The power of one qubit in quantum simulation algorithms
Polla, S.

The power of one qubit in quantum simulation algorithms

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl
volgens besluit van het college voor promoties
te verdedigen op donderdag 22 februari 2024
klokke 15:00 uur

door

Stefano Polla

geboren te Bergamo, Italië
in 1994

On the cover: *an artistic representation of quantum computing emerging as a new technology, inspired by the rising cities painted by the futurist movement.* — by Jeanne M. Viet [Miss J Art]

To those who brightened my hardest days

Contents

Contents

Introduction

1.1 Preface

Quantum theory, the mathematical framework that governs the behavior of elementary physical systems, stands as a cornerstone of modern science. While its first principles allow modeling complex systems like molecules and materials, managing the complexity and predicting emergent behaviors demands additional tools [1]. Analytical approximations played a key role in early successes of nuclear, atomic, and solid-state physics; together with computational simulation techniques they now drive the fields of quantum chemistry and material science. Since the 1980s *quantum computers* have been proposed as an additional tool, promising access to simulations impossible for classical methods [2]. Recent years saw the realization of the first programmable *quantum devices* that can outperform classical computers in benchmark tasks [3–7], but lack practical applications yet. The next quantum computing milestone is demonstrating the ability to solve genuinely useful problems beyond the capabilities of other tools. Achieving this requires hardware enhancements, but also research on pertinent target problems, and the development of quantum algorithms that exploit the hardware to solve the problem. The study of complex quantum systems offer a variety of challenging problems, supported by heuristic insight from physics: a valuable resource in the effective development of

1

quantum algorithms.

This thesis collects the proposals of quantum algorithms developed in this context, tackling some known hurdles in simulating complex systems with near-term quantum devices. The studied techniques focus on questions like the preparation of ground states of natural systems, the extraction of information about observables from quantum states, and and the resilience of quantum simulations to noise. A red line connecting the proposed algorithms is the focus on the support which a single qubit – the smallest unit of quantum information – can provide in implementing a variety of tasks in quantum simulation.

The remainder of this chapter is structured as follows: Section 1.2 provides an overview of key concepts in quantum information processing, discussing the advantages and limitations of quantum devices. Section 1.3 introduces quantum simulation by examining the computational problems naturally defined by a Hamiltonian description of a quantum system. Section 1.4 offers a high-level description of some significant quantum simulation algorithms relevant to this thesis. Section 1.5 delves deeper into a specific quantum simulation target, namely, the study of molecular systems. Finally, in Section 1.6, we present an overview of the chapters that constitute the main body of this work.

1.2 Processing quantum information

A *state* is a specific configuration that a physical system can assume at some point in time. In classical physics, states are described by variables taking defined values, such as the position $\boldsymbol{x}$ and velocity $\dot{\boldsymbol{x}}$ of a cannonball in Newtonian mechanics, or the ON/OFF state of a transistor in digital electronics. Conversely, the state of a quantum system can consist of a *superposition* of multiple, distinguishable states. In Dirac notation [8]

$$|\psi\rangle = \alpha_A |A\rangle + \alpha_B |B\rangle + \dots, \quad |\alpha_A|^2 + |\alpha_B|^2 + \dots = 1, \qquad (1.1)$$

where $\{A, B, \dots\}$ are a set of states that can be distinguished deterministically by a sufficiently precise measurement, each representing a different set of values of the classical variables that characterize the system. The state of a quantum switch will then have the form $\alpha_{\text{ON}} |\text{ON}\rangle + \alpha_{\text{OFF}} |\text{OFF}\rangle$, while the description of a quantum particle will need one value of $\alpha_{\boldsymbol{x}}$ for each possible position $\boldsymbol{x}$.

Consider a system which can assume N distinguishable states. A deterministic description of a classical state of such a system just needs to

identify one out of the N possible states; this information can be encoded in $\log_2(N)$ bits of memory. A full description of a quantum state for the same system will require $N - 1$ complex amplitudes, each stored to precision $\epsilon_\mathbb{C}$ in $m = 2\log_2(\epsilon_\mathbb{C}^{-1})$ bits of classical memory. For a composite system, made up of n elements, the number of total distinguishable states scales exponentially, i.e. $N = \mathcal{O}(2^n)$. A classical state of such system can be described in $\mathcal{O}(n)$ bits, but a full representation of the quantum amplitudes requires $\mathcal{O}(m2^n)$ bits of classical memory. This exponential scaling quickly makes unfeasible to fully represent quantum states of complex systems.

In contrast to a classical computer, a quantum device can natively store and process quantum states of the form Eq. (1.1). Typical quantum devices are constructed by assembling a number of *two-level systems* – elements with two distinguishable states labeled $|0\rangle$ and $|1\rangle$. Combining n two-level systems allows to store a quantum superposition of 2^n distinguishable states, or n qubits (a quantum analogue for bits) of quantum information. Each two-level system stores a qubit of quantum information; each additional qubit doubles the dimension of the Hilbert space of states available to the quantum device.

1.2.1 Classical input-output

While quantum theory introduces one significant extension to the concept of information by allowing for superposition storage, this comes with a notable limitation. Contrary to a classical state, a single copy of an arbitrary quantum state cannot be fully characterized by observation. The conversion of quantum information to classical information implies a loss, typically associated with the randomness of a measurement's outcome. Holevo's theorem [9] quantifies this, stating the amount of classical information extracted from a n-qubit quantum state is bounded by n bits. As a corollary, it is impossible to copy an arbitrary state of a quantum device (*no-cloning*), unlike for classical information which can always be observed and transcribed.

A quantum algorithm needs to ensure that, at the end of the computation, the quantum state of the device encodes the relevant information in a way that makes its extraction easy. The whole n qubit Hilbert space can be used to perform computation, but the output should be encoded in $O(n)$ bits of classical information. The final measurement performed on the state needs to be designed such that it can extract the necessary information from the state as efficiently as possible. In Chapter 4, we explore this optimization of measurements for the expectation value of observables,

under a further restriction on the amount of classical information that can be extracted from the device.

Similarly, the input to any quantum computation – the list of instructions – should be given in terms of classical information. A common framework to describe this list of instructions is the *quantum circuit model*. There, operations on the quantum memory are described as *gates* involving a small number of qubits at a time. Gates are typically unitary operations, although in some cases it is convenient to consider quantum channels to model noisy gates or non-unitary gates such as *reset* (a key element in the algorithm presented in Chapter 2). The computation is then realized by applying a sequence of gates (a circuit) to a fiducial initial state.

Another model of computation relevant to this thesis is that employed by *analog quantum simulators*. These quantum devices can natively implement the time evolution generated by a Hamiltonian $H(\boldsymbol{\theta})$, function of parameters $\boldsymbol{\theta}$ that can eventually be changed during the computation The Hamiltonians acts on the whole system at once, but it can be described in terms of classical information as explained in Section 1.3. In contrast to the circuit model, here the computation is described by stating the Hamiltonian throughout the computation time. Quantum simulators allowing for general enough Hamiltonian are universal [10]. In fact, gate-based quantum computers are practically realized as specialized simulators, where each gate is generated by a pre-calibrated time-dependent Hamiltonian. The term *analog quantum simulators* tends to be reserved for devices that can implement a limited set of Hamiltonians, modeling analogously the evolution of some quantum system of interest.

1.2.2 Noise, error correction and mitigation

All physical processes, including computation, suffer from a certain amount of noise due to unpredictable perturbations. Classical computers can deal with noise by storing information redundantly, both at the hardware level (e.g. using macroscopic bistable systems) and by backing up data. Redundancy-based error correction cannot be trivially applied in quantum computation, as quantum information cannot be copied (no-cloning) and compared throughout the computation. This leads to an accumulation of noise over the computation, leading to degradation of quantum information and, ultimately, a success probability that decays exponentially with the space-time volume of the computation. [11, 12]

Quantum error correction codes aim to protect quantum information by encoding it in a subspace of *logical states*, living in a much larger space of *physical states* of the device. The logical subspace is engineered to ensure

that physical noise processes scattering one logical state to the other are exceedingly rare. For example, in *topological codes* [13], information is encoded non-locally in terms of correlations between a large number of two-level systems (a.k.a. *physical qubits*). A perturbation can only convert one logical state into the other by acting on many two-level systems in a globally-correlated fashion. The probability of such process occurring due to natural noise is exponentially small in the number of physical qubits. Non-logical states can be led back to logical states by following the geodesic defined by error probability, effectively correcting the most-probable errors.

