2023. 60, 138, 188, 201, 204

[62] Luca D'Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics*, 65(3):239–362, May 2016. ISSN 1460-6976. doi: 10.1080/00018732.2016.1198134. URL http://dx.doi.org/10.1080/00018732.2016.1198134. 10

[63] Samuel Frederick Edwards and Phil W Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975. 197

[64] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. arXiv: quant-ph/0001106, 2000. 9, 169

[65] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. *arXiv preprint quant-ph/0001106*, 2000. 197

[66] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. arXiv: 1411.4028, 2014. 169

[67] Francesco Ferrari, Federico Becca, and Juan Carrasquilla. Neural gutzwiller-projected variational wave functions. *Phys. Rev. B*, 100:125131, 2019. doi: 10.1103/PhysRevB.100.125131. URL https://link.aps.org/doi/10.1103/PhysRevB.100.125131. 204

[68] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6-7):467–488, 1982. doi: 10.1007/BF02650179. 9, 138

[69] Veronica Umanità Franco Fagnola. Generators of detailed balance quantum Markov semigroups. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 10(03):335–363, 2007. doi: 10.1142/S0219025707002762. arXiv: 0707.2147 152, 191, 192

[70] Sevag Gharibian and François Le Gall. Dequantizing the quantum singular value transformation: hardness and applications to quantum chemistry and the quantum PCP conjecture. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 19–32, 2022. 9, 197, 203

[71] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/gilmer17a.html. 197

[72] András Gilyén and Or Sattath. On preparing ground states of gapped Hamiltonians: An efficient quantum Lovász local lemma. In *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 439–450, 2017. doi: 10.1109/FOCS.2017.47. URL http://ieee-focs.org/FOCS-2017-Papers/3464a439.pdf. arXiv: 1611.08571 38

[73] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st ACM Symposium on the Theory of Computing (STOC)*, pages 193–204, 2019. doi: 10.1145/3313276.3316366. arXiv: 1806.01838 128, 131, 137, 140

[74] Ivan Glasser, Nicola Pancotti, Moritz August, Ivan D. Rodriguez, and J. Ignacio Cirac. Neural-network quantum states, string-bond states, and chiral topological states. *Phys. Rev. X*, 8:011006, 2018. doi: 10.1103/PhysRevX.8.011006. URL https://link.aps.org/doi/10.1103/PhysRevX.8.011006. 204

[75] Roy J Glauber. Time-dependent statistics of the ising model. *Journal of mathematical physics*, 4(2):294–307, 1963. doi: 10.1063/1.1703954. 4

[76] Daniel Gottesman and Sandy Irani. The quantum and classical complexity of translationally invariant tiling and Hamiltonian problems. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 95–104. IEEE, 2009. 9

[77] Harper R Grimsley, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications*, 10(1):3007, 2019. doi: 10.1038/s41467-019-10988-2. 199, 202, 221, 222

[78] Jinkang Guo, Oliver Hart, Chi-Fang Chen, Aaron J. Friedman, and Andrew Lucas. Quantum active matter, in preparation. 151

[79] Jeongwan Haah. Product Decomposition of Periodic Functions in Quantum Signal Processing. *Quantum*, 3:190, 2019. doi: 10.22331/q-2019-10-07-190. arXiv: 1806.10236 51

[80] Jeongwan Haah, Matthew B. Hastings, Robin Kothari, and Guang Hao Low. Quantum algorithm for simulating real time evolution of lattice hamiltonians. In *Proceedings of the 59th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 350–360, 2018. doi: 10.1109/FOCS.2018.00041. arXiv: 1801.03922 187, 188

[81] Howard E. Haber. Notes on the matrix exponential and logarithm, 2021. URL http://scipp.ucsc.edu/~haber/webpage/MatrixExpLog.pdf. 116

[82] Reza Haghshenas, Matthew J. O'Rourke, and Garnet Kin-Lic Chan. Conversion of projected entangled pair states into a canonical form. *Phys. Rev. B*, 100:054404, 2019. doi: 10.1103/PhysRevB.100.054404. URL https://link.aps.org/doi/10.1103/PhysRevB.100.054404. 197, 204