Quantum computation can be made completely *fault-tolerant* by ensuring information is stored in a large enough logical subspace throughout the computation (including initialization, gate operations and measurements). The implementation of a large scale fault tolerant quantum computer is the only known way towards many key applications of quantum computers, including breaking cryptosystems using Shor's algorithm [14] and performing challenging chemistry simulations of industrial importance [15]. However, such a computer requires an underlying physical device of very large scale (millions of qubits) and with sufficiently small physical error rates, far out of reach of today's technologies. Large amounts of resources are being invested towards the goal of fault-tolerant quantum computing, and the last two years saw the first small scale proof-of-concept demonstration of successful error correction [16] and fault tolerant operation of a quantum device [17]. While defining the timeline to full-scale fault tolerant quantum computing is not yet possible, even the most optimistic estimates accepted by the community measure in the order of decades.

Another approach, more adapted to today's noisy intermediate-scale quantum (NISQ, [18]) devices, consists in keeping computations short to prevent accumulated noise to reach a disruptive level. Short computations are obviously limited in capabilities, but have been demonstrated to outperform classical devices on benchmark tasks [3–7]. The NISQ paradigm can be summarized as *prepare, sample, repeat*, and it focuses on performing many short-time quantum computations (called *circuit runs* or *shots*), used as subroutines by a classical algorithm. This results in a hybrid quantum-classical computation, where the usefulness of the quantum subroutine is typically heuristic. An archetypal category of NISQ algorithms are variational quantum algorithms, such as the variational quantum eigensolver [19] described in Section 1.4.4 and considered in many chapters of this thesis.

1.3 Targets of quantum simulation

The term *quantum simulation*, in its most general sense, refers to the use of a quantum device to study a physical phenomenon through a representative model. Phenomena happening within an isolated quantum system can be modeled through the Hamiltonian picture. The Hamiltonian H is a Hermitian linear operator on the Hilbert space of states, which defines its *dynamics* (i.e. how its state changes in time) through the *time-dependent Schrödinger equation*

$$-i\frac{\partial}{\partial t}\left|\psi(t)\right\rangle = H\left|\psi(t)\right\rangle. \tag{1.2}$$

If the system is influenced by a classical environment, the Hamiltonian could depend on external parameters, possibly changing in time. In the case of more complex interactions with the environment, describing the dynamics of the system will require a more complex equation (such as the Liouville equation, or a Schrödinger equation for a larger model including the environment). Nevertheless, weak interactions with the environment can often be studied through linear response theory, with Hamiltonian eigenstates and dynamics playing a key role.

A system's Hamiltonian naturally defines a set of natural targets for quantum simulation: the synthesis of its dynamics (Section 1.3.1), the preparation of thermal states and eigenstates (Section 1.3.2), and the measurement of the system's energy (Section 1.3.3). In this section, we formalize these primitive target as problems with classical input-output relations (where one of the inputs is always the Hamiltonian). We define the size of each problem as the number qubits n defining the model's Hilbert space. We require all inputs and outputs to have an efficient classical description, i.e. be representable in $\mathcal{O}(\text{poly}(n))$ bits. When we require *quantum inputs* such as a state $\left|\psi\right\rangle$ or an observable O, we assume that they are given in terms of an efficient procedure that prepares $\left|\psi\right\rangle$ or samples O. All problems are formulated such that output is an estimate of a defined quantity, accurate to precision $\epsilon > 0$ with high probability. These problems are summarized in table 1.1.

A generic Hamiltonian cannot be described efficiently, as the definition of a Hermitian operator on an n-qubit Hilbert space $\mathcal{H}$ requires 2^{2n-1} complex parameters. Natural Hamiltonians aren't however fully generic, as they inherit locality from the underlying physics. In local systems, the Hamiltonian can be described as a sum of local interaction terms h_j, each of which only acts non-trivially on a small Hilbert subspace $\mathcal{H}_j$ and as the

Problem	Input	Target	Output
Dynamics	$\lvert\psi\rangle, O, t$	$U(t)$	$\tilde{O} : \lvert\tilde{O} - \langle\psi(t)\lvert O\rvert\psi(t)\rangle\rvert < \epsilon$
Gibbs	O, T	$\rho_T \propto e^{-\frac{H}{T}}$	$\tilde{O} : \lvert\tilde{O} - \mathrm{Tr}\{O\rho_T\}\rvert < \epsilon$
Eigenstate	O, j	$\lvert E_j\rangle$	$\tilde{O} : \lvert\tilde{O} - \langle E_j\lvert O\rvert E_j\rangle\rvert < \epsilon$
Expectation	$\lvert\psi\rangle$	Expval $\langle H\rangle$	$\tilde{E} : \lvert\tilde{E} - \langle\psi\lvert H\rvert\psi\rangle\rvert < \epsilon$
Eigenvalue sampling	$\lvert\psi\rangle$	Projective meas. of H	$\tilde{E} : \lvert\tilde{E} - E_j\rvert < \epsilon,$ with prob. $p_j = \lvert\langle E_j\lvert\psi\rangle\rvert^2$
Eigenvalue estimation	j	E_j	$\tilde{E} : \lvert\tilde{E} - E_j\rvert < \epsilon$

Table 1.1: Summary of some primitive problems in quantum simulation. The input of an efficient representation of H and of the required accuracy ϵ is implicit for every problem.

$\mathbb{1}$ on the complementary subspace $\mathcal{H} \setminus \mathcal{H}_j$[1]:

$$H = \sum_{j}^{J\sim\mathrm{poly}(n)} h_j; \quad h_j := [h_j]_{\mathcal{H}_j} \otimes \mathbb{1}_{\mathcal{H}\setminus\mathcal{H}_j}. \tag{1.3}$$

This description is efficient as long as the number of terms in the sum scales polynomially $J \sim \mathcal{O}(poly(n))$ and the dimension of each local Hilbert space is constant $\dim[\mathcal{H}_j] \sim \mathcal{O}(1)$.

The two most common models of locality are illustrated in Fig. 1.1. A *geometrically local* system is composed of subsystem arranged in space, for example in a lattice. Each element can only couple to its geometric neighbors. As the number of neighbors of each subsystem is bounded by a constant, the number of couplings we need to describe the Hamiltonian can only scale linearly in the size of the system. Sometimes it is convenient to think of subsystems as particles, which can be geometrically delocalized and can interact with all other particles; in these cases, interactions typically involve no more than k-particles at a time. This defines an a *k-local* model, the number of possible interactions scales as n^k (thus still polynomial). Note that geometric locality is more stringent than k-locality. In both

---

[1]This description is only rigorous for local commuting systems (qudits). The action of a local fermionic Hamiltonian on the Hilbert space is more complex, as fermionic operators acting on different locations need anticommute. However, the number of terms that can appear in a fermionic local Hamiltonian remains the same.
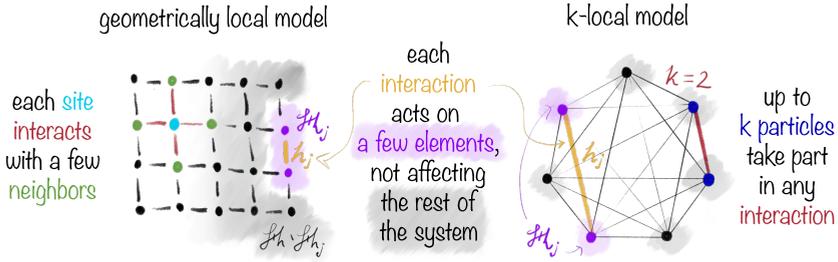
Figure 1.1: Visualization of the Hamiltonian Eq. (1.3) in a 2-dimensional lattice model (geometric locality) and a two-body interacting particle model (*k*-locality).

cases, the Hilbert space on which each interaction acts $\mathcal{H}_j$ is composed of up to a constant number k of subsystems, thus each h_j can be defined with $\mathcal{O}(2^k)$ parameters (constant in n). Other equivalently efficient descriptions of sparse Hamiltonians exist; we focus on local models as they mostly cover the natural systems of interest.

1.3.1 Hamiltonian dynamics

Reproducing the effect of time evolution of a state is the essence of simulation, in the stricter sense of the term. We refer to this problem to as *Hamiltonian simulation* or *dynamical simulation*. The inputs of the problem, along with H and ϵ, are the initial state $|\psi(0)\rangle \coloneqq |\psi\rangle$, the observable O, and the evolution time t. The goal is producing an estimate $\tilde{O}$ of the the expectation value of O on the state $|\psi(t)\rangle$ solution of the time-dependent Schrödinger equation (1.2), up to the required accuracy accuracy $|\tilde{O} - \langle\psi(t)|O|\psi(t)\rangle| < \epsilon$.

Solving Hamiltonian simulation boils down to synthesizing the propagator $U(t)$, the unitary operator that sends $|\psi\rangle$ to $|\psi(t)\rangle$. More generally, the propagator solves the operator-Schrödinger equation, $-i\partial_t U(t) = H(t)U(t)$. A quantum algorithm for Hamiltonian simulation needs to define a procedure to construct an approximation of $U(t)$, starting from the description of the Hamiltonian. In the case of a time-independent Hamiltonian the target operator is $U(t) = e^{-iHt}$, which is not a sparse operator (unlike its generator H).