[83] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009. doi: 10.1103/PhysRevLett.103.150502. arXiv: 0811.3171 2

[84] Matthew B. Hastings. An area law for one-dimensional quantum systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2007:P08024 – P08024, 2007. 60, 168

[85] Matthew B Hastings and Xiao-Gang Wen. Quasiadiabatic continuation of quantum states: The stability of topological ground-state degeneracy and emergent gauge invariance. *Physical review b*, 72(4):045141, 2005. 185

[86] Wilfred Keith Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi: 10.1093/biomet/57.1.97. 3

[87] Patrick Hayden, Debbie W Leung, and Andreas Winter. Aspects of generic entanglement. *Commun. Math. Phys.*, 265:95–117, 2006. 226

[88] Mohamed Hibat-Allah, Martin Ganahl, Lauren E Hayward, Roger G Melko, and Juan Carrasquilla. Recurrent neural network wave functions. *Phys. Rev. Res.*, 2(2):023358, 2020. doi: 10.1103/PhysRevResearch.2.023358. 197, 204

[89] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964. doi: 10.1103/PhysRev.136.B864. URL https://link.aps.org/doi/10.1103/PhysRev.136.B864. 1, 197

[90] Zoe Holmes, Gopikrishnan Muraleedharan, Rolando D Somma, Yigit Subasi, and Burak Şahinoğlu. Quantum algorithms from fluctuation theorems: Thermal-state preparation. *Quantum*, 6:825, 2022. 137

[91] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613), 2022. ISSN 1095-9203. doi: 10.1126/science.abk3333. URL http://dx.doi.org/10.1126/science.abk3333. 197

[92] William J. Huggins, Kianna Wan, Jarrod McClean, Thomas E. O'Brien, Nathan Wiebe, and Ryan Babbush. Nearly optimal quantum algorithm for estimating multiple expectation values. arXiv: 2111.09283, 2021. 43

[93] Katharine Hyatt and E. M. Stoudenmire. DMRG Approach to Optimizing Two-Dimensional Tensor Networks. *arXiv preprint arXiv:1908.08833*, 2019. 197, 204

[94] Chi Jin, Praneeth Netrapalli, and Michael I. Jordan. Accelerated Gradient Descent Escapes Saddle Points Faster than Gradient Descent. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 1042–1085. PMLR, 06–09 Jul 2018. URL https://proceedings.mlr.press/v75/jin18a.html. 199

[95] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac. Classical simulation of infinite-size quantum lattice systems in two spatial dimensions. *Phys. Rev. Lett.*, 101:250602, 2008. doi: 10.1103/PhysRevLett.101.250602. URL https://link.aps.org/doi/10.1103/PhysRevLett.101.250602. 204

[96] Michael J. Kastoryano and Fernando G. S. L. Brandao. Quantum Gibbs samplers: the commuting case, 2016. arXiv: 1409.3435 5, 6, 16, 22, 168, 261

[97] Michael J. Kastoryano and Kristan Temme. Quantum logarithmic Sobolev inequalities and rapid mixing. *Journal of Mathematical Physics*, 54(5):052202, 2013. 16, 32, 117

[98] Julia Kempe, Alexei Kitaev, and Oded Regev. The Complexity of the Local Hamiltonian Problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006. 197

[99] Scott Kirkpatrick and David Sherrington. Infinite-ranged models of spin-glasses. *Phys. Rev. B*, 17(11):4384, 1978. doi: 10.1103/PhysRevB.17.4384. 197

[100] A. Yu Kitaev, A. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, 2002. 197, 201, 208, 269

[101] Alexei Yu Kitaev, Alexander Shen, Mikhail N Vyalyi, and Mikhail N Vyalyi. *Classical and quantum computation*. Graduate studies in mathematics. American Mathematical Society, 2002. ISBN 9780821832295. 1, 9

[102] Emanuel Knill, Gerardo Ortiz, and Rolando D Somma. Optimal quantum measurements of expectation values of observables. *Physical Review A—Atomic, Molecular, and Optical Physics*, 75(1):012328, 2007. 28