An example use case of Hamiltonian dynamics is studying the evolution of a known equilibrium state after an external field quenches (i.e. suddenly

changes) the system's Hamiltonian. The synthesis of the evolution operator $U(t)$ is also an key subroutine in many other simulation algorithms, such as simulated thermalization (a version of which is studied in Chapter 2), eigenvalue sampling (Section 1.4.5), quantum-assisted Hamiltonian learning and many others. Furthermore, any efficient quantum computation can be encoded in a Hamiltonian simulation problem [2, 20] – in complexity theory terms, Hamiltonian simulation is BQP-complete.

The first proposal for a procedure to approximately synthesize $U(t)$ on a universal quantum computer was put forward by Lloyd in 1996 [21], and it is based on Suzuki-Trotter product formulas [22, 23]. Through this procedure Lloyd showed that a universal quantum computer can solve the Hamiltonian dynamics problem for any k-local Hamiltonian efficiently. The last years saw many proposals of improved algorithms for Hamiltonian simulation, most of which lead to a significant reduction of the asymptotic costs. Finally, analog quantum simulators can implement natively the dynamics generated by some restricted class of Hamiltonian. We expand on these in Section 1.4.2.

1.3.2 State preparation

Thermal state preparation

A physical system weakly interacting with an environment tends to thermalize, that is to reach an equilibrium state ensemble with a fixed temperature. Analogously to the classical case, a quantum thermal ensemble distribution mirrors the Gibbs measure, with the probability of a state decaying exponentially with the state's energy divided by the temperature. As the energy is described by the Hamiltonian H, the density matrix for the *Gibbs state* at temperature T is defined by

$$\rho_T := \frac{1}{Z(T)} e^{-\frac{H}{T}}, \tag{1.4}$$

where $Z(T) := \text{Tr}\left[e^{-H/T}\right]$ is the partition function which normalizes the density matrix. As most systems in Nature are found in a thermal equilibrium state, studying the properties of thermal states has intrinsic interest. Characterizing thermal state properties as a function of temperature and other external parameters can lead to the identification of transitions between phases of matter, whose study is of special importance to material science.

The inputs to the problem of thermal state preparation, along with H and ϵ, are the observable O and the temperature T. The goal is producing

an estimate $\tilde{O}$ of the the expectation value of O on the Gibbs state, up to the required accuracy accuracy $|\tilde{O} - \text{Tr}[\rho_T O]| < \epsilon$. To solve this problem, one can attempt to prepare an approximation of the Gibbs ensemble ρ_T on the quantum device.

Thermal state preparation algorithms include methods that simulate thermalization, by coupling the model system with a model bath. The algorithm we introduce in Chapter 2 fits in this class, although we specifically explore its application to ground state preparation. Other algorithms for thermal state simulation include a quantum version of metropolis sampling [24], variational approaches [25, 26], and filtering-based techniques [27].

Eigenstate preparation

The eigenstates of the Hamiltonian represent the equilibrium pure states of an isolated system, left invariant by time evolution. Each eigenstate $|E_j\rangle$ has a well-defined energy E_j, matching an eigenvalue of the Hamiltonian and satisfying the time-independent Schrödinger equation

$$H |E_j\rangle = E_j |E_j\rangle. \tag{1.5}$$

We assume an ordering of the eigenvalues $E_j < E_{j-1}$ without loss of generality.

The lowest-energy eigenstate denoted $|E_0\rangle$ or the *ground state*, holds particular significance, as it represent the zero-temperature limit of the thermal state Eq. (1.4). When the lowest excitation energy, also known as the "ground-state gap" of a system ($\Delta_{\text{GS}} = E_1 - E_0$), significantly exceeds the temperature of the environment T, thermal states closely resembling the ground state naturally emerge. For instance, this is the case for electrons within a molecule or semiconductor, where typical excitation energies are on the order of 1eV (equivalent to 10^4K, two orders of magnitude higher than room temperature). Simulating the electronic ground state in these systems allows for the prediction of molecular and material properties in their natural state. Moreover, low-lying excited eigenstates are relevant to the study of spectroscopic properties. Further insights into the quantum simulation of molecular ground states are provided in Section 1.5.1.

The inputs to the problem of eigenstate preparation, along with H and ϵ, are the observable O and the eigenvalue index j. The goal is to produce an estimate $\tilde{O}$ of the the expectation value of O on the j-th eigenstate, up to the required accuracy accuracy $|\tilde{O} - \langle E_j|O|E_j\rangle| < \epsilon$. Solving this problem requires the quantum simulation to aim at preparing an approximation of the eigenstate $|E_j\rangle$.

Many algorithms for eigenstate preparation have been proposed, each employing very diverse techniques and heuristics, due to the intrinsic difficulty of the problem. The variational quantum eigensolver (Sec. 1.4.4) and adiabatic state preparation (Sec. 1.4.3) are important heuristic techniques for eigenstate preparation, discussed in more detail in the next section. Projective variants of quantum phase estimation (Sec. 1.4.5) can also be used for heralded preparation of eigenstates. Other ground state preparation techniques include among others simulated thermalization (a variant of which is studied in Chapter 2), and various flavors of quantum imaginary time evolution [28, 29].

1.3.3 Energy measurements

The Hamiltonian H serves as an observable representing the system's energy. In the preceding sections, we assumed the presence of an observable O as input, along with an efficient method for its sampling. This section addresses the challenge of implementing procedures to measure the Hamiltonian observable, starting from its efficient classical description. There are two distinct tasks, both referred to as energy measurement: *expectation value estimation* and *eigenvalue sampling.*

Expectation value estimation

Expectation value estimation involves an input state $|\psi\rangle$, along with the description of H and the target precision ϵ. The objective is to provide an estimate $\tilde{E}$ of the expected energy of $|\psi\rangle$ to the required precision $|\tilde{E} - \langle\psi|H|\psi\rangle| < \epsilon$. Expectation value estimation is a key subroutine in the variational quantum eigensolver (discussed in Section 1.4.4). Additionally, the same technique can be applied to estimate the expectation value of observables other than the system Hamiltonian.

The most basic algorithm to estimate $\langle\psi|H|\psi\rangle$ for H defined in Eq. (1.3) relies on the linearity of the expectation value:

$$\langle\psi|H|\psi\rangle = \sum_{j}^{J} \langle\psi|h_j|\psi\rangle. \qquad (1.6)$$

Given that h_j acts non-trivially on a subsystem of constant size, we can implement a quantum operation to sample it in constant time. In Chapter 4 we perform a thorough study of expectation value estimation under a restricted model of quantum computation, with applications in error-mitigated NISQ computing and metrology.

Eigenvalue sampling

The task of implementing projective measurement of H (up to target accuracy ϵ) on an input state $|\psi\rangle$ is more complex. The aim of eigenvalue sampling algorithm is to produce an estimate $\tilde{E}$ of an eigenvalue E_j, sampled from the probability distribution described by the Born rule $p_j = |\langle E_j|\psi\rangle|^2$, up to the required accuracy $|\tilde{E} - E_j| < \epsilon$. Note that this task entails two objectives: sampling from the probability distribution p_j and estimating the eigenvalue E_j, which is not known in advance based on the description (1.3) of the Hamiltonian. The entanglement-based quantum phase estimation algorithm, utilizing Hamiltonian dynamics simulation as a subroutine, can be used to implement this measurement efficiently (see Section 1.4.5).

1.3.4 The ground state energy problem

The eigenvalue estimation problem revolves around the estimation of a specific eigenvalue E_j of H, with j being the sole input to the problem in addition to H and ϵ. Frequently, this problem is specifically addresses the ground state, where it is known as the *ground state energy problem* or sometimes simply the *ground state problem*. Solving the ground state energy problem involves a combination of ground state preparation and energy measurement algorithms. A good approximation of the ground state energy can be achieved either by measuring $\langle H \rangle$ on a well-approximated $|E_0\rangle$, or by repeatedly sampling eigenvalues of H from an input state that demonstrates a substantial fidelity $|\langle E_0|\psi\rangle|^2$ with the ground state.

The ground state energy itself plays a pivotal role in the simulation of chemistry and material science, as exemplified in Section 1.5.1. Furthermore, any quantum computation that can be efficiently verified can be mapped into a ground state problem using Kitaev's history-state Hamiltonian construction [13, 30]; in complexity theory terms, the ground state problem is QMA-complete. The history state construction has been iteratively improved upon in literature, allowing for the encoding of any QMA problem into Hamiltonians that describe systems as straightforward as one-dimensional chains of nearest-neighbor-interacting subsystems, each with Hilbert subspace of constant size $N = 12$ [31]. Given that the quantum phase estimation algorithm implements eigenvalue sampling efficiently, the task of preparing a state $|\psi\rangle$ with a good fidelity to the ground state [i.e. $|\langle\psi|E_0\rangle|^2 \gtrsim 1/\text{poly}(n)$] must also be QMA-hard. Nevertheless, the statement on the complexity of ground state preparation is only true for worst-case. History state Hamiltonians have very specific forms, signifi-

1

cantly different from most natural system's Hamiltonians; for the latter, heuristic ground state preparation are more likely to succeed.

1.4 Algorithms for quantum simulation

In this section, we introduce some of several prominent algorithms designed to address the quantum simulation problems presented in the previous section. Since a primary objective of quantum algorithm research is to identify problems and methodologies that enable quantum devices to exhibit an advantage over classical numerical approaches, we will start by providing a concise conceptual introduction to prominent classical techniques used to solve quantum simulation problems. Subsequently, we will provide an overview of three fundamental quantum algorithms (or, more precisely, classes thereof), each of particular significance to the context of this thesis.

1.4.1 Classical algorithms

All the problems discussed in the preceding section can be framed in terms of classical input-output relationships, making them amenable to both classical and quantum algorithms. However, all the problems we described implicitly involve the storage and manipulation of states of the simulated quantum system. While quantum algorithms inherently handle this task, classical simulation algorithms must work with highly compressed state representations to maintain memory and time efficiency. Any compression of the quantum state introduces approximations, and tends to be accurate and efficient only in specific cases. Here, we highlight some successful classical algorithms used to address the ground state problem, emphasizing their primary assumptions.

Mean-field methods represent the state as a tensor product of linearly-many local subsystem states. By neglecting entanglement between different parts of the system, these methods offer an efficient yet rough representation suitable for approximating ground states in weakly-interacting systems. Even in such cases, mean-field methods often require to be aided by perturbation theory to achieve quantitative results. An example of a popular mean-field method is the **Hartree-Fock self-consistent field method**, frequently employed for Fermionic particle systems like molecular and material models. It is often used as a starting point for other methods, including ground state quantum simulations (see Section 1.5.1 for an example).

Similarly, **tensor-network methods** address ground state problems for local Hamiltonians by constraining the level of entanglement between local subsystems. This is accomplished by compressing the 2^n amplitudes of the state vector into a set of small tensors, along with a set of rules for contracting their indices structured to replicate the locality of the system. Tensor network techniques are particularly well-suited to treating geometrically local systems (see Fig. 1.1) when entanglement between distant parts of the system is small. One-dimensional tensor networks, also known as matrix product states, excel in describing ground states of gapped one-dimensional systems. Extending tensor network methods to higher-dimensional systems is more challenging due to the computational complexity of arbitrary index contractions. However, heuristic tensor network methods continue to improve, providing some of the most accurate results for correlated condensed matter systems [32] and chemistry applications [33, 34].

In the realm of Fermionic models of natural systems, **density functional theory (DFT)** offers a set of highly efficient algorithms for studying ground states. DFT achieves compression by storing only the local density of particles, a classical variable whose information content scales linearly with the system's size. The Hohenberg-Kohn theorems establish that such a description suffices to fully characterize a ground state for a two-body Fermionic Hamiltonian [35, 36]. Nevertheless, the computational complexity of exactly retrieving the energy from this compressed representation remains exponential. To simplify this computation, various mathematically and/or empirically motivated approximations have been developed, resulting in a diverse range of *approximate functionals*. Thanks to the compact nature of the compressed representation and the relatively straightforward computation of these approximate functionals, DFT proves to be exceptionally computationally efficient. Consequently, it stands out as the most practical method for handling exceedingly large systems. The accuracy of DFT outcomes is limited to systems where particle correlations remain relatively weak, rendering it particularly suitable for addressing many practical problems in fields such as organic chemistry.

Quantum Monte Carlo (QMC) encompasses a broad array of classical techniques that apply stochastic, or Monte Carlo, methods to quantum simulation problems. QMC methods typically circumvent the need to store an accurate representation of the ground state by recasting the ground state problem as a high-dimensional integral involving simple wave functions and operators. The complexity of solving this integral is managed through an approximate solution based on stochastic integration, facilitated by cleverly designed importance sampling. A poor choice in the

definition of the integral or of the importance sampling step can lead to a divergence in the variance of the result,thus requiring exponential sampling time to achieve the target accuracy. While this divergence often occurs in Fermionic systems due to the so-called *sign problem* [37, 38], techniques based on heuristic constraints enable QMC to obtain competitive results for complex problems in condensed matter physics and chemistry.

1.4.2 Hamiltonian simulation algorithms

The first approach to efficiently synthesize an evolution operator $U(t)$ on a quantum device, addressing the Hamiltonian simulation problem (see Section 1.3.1), was proposed by Lloyd in 1996 [21]. This technique, often referred to as Trotterization, adapts expansion methods for exponentials of operators developed by Trotter [22] and Suzuki [23]. The central concept revolved around recognizing that, for small time increments δt, the linearized evolution operator generated by the Hamiltonian $H(t)$ (decomposed as in Eq. (1.3)) can be approximated as

$$U(t, \delta t) = e^{-iH(t)\delta t} = \prod_{j \in [J]} e^{-ih_j(t)\delta t} + \mathcal{O}(\delta t^2). \qquad (1.7)$$

The second order error in δt is due to the non-commuting terms $[h_j, h_{j'}] \neq 0$. Since each $e^{-ih_j \delta t}$ only acts non-trivially on a constant-size subsystem $\mathcal{H}_j$, it can typically implemented in constant or linear time on a quantum device. To achieve the desired $U(t) = \prod_{\tau=1}^{t/\delta t} U(\tau \delta t, \delta t)$, the short-time evolution operator is synthesized with the desired accuracy and applied $\frac{t}{\delta t}$ times.

The approximation error in Trotterization can be controlled in various ways. This includes reducing δt, modifying the approximation (1.7) increasing the order of the remainder, or randomizing the sequence of the local operators $e^{ih_j \delta t}$ using a technique known as quantum stochastic drift protocol (qDRIFT) [39]. While analytical bounds for the Trotter approximation error are less favorable than for more advanced methods, they can sometimes be tightened by leveraging the Hamiltonian's locality through Lieb-Robinson bounds [40–42]. Extensions of these methods that aim to counteract the approximation using locality provide highly efficient techniques for simulating geometrically local Hamiltonians.

Trotter-like methods represent the simplest algorithms for Hamiltonian simulation, and to date they are the only ones experimentally tested on quantum devices. In some of these experiments, the approximation error is embraced and included in the models as a periodic drive potential,

resulting in so-called *Floquet models.*

Over the past decade, various techniques have been developed that take a fundamentally different approach than Trotter expansion to synthesize $U(t)$. These methods are typically based on the concept of block encodings of a sub-unitary operator H/λ (with $\lambda \geq \|H\|$), constructed by expressing H as a linear combination of unitary operators (LCU) [43–45]. The evolution $U(t)$ is then synthesized by transforming the block-encoded operator through techniques derived from quantum walk theory [43] and nuclear magnetic resonance control. These techniques are known as **qubitization** [46], **quantum signal processing** [47], and the quantum singular value transform [48].

All the techniques mentioned above are designed for universal quantum computers and can be applied to simulate any physical model given a fault-tolerant quantum device of sufficiently large scale. In contrast, *analog quantum simulators* natively implement the evolution according to a restricted class of Hamiltonians, defined by the underlying physical implementation and the experimental control on the device. Due to these limitations, each quantum simulator can only address a subset of simulation problems, with its capabilities heavily dependent on its specific implementation.

1.4.3 Adiabatic state preparation

The adiabatic algorithm [49] is an early technique for eigenstate preparation (see Section 1.3.2), grounded in the adiabatic theorem introduced by Born and Fock in 1928 [50]. According to this theorem, a system evolving under a time-dependent Hamiltonian $H(t)$ will remain in its instantaneous eigenstate $|E_j(t)\rangle$ if the Hamiltonian changes gradually enough (i.e., *adiabatically*) and there is a persistent energy gap between $E_j(t)$ and its neighboring eigenvalues $E_{j+1}(t), E_{j-1}(t)$. For simplicity, hereon we focus on ground states, considering only the ground state gap $\Delta = E_1 - E_0$.