[103] W. Kohn. Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.*, 71:1253–1266, 1999. doi: 10.1103/RevModPhys.71.1253. URL https://link.aps.org/doi/10.1103/RevModPhys.71.1253. 197

[104] Tomotaka Kuwahara, Álvaro M Alhambra, and Anurag Anshu. Improved thermal area law and quasilinear time algorithm for quantum gibbs states. *Physical Review X*, 11(1):011047, 2021. 187

[105] Thomas William Körner. *Fourier Analysis*. Cambridge University Press, 1988. doi: 10.1017/CBO9781107049949. 74

[106] Zeph Landau, Umesh Vazirani, and Thomas Vidick. A polynomial time algorithm for the ground state of one-dimensional gapped local hamiltonians. *Nature Physics*, 11(7):566–569, 2015. 60, 204

[107] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush. Even more efficient quantum computations of chemistry through tensor hypercontraction. *PRX Quantum*, 2 (3), Jul 2021. ISSN 2691-3399. doi: 10.1103/prxquantum.2.030305. URL http://dx.doi.org/10.1103/PRXQuantum.2.030305. 1, 9, 138, 168

[108] Seunghoon Lee, Joonho Lee, Huanchen Zhai, Yu Tong, Alexander M Dalzell, Ashutosh Kumar, Phillip Helms, Johnnie Gray, Zhi-Hao Cui, Wenyuan Liu, Michael Kastoryano, Ryan Babbush, John Preskill, David R. Reichman, Earl T. Campbell, Edward F. Valeev, Lin Lin, and Garnet Kin-Lic Chan. Is there evidence for exponential quantum advantage in quantum chemistry?, 2022. arXiv: 2208.02199 1, 9, 138, 197

[109] Seunghoon Lee, Joonho Lee, Huanchen Zhai, Yu Tong, Alexander M Dalzell, Ashutosh Kumar, Phillip Helms, Johnnie Gray, Zhi-Hao Cui, Wenyuan Liu, et al. Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry. *Nature Communications*, 14(1):1952, 2023. doi: 10.1038/s41467-023-37587-6. 203

[110] David Asher Levin, Yuval Peres, Elizabeth L. Wilmer, James Propp, and David B. Wilson. *Markov chains and mixing times*. American Mathematical Society, 2017. 3, 6, 10, 11, 31, 138

[111] Laura Lewis, Hsin-Yuan Huang, Viet T. Tran, Sebastian Lehner, Richard Kueng, and John Preskill. Improved machine learning algorithm for predicting ground state properties. *Nature Communications*, 15(1), 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-45014-7. URL http://dx.doi.org/10.1038/s41467-024-45014-7. 197

[112] Xiantao Li and Chunhao Wang. Simulating Markovian open quantum systems using higher-order series expansion. arXiv: 2212.02051, 2022. 24, 38, 127, 159

[113] Xiantao Li and Chunhao Wang. Succinct description and efficient simulation of non-Markovian open quantum systems. *Communications in Mathematical Physics*, 2023. doi: 10.1007/s00220-023-04638-4. arXiv: 2111.03240 38

[114] Lin Lin and Yu Tong. Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers. *PRX Quantum*, 3(1):010318, 2022. 9, 197

[115] Goran Lindblad. On the generators of quantum dynamical semigroups. *Commun. Math. Phys.*, 48:119–130, 1976. 197, 199

[116] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996. doi: 10.1126/science.273.5278.1073. 9, 26

[117] Guang Hao Low and Isaac L. Chuang. Optimal Hamiltonian simulation by quantum signal processing. *Physical Review Letters*, 118(1):010501, 2017. doi: 10.1103/PhysRevLett.118.010501. arXiv: 1606.02685 26, 140

[118] Guang Hao Low and Isaac L. Chuang. Hamiltonian simulation by uniform spectral amplification. arXiv: 1707.05391, 2017. 131

[119] Sirui Lu, Mari Carmen Bañuls, and J. Ignacio Cirac. Algorithms for quantum simulation at finite energies. *PRX Quantum*, 2(2):020321, 2021. doi: 10.1103/PRXQuantum.2.020321. arXiv: 2006.03032 19