The adiabatic theorem can be harnessed by constructing a time-dependent Hamiltonian

$$H(t) = \left(1 - \frac{t}{t_{\max}}\right) H_{\text{init}} + \frac{t}{t_{\max}} H_{\text{target}}, \qquad (1.8)$$

Here, H_{target} represents the Hamiltonian whose ground state $|E_0^{\text{target}}\rangle$ we aim to prepare, while H_{init} is a Hamiltonian whose ground state $|E_0^{\text{init}}\rangle$ is known and easy to prepare. After initializing the quantum device in the state $|E_0(0)\rangle = |E_0^{\text{init}}\rangle$, we can simulate the evolution dictated by
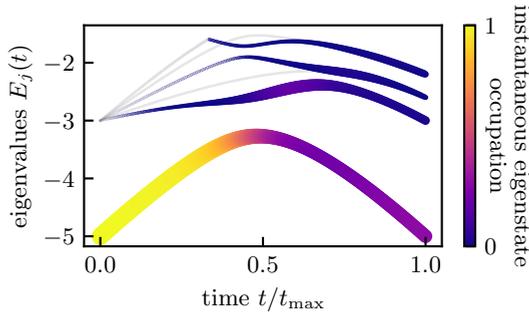
Figure 1.2: Example of nonadiabatic transitions. The plot shows the low-energy spectrum of an adiabatic algorithm Hamiltonian $H(t)$ [see Eq. (1.8)], for the same Ising chain model used in the simulations of Chapter 5. We plot the fidelity $|\langle E_j(t)|\psi(t)\rangle|$ of the adiabatic state $|\psi\rangle(t)$ with the instantaneous eigenstates $|E_j(t)\rangle$ (both in color and line size) throughout the progress of the quasi-adiabatic evolution. Note how around the gap closing at $\frac{t}{t_{\max}} \approx 0.5$ the fidelity with the instantaneous ground state is reduced in favor of the first excited state.

$H(t)$, ensuring $T_{\max}$ is adequately large to ensure the Hamiltonian changes adiabatically. If the system remains in the instantaneous eigenstate $|E_0(t)\rangle$, we will have successfully prepared the target ground state at time $t = t_{\max}$.

This form of the adiabatic theorem holds exactly in the infinite $t_{\max}$ limit, which is clearly unpractical. Finite rates of change in $H(t)$ result in *nonadiabatic transitions*, representing a transfer of some amplitude between instantaneous eigenstates, as depicted in Figure 1.2. The rate of transition depends on the rate of change of the Hamiltonian $\|\frac{d}{dt}H(t)\|$ (faster changes lead to larger transitions) and on the ground state gap $\Delta(t) = E_1(t) - E_0(t)$. The nonadiabatic transitions are stronger when the gap $\Delta(t)$ is small. While it is possible to enhance adiabatic algorithm outcomes by employing a schedule that adjusts the rate of change in $H(t)$ based on the gap, determining the gap a priori is challenging.

Rigorous bounds on the final state's approximation quality, dependent on the schedule and gap, have been established in a series of adiabatic theorems [51–53]. A popular version of the adiabatic theorem states that a time proportional to $[\min_t \Delta(t)]^{-2}$ is required to prepare the target ground state with fidelity $1 - \epsilon$.

The adiabatic algorithm is particularly suitable for analog quantum

simulators, which can natively implement dynamics generated by $H(t)$. On a universal quantum computer, implementing the adiabatic algorithm would entail simulating a time-dependent Hamiltonian, utilizing one of the algorithms discussed in the previous section and thus incurring additional computational overhead.

1.4.4 The variational quantum eigensolver

The variational quantum eigensolver (VQE) is a ground state preparation algorithm (see Section 1.3.2) tailored to the capabilities of NISQ devices, originally introduced in 2014 [19].

The VQE, illustrated in Fig. 1.3, is based on the interplay between a quantum device (or *quantum processing unit*, QPU) and a classical computer. The quantum device is used to run a subroutine that involves two main steps: 1. Preparation a parametrized *ansatz state* $|\psi(\boldsymbol{\theta})\rangle$; and 2. Estimation of the Hamiltonian expectation value $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle$ (as in Sec. 1.3.3). The ansatz state is generated through a set of parametrized operations on the quantum device, forming what is known as a *parametrized quantum circuit* (PQC). This subroutine is used by a classical algorithms, running an outer optimization loop with the objective of minimizing $E(\boldsymbol{\theta})$. Notably, since no state possesses lower energy than the ground state $|E_0\rangle$, the result of the VQE is *variationally bounded*, meaning $\min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) \geq E_0$.

An essential factor for the VQE's effectiveness is the design of the PQC preparing the ansatz state $|\psi(\boldsymbol{\theta})\rangle$. The set of ansatz states cannot cover the entire Hilbert space, since a classical variable $\boldsymbol{\theta}$ containing the same amount of information as the amplitudes characterizing the quantum state (which grows exponentially with the system size) would defeat the purpose of employing a quantum device. Instead, the choice of the PQC is guided by various heuristics. These can include extensions of perturbation theory (e.g., unitary coupled cluster [19]), constructions inspired by the problem's Hamiltonian (e.g., variational Hamiltonian ansatz [54], quantum alternating operator ansatz [55, 56]), or its symmetries (e.g., quantum-number preserving fabrics [57]). Adaptive ansätze have also been proposed, where the PQC is generated dynamically at runtime [58]. Given the importance of keeping the PQC compact, ansätze optimized for a specific hardware architecture are common in current proof-of-concept experiments [59].

While VQE provides an invaluable testing platform for quantum algorithm in the NISQ era, its potential to achieve practical quantum advantage is a topic of debate. Even with strong heuristics in the ansatz construction, low-dimensional ansatz manifolds may not accurately represent complex
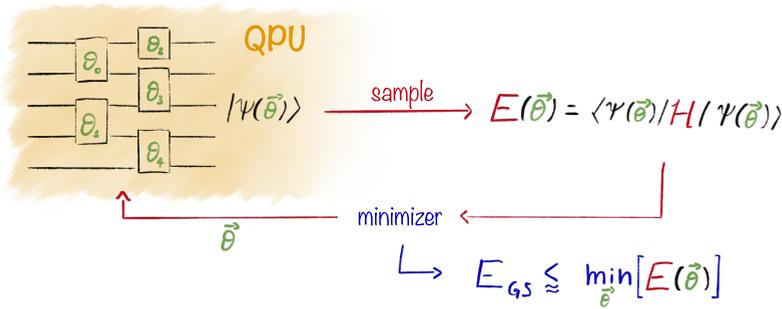
Figure 1.3:   Block scheme of the variational quantum eigensolver. The subroutine run on the quantum device (labeled QPU) prepares an ansatz state $|\psi(\boldsymbol{\theta})\rangle$ through a quantum circuit whose gates are parameterized by the elements θ_j of $\boldsymbol{\theta}$. The expected energy $E(\boldsymbol{\theta})$ of $|\psi(\boldsymbol{\theta})\rangle$ is measured by sampling on many repetitions of the circuit, and minimized by a classical outer loop.

correlated systems; certifying the quality of the result is also complicated. Minimizing the cost function $E(\boldsymbol{\theta})$ is difficult due to the complex, multimodal nature of the optimization landscape, occasional vanishing gradients, and added challenges from hardware and sampling noise. Even when VQE has the potential to prepare the ground state of a system of interest, obtaining a high.accuracy estimate of the final energy estimate necessitates a substantial number of shots. This cost is further increased by the necessary application of error mitigation techniques.

1.4.5 Quantum phase estimation algorithms

The problem of quantum phase estimation is closely related to that of eigenvalue sampling described in Section 1.3.3. The objective of phase estimation is to sample eigenvalues φ_j of a unitary operator U with probabilities proportional to the fidelity $p_j = |\langle \varphi_j | \psi \rangle|^2$ with an input state $|\psi\rangle$. When considering $U = e^{-iHt}$, the sampled unit eigenvalues φ_j can be mapped to corresponding eigenvalues of the generator H as $E_j = \arg(\varphi_j)/t$, provided that t is chosen appropriately to ensure the argument function is single-valued (typically $t < \pi\|H\|$).

   To gain insight into the relationship between the dynamics generated by H and the measurement of the observable H, we can draw parallels to an

early model of measurements introduced by Von Neumann [60][2]. In this model, a *pointer* representing the measurement apparatus is coupled to the system, with the aim to map the quantity we want to measure onto the pointer's state. This is represented in the left panel of Fig. 1.4. The pointer is described by a single continuous variable q, with generalized eigenstates $|x\rangle_{\mathrm{P}}$, initially in the state $|0\rangle_{\mathrm{P}}$. The displacement operator $D(\alpha) = e^{-i\alpha p}$ generated by the conjugate momentum p can be used to encode a number α on the pointer's state $D(\alpha)|0\rangle_{\mathrm{P}} = |\alpha\rangle_{\mathrm{P}}$. Correspondingly, the coupling the pointer and the system through the generator $p \otimes H$ yields

$$e^{-ip\otimes H\,t} |0\rangle_{\mathrm{P}} \otimes |\psi\rangle = \sum_j |E_j\rangle_{\mathrm{P}} \otimes |E_j\rangle \langle E_j|\psi\rangle \qquad (1.9)$$

Tracing out the original system leaves the pointer in the ensemble state $p_j |E_j\rangle\langle E_j|_{\mathrm{P}}$, yielding the desired result of eigenvalue sampling.