[120] Di Luo, Zhuo Chen, Kaiwen Hu, Zhizhen Zhao, Vera Mikyoung Hur, and Bryan K. Clark. Gauge-invariant and anyonic-symmetric autoregressive neural network for quantum lattice models. *Phys. Rev. Res.*, 5:013216, Mar 2023. doi: 10.1103/PhysRevResearch.5.013216. URL https://link.aps.org/doi/10.1103/PhysRevResearch.5.013216. 204

[121] Christian Majenz, Tameem Albash, Heinz-Peter Breuer, and Daniel A. Lidar. Coarse graining can beat the rotating-wave approximation in quantum Markovian master equations. *Phys. Rev. A*, 88:012103, Jul 2013. doi: 10.1103/PhysRevA.88.012103. URL https://link.aps.org/doi/10.1103/PhysRevA.88.012103. 3, 15, 16, 60, 101

[122] Juan Maldacena and Leonard Susskind. Cool horizons for entangled black holes. *Fortschritte der Physik*, 61, 2013. 28, 140

[123] Chris Marriott and John Watrous. Quantum Arthur–Merlin games. *Computational Complexity*, 14(2):122–152, 2005. doi: 10.1007/s00037-005-0194-x. arXiv: cs/0506068 17, 19

[124] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics*, pages 93–191. Springer, 1999. doi: 10.1007/978-3-540-48115-7_2. 5, 6, 10, 60

[125] John Martyn and Brian Swingle. Product spectrum ansatz and the simplicity of thermal states. *Physical Review A*, 100(3):032107, 2019. doi: 10.1103/PhysRevA.100.032107. arXiv: 1812.01015 19

[126] John M Martyn, Zane M Rossi, Andrew K Tan, and Isaac L Chuang. Grand unification of quantum algorithms. *PRX Quantum*, 2(4):040203, 2021. 247, 249

[127] Sam McArdle, András Gilyén, and Mario Berta. Quantum state preparation without coherent arithmetic, 2022. arXiv: 2210.14892 27, 49, 72

[128] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9, 2018. doi: 10.1038/s41467-018-07090-4. arXiv: 1803.11173 9

[129] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):1–6, 2018. doi: 10.1038/s41467-018-07090-4. 210

[130] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: 10.1063/1.1699114. 3

[131] Stewart Morawetz, Isaac J. S. De Vlugt, Juan Carrasquilla, and Roger G. Melko. U(1)-symmetric recurrent neural networks for quantum state reconstruction. *Phys. Rev. A*, 104:012401, 2021. doi: 10.1103/PhysRevA.104.012401. URL https://link.aps.org/doi/10.1103/PhysRevA.104.012401. 204

[132] Mario Motta, Chong Sun, Adrian TK Tan, Matthew J OâRourke, Erika Ye, Austin J Minnich, Fernando GSL Brandão, and Garnet Kin-Lic Chan. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nat. Phys.*, 16(2):205–210, 2020. 137

[133] Jonathan Edward Moussa. Low-depth quantum Metropolis algorithm. arXiv: 1903.01451, 2019. 137, 142, 143, 169, 329

[134] Evgeny Mozgunov and Daniel Lidar. Completely positive master equation for arbitrary driving and small level spacing. *Quantum*, 4:227, February 2020. ISSN 2521-327X. doi: 10.22331/q-2020-02-06-227. URL https://doi.org/10.22331/q-2020-02-06-227. 3, 13, 15, 16, 18, 21, 60, 101, 103, 138

[135] Evgeny Mozgunov and Daniel Lidar. Completely positive master equation for arbitrary driving and small level spacing. *Quantum*, 4:227, February 2020. ISSN 2521-327X. doi: 10.22331/q-2020-02-06-227. URL https://doi.org/10.22331/q-2020-02-06-227. 197, 199, 213, 242, 243

[136] John A Mydosh. *Spin glasses: an experimental introduction*. CRC Press, 1993. 197

[137] Michael Müger. Lebesgue's characterization of Riemann integrable functions, 2006. URL https://www.math.ru.nl/~mueger/Lebesgue.pdf. 95

[138] Daniel Nagaj, Pawel Wocjan, and Yong Zhang. Fast amplification of QMA. *Quantum Information and Computation*, 9(11&12):1053–1068, 2009. doi: 10.26421/QIC9.11-12. arXiv: 0904.1549 16

[139] Frederik Nathan and Mark S Rudner. Universal lindblad equation for open quantum systems. *Physical Review B*, 102(11):115109, 2020. 101

[140] Michael A Nielsen, Mark R Dowling, Mile Gu, and Andrew C Doherty. Quantum computation as geometry. *Science*, 311(5764):1133–1135, 2006. doi: 10.1126/science.1121541. 202, 222

[141] Yusuke Nomura, Andrew S. Darmawan, Youhei Yamaji, and Masatoshi Imada. Restricted boltzmann machine learning for solving strongly correlated quantum systems. *Phys. Rev. B*, 96:205152, 2017. doi: 10.1103/PhysRevB.96.205152. URL https://link.aps.org/doi/10.1103/PhysRevB.96.205152. 204

[142] Roberto Oliveira and Barbara M. Terhal. The complexity of quantum spin systems on a two-dimensional square lattice. *Quantum Inf. Comput.*, 8:900–924, 2008. 201, 208, 237, 267, 269

[143] Peter JJ O'Malley, Ryan Babbush, Ian D Kivlichan, Jonathan Romero, Jarrod R McClean, Rami Barends, Julian Kelly, Pedram Roushan, Andrew Tranter, Nan Ding, et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X*, 6:031007, Jul 2016. doi: 10.1103/PhysRevX.6.031007. URL https://link.aps.org/doi/10.1103/PhysRevX.6.031007. 199, 202

[144] Panos M Pardalos and Stephen A Vavasis. Quadratic programming with one negative eigenvalue is np-hard. *Journal of Global optimization*, 1(1):15–22, 1991. 199, 214

[145] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix product state representations. *Quantum Info. Comput.*, 7(5):401–430, 2007. ISSN 1533-7146. 1, 197, 204

[146] ProofWiki. Laplace transform of error function. Accessed on May 1, 2023. URL https://proofwiki.org/wiki/Laplace_Transform_of_Error_Function. 179

[147] Zhuoran Qiao, Matthew Welborn, Animashree Anandkumar, Frederick R. Manby, and Thomas F. Miller. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *The Journal of Chemical Physics*, 153(12), 2020. ISSN 1089-7690. doi: 10.1063/5.0021955. URL http://dx.doi.org/10.1063/5.0021955. 197

[148] Patrick Rall. Faster coherent quantum algorithms for phase, energy, and amplitude estimation, 2021. arXiv: 2103.09717 15

[149] Patrick Rall, Chunhao Wang, and Pawel Wocjan. Thermal state preparation via rounding promises. *Quantum*, 7:1132, 2023. 4, 12, 15, 18, 19, 38, 43, 57, 137, 138, 139, 143, 201, 204

[150] Alfred Guillou Redfield. The theory of relaxation processes. In John S. Waugh, editor, *Advances in Magnetic Resonance*, volume 1 of *Advances in Magnetic and Optical Resonance*, pages 1–32. Academic Press, 1965. doi: https://doi.org/10.1016/B978-1-4832-3114-3.50007-6. 3, 15, 101

[151] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56:34:1–34:40, 2009. 73

[152] Ángel Rivas and Susana F. Huelga. *Open Quantum Systems*. Springer, 2012. doi: 10.1007/978-3-642-23354-8. arXiv: 1104.5242 9, 13, 18, 101

[153] Cambyse Rouzé, Daniel Stilck França, and Álvaro M. Alhambra. Efficient thermalization and universal quantum computing with quantum Gibbs samplers. arXiv: 2403.12691, 2024. 6

[154] Walter Rudin. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 3rd edition, 1976. ISBN 9780070856134. URL https://books.google.hu/books?id=kwqzPAAACAAJ. 95

[155] Anders W. Sandvik. Stochastic series expansion method with operator-loop update. *Phys. Rev. B*, 59:R14157–R14160, 1999. doi: 10.1103/PhysRevB.59.R14157. URL https://link.aps.org/doi/10.1103/PhysRevB.59.R14157. 197