The classic quantum phase estimation algorithm (QPEA), introduced by Kitaev in 1995 [61], can be described as a Von Neumann measurement model with a digital pointer. The variable representing the pointer's position q is discretized and encoded in a binary register of m qubits, initialized to the all-zero state $|0\rangle_{P_{m-1}} ... |0\rangle_{P_0}$. A quantum Fourier transform (QFT) [62] is applied to the pointer register, transforming its basis to represent a discrete counterpart of the conjugate momentum p Subsequently, each qubit P_k in the pointer register is employed to control 2^k iterations of the operator $U = e^{-iHt}$, specifically using the operation $|1\rangle\langle P_k| U^{2^k} + |0\rangle\langle P_k| \mathbb{1}$. This effectively implements the discrete analog of the evolution $e^{-ip\otimes H\,t}$. Finally, an inverse QFT returns the pointer register to representing the pointer's position. The resulting state encodes an approximation of E_j, with precision depending on the size of the pointer register m and with probability proportional to the input state fidelity. A block diagram of the circuit for QPEA is shown in the top-right panel of Figure 1.4.

An extremely reduced version of the QPEA circuit can be created by modeling the pointer with a single qubit. The QFT on a single qubit coincides with a Hadamard gate, and the output of each run of the circuit is binary (0 or 1), hence the common name for this circuit is the **Hadamard Test (HT)**. The HT circuit is represented in the bottom-right panel of Figure 1.4. The Hadamard test is explored in depth as a model of a generalized measurement with binary output in Chapter 4. Although a single run of the HT is not particularly informative, the expectation value of

---

[2]The connection between Von Neumann measurement model and quantum phase estimation introduced in these pages was inspired by a presentation given by Seth Lloyd during his Lorentz professorship in Leiden, in 2019.
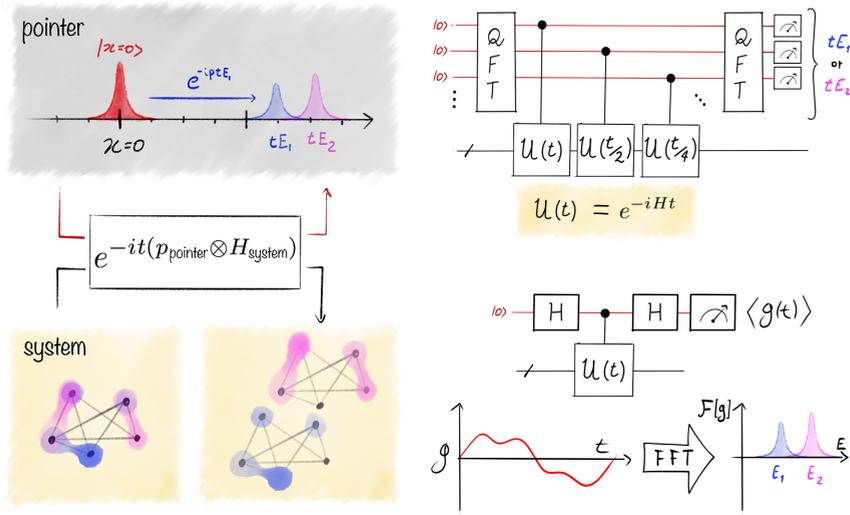
Figure 1.4:   (Left) Von Neumann measurement of the system energy H_{system} using a continuous-variable pointer. (Top right) Entanglement-based quantum phase estimation algorithm, employing the quantum Fourier transform (QFT). The binary-encoded result is read out on the pointer qubits, marked in red. (c) Single-control quantum phase estimation. The Hadamard test circuit it used to measure a signal function $g(t)$. Processing the signal with a fast Fourier transform (FFT) or a similar method yields a spectral density function.

the results (interpreted as -1 and $+1$ instead of 0 and 1) matches $\langle\psi|U|\psi\rangle$, with U being the controlled operator. Measuring this for the evolution operator $U(t) = e^{-iHt}$ at different times t allows us to reconstruct a *signal function* $g(t) \coloneqq \langle\psi|e^{-iHt}|\psi\rangle$. The frequencies of the Fourier components of $g(t)$, which can be extracted using various signal processing techniques, approximate the eigenvalues E_j, and their amplitudes approximate p_j. While this result is fundamentally different from eigenvalue sampling, both methods can be applied to the eigenvalue estimation problem described in Section 1.3.4. The algorithms based on this idea are commonly referred to as **single-control QPEA**s, and due to the shorter and simpler circuits they required they are more suited to the NISQ regime. In Chapter 3, we introduce an error mitigation technique specially tailored to single-control QPEAs.

1.5 Molecular simulation on a quantum computer

A quantum description of molecules – objects emerging from the interaction between atomic nuclei and electrons – stands as one of the early successes of quantum mechanics [63]. The theoretical study of molecules is one of the fields that most benefited from the development of classical computational simulation. Computational chemistry, a mature discipline, employs a diverse range of methods with applications spanning from fundamental research to industrial process development. Established classical techniques such as density functional theory, mean field and perturbation theory methods (described in Section 1.4.1) allow for efficient simulation of many large and complex molecules, particularly in organic chemistry. This empowers chemists to make predictions regarding molecular properties, geometries, reaction pathways and rates. Importantly, this bypasses the costly and sometimes unfeasible task of physically isolating or synthesizing the target molecule in the laboratory. Predictions from computational simulations can guide experiments, and provide valuable insight for chemical engineers, biochemists and other scientists in the development of industrial processes or the advancement of our understanding of the natural world.

Nonetheless, there are questions within the realm of quantum chemistry that pose challenges for classical techniques. These typically pertain the study of the electronic structure of *strongly correlated* molecules. In such systems, strong interactions disrupt perturbative assumptions, and electrons become entangled in complex ways that cannot be adequately described by the approximate models commonly used in computational simulations (like those described in Section 1.4.1). Quantum simulation holds the promise of providing a novel and distinct tool to tackle the unresolved questions in correlated quantum chemistry. The ability to precisely represent strongly-correlated quantum states renders quantum devices particularly attractive in this context.

While quantum hardware has not yet reached a point where it can support practical computations, ongoing research in quantum simulation is bringing the prospect of useful quantum advantage closer. This comprehensive research endeavor encompasses many stages: the identification of open problems that pose challenges to classical methods, the development of techniques to simplify and adapt these problems to quantum methods, the improvement of quantum algorithms tailored to address these specific challenges, and the establishment of benchmarks to test the algorithms and certify the results. The domain of computational quantum chemistry

offers a multitude of open questions with evident implications for both research and commercial applications.

1.5.1 The pipeline for electronic structure on a quantum computer

In this section, we introduce a key problem in quantum chemistry: the identification of the electronic structure ground state. Various quantum simulation methods aiming to address this have been proposed and explored at length. We show that this problem forms the basis for computing chemical reaction rates, probing molecular properties, and conducting molecular dynamics simulation. Following this, we outline the widely employed second-quantization approach, which simplifies the molecular model and facilitates its mapping onto a quantum device. The resulting quantum simulation problem can be addressed using some of the quantum algorithms introduces in Section 1.4.

The electronic structure ground state problem

A molecule is a system composed of electrons (here labeled as e) and atomic nuclei (labeled as N). The Hamiltonian of such a system can be decomposed in kinetic and electrostatic interaction terms[3],

$$H = \underbrace{T_{\mathrm{N}} + T_{\mathrm{e}}}_{\text{Kinetic energy}} + \underbrace{U_{\mathrm{NN}} + U_{\mathrm{ee}} + U_{\mathrm{eN}}}_{\text{Coulomb electrostatic energy}} . \tag{1.10}$$

The dynamics of these two groups of particles can be treated separately, following the approximation introduced by Born and Oppenheimer in 1927 [63], which is founded in large mass difference between nuclei and electrons. Initially, the heavier and slower nuclei are treated as immobile charges fixed at coordinates $\boldsymbol{R}$. Their dynamics can be reintroduced later, either classically (treating $\boldsymbol{R}$ as a classical variable) or semiclassically (quantizing $\boldsymbol{R}$ but neglecting its correlations with the electronic state). An *electronic structure Hamiltonian* can then be expressed as a function of $\boldsymbol{R}$

$$H_{\mathrm{e}}(\boldsymbol{R}) = U_{\mathrm{NN}}(\boldsymbol{R}) + U_{\mathrm{eN}}(\boldsymbol{R}) + T_{\mathrm{e}} + U_{\mathrm{ee}}. \tag{1.11}$$

---

[3]This molecular structure Hamiltonian neglects relativistic effects (including spin-orbit) and nuclear spins, which is often a reasonable approximation. In areas of theoretical chemistry where these effects are significant, the treatment can be adjusted accordingly.

1

The first term $U_{\mathrm{NN}}(\boldsymbol{R})$ is a nuclear repulsion constant which does not influence electronic dynamics, like the nuclear kinetic term T_{N} which we removed by fixing the nuclear positions. The term $U_{\mathrm{eN}}(\boldsymbol{R})$ now represents the potential experienced by electrons due to nuclear charges. The remaining operators T_{e} and U_{ee}, are independent of $\boldsymbol{R}$ and represent the kinetic energy of electrons and the electrostatic repulsion between electron pairs, respectively.

Solving the ground state problem for the electronic structure Hamiltonian Eq. (1.11) yields the energy $E_0(\boldsymbol{R})$ as a function of nuclear coordinates. The energy required to excite electrons to higher eigenstates is typically on the order of an electronvolt, which is two orders of magnitude larger than both the ambient temperature and the typical nuclear kinetic energy. This difference justifies the approximation of considering only electronic ground states, as well as the Born-Oppenheimer approximation. The local minima of $E_0(\boldsymbol{R})$ define the equilibrium geometries for the molecule. The shape of the potential around the equilibrium points defines the vibrational and rotational spectra of the molecule. More generally, nuclear motion can be reintroduced as governed by the potential $E_0(\boldsymbol{R})$, assuming that electrons follow their instantaneous ground state adiabatically. The $\boldsymbol{R}$-derivatives of $E_0(\boldsymbol{R})$ represent the forces acting on nuclei, which can be used in *molecular dynamics* simulations. Additionally, other molecular properties, such as polarizability, can be predicted based on the representation of the electronic structure ground state.

One relevant target of electronic structure calculations is the prediction of reaction rates. A chemical reaction can be viewed as a trajectory in nuclear coordinate space, starting from a stable state $\boldsymbol{R}_{\mathrm{R}}$, with the nuclei arranged as reactant sub-molecules, and ending at another stable state $\boldsymbol{R}_{\mathrm{P}}$ representing the products. Along this reaction path connecting $\boldsymbol{R}_{\mathrm{R}} \to \boldsymbol{R}_{\mathrm{P}}$, a saddle point $\boldsymbol{R}_{\mathrm{T}}$ signifies a transition state: the highest energy point along the path. The activation energy of the reaction $\Delta E_{\mathrm{A}} = E_0(\boldsymbol{R}_{\mathrm{T}}) - E_0(\boldsymbol{R}_{\mathrm{R}})$ is a key factor in determining the reaction rate constant k[4]. Both the (empirical) Arrhenius equation and the (first-principle) Erying equation, reveal that k depends on the activation energy exponentially, $k \propto e^{-E_{\mathrm{a}}}$. Therefore, resolving the activation energy to high precision is essential.

---

[4]Complex reactions may occur along multiple pathways, each consisting of several steps between metastable states. The overall reaction rate can be computed by applying the concepts described here to each individual step of the reaction and combining the results.

Discretization

Addressing the challenge just described amounts to solving the ground state problem, described in Section 1.3.4, for the electronic structure Hamiltonian Eq. (1.11). Tackling this with quantum simulation first requires to translate this Hamiltonian into the form of an operator acting on the state space space of a quantum device. A common approach, assuming a second-quantized treatment of the problem, encompasses three steps: 1. discretization of the single-particle state space, 2. mean-field calculation, leading to the selection of single-particle orbitals relevant to correlations, and 3. mapping fermionic operators to qubit operators.

The first step, widely used in computational chemistry, involves discretizing space using a set of $2M_{\mathrm{ao}}$ single-particle wavefunctions called *atomic orbitals*[5] $\chi_\mu(\boldsymbol{x}, \sigma) \in \mathbb{C}$, with $\boldsymbol{x} \in \mathbb{R}^3$ representing the electron's position and $\sigma \in \mathbb{Z}_2$ its spin. Atomic orbitals are chosen to approximate the low-energy mean-field eigenstates for individual atoms, and to be numerically integrable; they need not be orthogonal to each other. Various methods exist for selecting atomic orbitals, and the resulting collections of orbitals are referred to as a *basis set*. By combining a large number of atomic orbitals from different atoms, more complex single-particle wavefunctions known as *molecular orbitals* ϕ_p can be constructed:

$$\phi_p(\boldsymbol{x}, \sigma) = \sum_{\mu \in [2M_{\mathrm{ao}}]} C_{\mu p} \chi_\mu(\boldsymbol{x}, \sigma). \tag{1.12}$$

The combination of atomic orbitals to form molecular orbitals is illustrated in Fig. 1.5 (left and center). Molecular orbitals are typically chosen to be orthonormal.

An anti-symmetrized product of orthogonal molecular orbitals defines a many-electron state known as *Slater determinant*

$$\Phi\big(\boldsymbol{x}_1, \sigma_1; ...; \boldsymbol{x}_{N_{\mathrm{e}}}, \sigma_{N_{\mathrm{e}}}\big) = \sum_{\boldsymbol{\pi} \in \mathcal{P}_{N_{\mathrm{e}}}} \mathrm{sgn}(\boldsymbol{\pi}) \bigotimes_{p=0}^{N_{\mathrm{e}}} \phi_p(\boldsymbol{x}_{\boldsymbol{\pi}(p)}, \sigma_{\boldsymbol{\pi}(p)}), \tag{1.13}$$

where N_{e} is the number of electrons in the state and $\mathcal{P}_m$ is the set of permutations of size m. The state depends on the molecular orbital coeffi-

---

[5]The orbitals we define here are often called *spin-orbitals*, as their definition includes the spinor component of the wavefunction. It is common to further factorize *spatial orbitals* and spinors. We avoid this for the sake of synthesis, but we add a factor 2 in the number of spin-orbitals to maintains the compatibility with the common definition of M as the number of spatial orbitals.
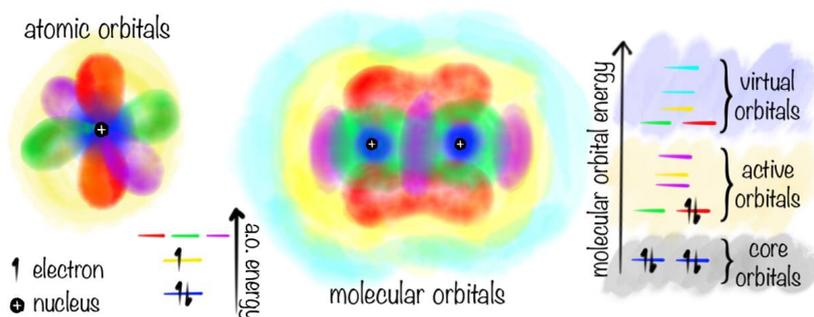
Figure 1.5: **Left**: artistic representation of the first few atomic orbitals and their energy levels. **Center**: atomic orbitals from two atoms combine to construct molecular orbitals. For the sake of clarity, we represent only a few of the first molecular orbitals generated by combining atomic orbitals from two atoms. **Right**: mean-field energy of each molecular orbital. One possible complete active space subdivision of orbitals is shown.

cients $C_{p\mu}$ through ϕ_p. A Slater determinant represents an uncorrelated many-body state, where each electron is not entangled to the others. Its energy can be computed efficiently in terms of integrals involving the atomic orbitals, the Coulomb operator $\frac{1}{\boldsymbol{x}_1 - \boldsymbol{x}_2}$, the kinetic operator $\nabla^2_{\boldsymbol{x}}$ and the nuclear positions and charges.

Reduction and second quantization

The Hartree-Fock self-consistent-field method (HF-SCF) efficiently optimizes the coefficients $C_{\mu p}$ to construct the minimal-energy Slater determinant Φ^{HF} and a set of orthogonal molecular orbitals ϕ_p^{HF}, ordered by their mean field energy $\epsilon_p^{\mathrm{HF}} \leq \epsilon_{p-1}^{\mathrm{HF}}$. To introduce correlations on top of the mean field state Φ^{HF}, one can consider superpositions of Slater determinants generated by transferring some of the electrons from *occupied* molecular orbitals (indexed by $p < N_{\mathrm{e}}$) to *unoccupied* molecular orbitals ($q \geq N_{\mathrm{e}}$). Transferring an electron between orbitals $p \to q$ always results in a positive change in mean-field energy $\Delta^{\mathrm{HF}}_{p \to q} = \epsilon_q^{\mathrm{HF}} - \epsilon_p^{\mathrm{HF}}$, but considering superpositions between such states can yield a lower, beyond-mean-field energy. However, when $\Delta^{\mathrm{HF}}_{p \to q}$ is large, the large increase in mean-field energy contrasts the gain from adding the excited Slater determinant to the

superposition. Therefore, when constructing the superposition of Slater determinants, we can limit to considering orbitals with energies within a selected window around Fermi energy $\epsilon_{\mathrm{F}} := \epsilon_{N_{\mathrm{e}}}^{\mathrm{HF}}$. The molecular orbitals are thus divided in three sets: a *core* set of fully-occupied low-energy orbitals for $p < 2M_{\mathrm{C}}$, an set of *active* orbitals $2M_{\mathrm{C}} \leq p < 2M_{\mathrm{C}} + 2M_{\mathrm{A}}$ used to describe correlations among $N_{\mathrm{A}} = N_{\mathrm{e}} - 2M_{\mathrm{C}}$ electrons, and the remaining space of high-energy *virtual* orbitals which are discarded in the calculation of correlations. This subdivision is illustrated in Figure 1.5. Calculations that prescribe this subdivision are called *complete active space* methods, and denoted by the abbreviation CAS($N_{\mathrm{A}}, M_{\mathrm{A}}$).