[156] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, 2011. ISSN 0003-4916. doi: 10.1016/j.aop.2010.09.012. URL http://dx.doi.org/10.1016/j.aop.2010.09.012. 197, 204

[157] Alexander Schuckert, Annabelle Bohrdt, Eleanor Crane, and Michael Jason Knap. Probing finite-temperature observables in quantum simulators with short-time dynamics, 2022. arXiv: 2206.01756 19

[158] Troy J. Sewell, Christopher David White, and Brian Swingle. Thermal multi-scale entanglement renormalization ansatz for variational Gibbs state preparation. arXiv: 2210.16419, 2022. 19

[159] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997. doi: 10.1137/S0097539795293172. Earlier version in FOCS'94. arXiv: quant-ph/9508027 2

[160] Oles Shtanko and Ramis Movassagh. Algorithms for Gibbs state preparation on noiseless and noisy random quantum circuits, 2021. URL https://arxiv.org/abs/2112.14688. 4, 18, 22, 26, 137, 138, 329

[161] Rolando Somma, Sergio Boixo, and Howard Barnum. Quantum simulated annealing. *arXiv preprint arXiv:0712.1008*, 2007. 18

[162] E.M. Stoudenmire and Steven R. White. Studying two-dimensional systems with the density matrix renormalization group. *Annu. Rev. Condens.*, 3(1):111–128, 2012. doi: 10.1146/annurev-conmatphys-020911-125018. URL https://doi.org/10.1146/annurev-conmatphys-020911-125018. 204

[163] Aaron Szasz, Johannes Motruk, Michael P. Zaletel, and Joel E. Moore. Chiral spin liquid phase of the triangular lattice hubbard model: A density matrix renormalization group study. *Phys. Rev. X*, 10:021042, 2020. doi: 10.1103/PhysRevX.10.021042. URL https://link.aps.org/doi/10.1103/PhysRevX.10.021042. 204

[164] Mário Szegedy. Quantum speed-up of Markov chain based algorithms. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 32–41, 2004. doi: 10.1109/FOCS.2004.53. arXiv: quant-ph/0401053 5, 18, 28, 53, 140

[165] Ewin Tang. Quantum principal component analysis only achieves an exponential speedup because of its state preparation assumptions. *Physical Review Letters*, 127(6):060503, 2021. doi: 10.1103/PhysRevLett.127.060503. arXiv: 1811.00414 2

[166] Kristan Temme, Michael James Kastoryano, Mary Beth Ruskai, Michael Marc Wolf, and Frank Verstraete. The χ^2-divergence and mixing times of quantum Markov processes. *Journal of Mathematical Physics*, 51(12):122201, 2010. doi: 10.1063/1.3511335. arXiv: 1005.2358 32, 44

[167] Kristan Temme, Tobias J. Osborne, Karl G. Vollbrecht, David Poulin, and Frank Verstraete. Quantum Metropolis sampling. *Nature*, 471(7336):87–90, 2011. doi: 10.1038/nature09770. arXiv: 0911.3635 4, 10, 12, 17, 18, 19, 31, 57, 134, 135, 137, 138, 139, 143, 201, 204

[168] Barbara M Terhal and David P DiVincenzo. Problem of equilibration and the computation of correlation functions on a quantum computer. *Physical Review A*, 61(2):022301, 2000. 3, 4, 9, 12, 18

[169] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H. Booth, and Jonathan Tennyson. The variational quantum eigensolver: A review of methods and best practices. *Physics Reports*, 2021. 9

[170] Anton Trushechkin. Unified Gorini-Kossakowski-Lindblad-Sudarshan quantum master equation beyond the secular approximation. *Physical Review A*, 103(6), Jun 2021. ISSN 2469-9934. doi: 10.1103/physreva.103.062226. URL http://dx.doi.org/10.1103/PhysRevA.103.062226. 101

[171] Karel Van Acoleyen, Michaël Mariën, and Frank Verstraete. Entanglement rates and area laws. *Physical Review Letters*, 111(17):170501, 2013. doi: 10.1103/PhysRevLett.111.170501. arXiv: 1304.5931 186, 187

[172] Laurens Vanderstraeten, Jutho Haegeman, Philippe Corboz, and Frank Verstraete. Gradient methods for variational optimization of projected entangled-pair states. *Phys. Rev. B*, 94:155123, 2016. doi: 10.1103/PhysRevB.94.155123. URL https://link.aps.org/doi/10.1103/PhysRevB.94.155123. 204

[173] F. Verstraete, V. Murg, and J.I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Adv. Phys.*, 57(2):143–224, 2008. doi: 10.1080/14789940801912366. URL https://doi.org/10.1080/14789940801912366. 197, 204

[174] Frank Verstraete, Michael M Wolf, and J Ignacio Cirac. Quantum computation and quantum-state engineering driven by dissipation. *Nature Physics*, 5(9): 633–636, 2009. doi: 10.1038/nphys1342. 203

[175] Tom Vieijra, Corneel Casert, Jannes Nys, Wesley De Neve, Jutho Haegeman, Jan Ryckebusch, and Frank Verstraete. Restricted boltzmann machines for quantum states with non-abelian or anyonic symmetries. *Phys. Rev. Lett.*, 124:097201, 2020. doi: 10.1103/PhysRevLett.124.097201. URL https://link.aps.org/doi/10.1103/PhysRevLett.124.097201. 204

[176] Chunhao Wang. Private communication via e-mail., March, 2023. 38, 127

[177] Steven R. White. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.*, 69:2863–2866, 1992. doi: 10.1103/PhysRevLett.69.2863. URL https://link.aps.org/doi/10.1103/PhysRevLett.69.2863. 1, 197, 204

[178] Steven R. White. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B*, 48:10345–10356, 1993. doi: 10.1103/PhysRevB.48.10345. URL https://link.aps.org/doi/10.1103/PhysRevB.48.10345. 197, 204

[179] Pawel Wocjan and Anura Abeyesinghe. Speedup via quantum sampling. *Physical Review A—Atomic, Molecular, and Optical Physics*, 78(4):042336, 2008. 46, 131

[180] Pawel Wocjan and Kristan Temme. Szegedy walk unitaries for quantum maps, 2021. arXiv: 2107.07365 10, 12, 15, 18, 19, 28, 50, 57, 58, 67, 137, 138, 139, 143, 146, 148, 329

[181] Michael M. Wolf. Quantum channels & operations: Guided tour. 2012. URL https://mediatum.ub.tum.de/download/1701036/1701036.pdf. 12, 114, 116, 118

[182] Ronald de Wolf. Quantum computing: Lecture notes, 2019. arXiv: 1907.09415 30

[183] Han-Qing Wu, Shou-Shu Gong, and D. N. Sheng. Randomness-induced spin-liquid-like phase in the spin-$\frac{1}{2}$ $J_1 - J_2$ triangular heisenberg model. *Phys. Rev. B*, 99:085141, 2019. doi: 10.1103/PhysRevB.99.085141. URL https://link.aps.org/doi/10.1103/PhysRevB.99.085141. 204

[184] Jingxiang Wu and Timothy H Hsieh. Variational thermal quantum simulation via thermofield double states. *Physical review letters*, 123(22):220502, 2019. 19

[185] Theodore J. Yoder, Guang Hao Low, and Isaac L. Chuang. Fixed-point quantum search with an optimal number of queries. *Physical Review Letters*, 113(21): 210501, 2014. doi: 10.1103/PhysRevLett.113.210501. arXiv: 1409.3305 132

[186] Man-Hong Yung and Alán Aspuru-Guzik. A quantum-quantum Metropolis algorithm. *Proceedings of the National Academy of Sciences*, 109(3):754–759, 2012. doi: 10.1073/pnas.1111758109. arXiv: 1011.1468 12, 18, 46, 131, 137, 138, 140, 143, 329

[187] Michael P. Zaletel and Frank Pollmann. Isometric tensor network states in two dimensions. *Phys. Rev. Lett.*, 124:037201, 2020. doi: 10.1103/PhysRevLett.124.037201. URL https://link.aps.org/doi/10.1103/PhysRevLett.124.037201. 204

[188] Daiwei Zhu, Sonika Johri, Norbert M Linke, KA Landsman, C Huerta Alderete, Nhunh H Nguyen, AY Matsuura, TH Hsieh, and Christopher Monroe. Generation of thermofield double states and critical ground states with a quantum computer. *Proceedings of the National Academy of Sciences*, 117(41):25402–25406, 2020. 19

[189] Vladimir Antonovich Zorich. *Mathematical Analysis II*. Universitext. Springer, 2nd edition, 2016. doi: 10.1007/978-3-662-48993-2. 179

[190] Antoni Zygmund. *Trigonometric Series*. Cambridge Mathematical Library. Cambridge University Press, 3rd edition, 2003. doi: 10.1017/CBO9781316036587. 73

[1]Here $\tilde{t}_{mix}$ refers to the mixing time of modified Lindbladians that forbid certain energies transitions (parameterized by an additional attenuation coefficient γ_{att}); it is unclear how this restricted connectivity impacts the mixing time. Also, our improved Lindbladian simulation results already improve their complexities from $\tilde{O}(\frac{\beta^3\tilde{t}_{mix}}{\epsilon^7})$ to $\tilde{O}(\frac{\beta\tilde{t}_{mix}}{\epsilon^2})$.

[2]Ref. [160] is similar to [38], but its proof of convergence assumes the Eigenstate Thermalization Hypothesis (ETH) and a maximally mixed initial state. [38] has guarantees assuming only the mixing time, and ETH is one way of bounding the mixing time.

[3]Ref. [180] did not include the algorithmic cost of quantum simulated annealing, so we fill in using our modernized version (section 2.12). Likewise, the simulated annealing cost of [186] could also be improved (still assuming perfect QPE).

[4]The proof of correctness assumes each eigenstate $\boldsymbol{H}$ can be perfectly distinguished.

[5]Assuming the Markov property and clustering quantities both decay exponentially.

[6]Their exactly detailed-balanced quantum channel seems to qualitatively differ from other quantum MCMC algorithms. It assumes certain efficiently implementable basis measurements with a *good overlap* with the energy basis. For example if measured in the computational basis, the 1D transverse field Ising model seems to exhibit an *exponential* mixing time at constant temperature, see [133, Page 5].

[7]A comparison of existing thermal state preparation algorithms. We focus on methods with provable guarantees for an ϵ-approximation of the Gibbs state (in trace distance) and list their cost, assumptions, and caveats. We use $\mathrm{Poly}(\cdot)$ to denote polynomials and $\tilde{O}(\cdot)$ to absorb logarithmic dependences. The first few algorithms are Monte Carlo-style methods, incoherent or coherent; we represent their costs by the total black-box Hamiltonian simulation time. The incoherent ones are based on semi-groups, with complexity being the cost of emulating the semi-group (which is basically dominated by the phase estimation time t_{QPE}) multiplied by the mixing time t_{mix}. Instead, the coherent version implements block-encoding for discriminants and prepares the purified Gibbs state via quantum simulated annealing. The number of discriminant calls is $\tilde{O}(\beta\|\boldsymbol{H}\|\sqrt{\lambda_{gap}^{-1}})$ (as the counterpart for the mixing time t_{mix}) where λ_{gap} refers to the minimum gap of discriminants along the adiabatic path. We calculate the costs for Theorem 2.1.3 and Theorem 2.1.4, assuming the algorithmic parameters and the mixing time t_{mix} or spectral gap λ_{gap} satisfy certain self-consistency constraint. The mixing time and spectral gap can be (loosely) converted to each other, assuming approximate detailed balance. Of course, as in all classical MCMC methods, the mixing time t_{mix} or spectral gap λ_{gap} can be exponentially small depending on the temperature and the system; indeed, there are systems whose thermal states are expected to have high complexity, such as certain spin glass models, and we do not expect efficient quantum algorithms to exist in general. Optimistically, we often care about the Gibbs state that appears in nature, which suggests the mixing time remains reasonably small.