The electronic structure Hamiltonian Eq. (1.11) can be constrained to the active space, accounting for the effect of frozen electrons in mean-field terms. The resulting Hamiltonian can be divided into terms that act on zero, one single, or pairs of electrons in the active space:

$$H_{\mathrm{A}} = \overbrace{U_{\mathrm{NN}} + U_{\mathrm{NC}} + U_{\mathrm{CC}} + T_{\mathrm{C}}}^{\text{constant}} + \underbrace{T_{\mathrm{A}} + U_{\mathrm{CA}} + U_{\mathrm{NA}}}_{\text{one active electron}} + \overbrace{U_{\mathrm{AA}}}^{\text{two active electrons}}. \qquad (1.14)$$

In turn, this can be written in terms of second-quantized fermionic operators acting on a single electron ($c_p^\dagger c_q$) and pairs of electrons ($c_p^\dagger c_q^\dagger c_r c_s$)

$$H_{\mathrm{A}} = \mathrm{const.} + \sum_{p,q \in [2M_{\mathrm{A}}]} g_{pq} c_p^\dagger c_q c_r^\dagger c_s + \sum_{p,q,r,s \in [2M_{\mathrm{A}}]} h_{pqrs} c_p^\dagger c_q^\dagger c_r c_s. \qquad (1.15)$$

The coefficients g_{pq} and h_{pqrs} are obtained through integrals of the atomic orbitals with the kinetic and Coulomb operators and are known as *one-electron* and *two-electron integrals*, respectively. The number of Slater determinants that can be obtained by distributing N_{A} electrons over $2M_{\mathrm{A}}$ orbitals is the binomial coefficient $\frac{(2M_{\mathrm{A}})!}{N_{\mathrm{e}}!(2M_{\mathrm{A}} - N_{\mathrm{e}})!}$, which grows exponentially in the size of the active space. The description of the Hamiltonian, instead, only requires the $\mathcal{O}(M_{\mathrm{A}}^4)$ integrals.

Describing electronic structure accurately is most challenging in systems that necessitate large active spaces. The appropriate active space size is determined by chemical intuition, often guided by approximate classical simulations. The size of the molecule is not a reliable predictor of the required active space size; many large organic molecules are not particularly challenging because the electrons predominantly occupy bonding orbitals, well described by mean-field theory. There, transferring electrons to empty orbitals has a large energy cost and thus correlations remain small. Methods like density functional theory predict the ground state energy of such

systems efficiently and accurately. An example of a challenging problem is given by molecules containing transition metals; these provide many atomic *open-shell* orbitals, which lead to a number of complex molecular orbitals with similar energies which need to be considered in the active space construction. One such case is nitrogenase, an enzyme crucial for converting atmospheric nitrogen into ammonia, containing iron and molybdenum atoms embedded in an organic matrix (FeMo cofactor). A deeper understanding of this enzyme could potentially lead to improvements in industrial process of ammonia synthesis (critical, among others, for fertilizer production), currently achieved through the costly Haber-Bosch process.

Mapping to a quantum computer

The Slater determinants defined over active molecular orbitals constitute a set of orthogonal states. These can be mapped onto the states of a quantum device, serving as a basis to represent more general correlated states. Similarly, the Fermionic creation and annihilation operators $c_p^\dagger, c_p$ that construct the Hamiltonian Eq. (1.15) need to be mapped onto operators on the device's Hilbert space. The Jordan-Wigner transformation, a common mapping between fermionic systems and qubit system, prescribes assigning one qubit to each molecular spin-orbital. An occupied spin-orbital is represented by the state $|1\rangle$ of the corresponding qubit, while an empty orbital corresponds to $|0\rangle$. The mapping of fermionic operators is defined as

$$c_p^\dagger = \bigotimes_{j=0}^{p-1} \sigma_j^z \otimes \sigma_j^+; \quad c_p = \bigotimes_{j=0}^{p-1} \sigma_j^z \otimes \sigma_j^-, \tag{1.16}$$

where σ_j^z is the Pauli Z operator on the j-th qubit and $\sigma_j^\pm = \frac{1}{2}(\sigma_j^x \pm i\sigma_j^y)$ are the raising/lowering operators. This mapping of fermionic operators ensures the correct anti-commutation relations, but sacrifices locality.

Solution by quantum simulation

With a model of the electronic structure problem represented in terms of states and operators on a quantum device, we can now explore the use of quantum simulation algorithms to solve it. If the goal is to solve the ground state problem on a NISQ device, one viable approach is to employ the Variational Quantum Eigensolver (VQE), as described in Section 1.4.4. Several VQE ansätze tailored to chemistry problems have been developed [64, 65], with the original proposal of the VQE being motivated by implementing a unitary version of the coupled-cluster ansatz

[19]. Various optimizations have been proposed for estimating expectation values (see Section 1.3.3) of a Hamiltonian of the form Eq. (1.15) [66–68]. An effective and practical method consists in measuring the *one-* and *two-electron reduced density matrices* of the input state $|\psi\rangle$, respectively

$$D_p^q = \langle\psi|c_p^\dagger c_q|\psi\rangle \quad \text{and} \quad D_{pq}^{rs} = \langle\psi|c_p^\dagger c_q^\dagger c_r c_s|\psi\rangle . \tag{1.17}$$

In practice, current quantum devices can only handle very small toy models of molecules using VQE. The largest VQE simulation for electronic structure with the final expectation value estimation step performed on a quantum computer utilized an active space of only 10 orbitals (with a further approximation simplifying the spin degree of freedom), and it relied heavily on error mitigation techniques [69], including the one introduced in Chapter 3.

Proposals for fault-tolerant quantum simulation of chemistry typically focus on the application of the quantum phase estimation algorithm (see Section 1.4.5) [70]. Using QPEA to solve the ground state problem requires a Hamiltonian simulation subroutine and a technique for preparing a state with a significant overlap with the ground state. Recent years have seen the proposal of numerous optimized Hamiltonian simulation algorithms for electronic structure, many of which employ qubitization and LCU techniques (see Section 1.4.2). These methods primarily concentrate on constructing efficient compressions of the highly structured information contained in the one- and two-electron integrals, making them more suitable for uploading onto quantum computers [71–74]. Preparing a state with a good overlap with the ground state remains a challenging problem. It has been suggested that this step could potentially hinder the achievement of exponential quantum advantages in simulating electronic structure ground states [75]. Since most molecules naturally approach their electronic ground state through processes like thermalization and adiabatic evolution along a reaction coordinate, we expect heuristic methods can be used to efficiently emulate these phenomena. Dissipative algorithms, such as the one proposed in Chapter 2, or adiabatic state preparation, could be optimized to approximate molecular ground states effectively.

1.5.2 The journey towards useful quantum advantage in chemistry

The example presented in the previous section illustrates the multitude of decisions involved in defining a pipeline for studying a physical phenomenon using quantum simulation. These decisions impact the level of

approximation, precision of results, and the practicality of the method. Research in simulation can address any step of the pipeline, from the very choice of the problem to study to the final algorithmic result.

Extensive research has been conducted on the problem of electronic structure ground states in quantum simulation, with efforts focused on optimizing molecule modeling, mapping, and various quantum algorithms – in the previous section, we only presented the most common approach. An important body of research today includes the estimation of the costs of running quantum simulation algorithms on prospective quantum computer, especially for fault-tolerant approaches. For instance, in [74], it was estimated that a fault-tolerant quantum computer could execute the Quantum Phase Estimation Algorithm (QPEA) for a model of the nitrogenase FeMo cofactor, with a precision sufficient for estimating reaction rates, in four days using approximately 4 million physical qubits (assuming reasonable but as yet unachieved physical error rates).

To attain useful quantum advantage, it is essential to identify problems that are easy to solve through quantum simulation but challenging for classical methods. The ground state problem for molecules containing transition metals is hard for classical methods, due to the large size of the active space in which correlations need to be accounted for non-perturbatively. However, classical approximate methods for electronic structure ground state continue evolving, with some of them being able to approximate more and more complex correlations [75], making the remaining challenging cases even more demanding for quantum solutions.

Another set of problems challenging for classical simulation is provided by photochemistry, where low-lying excited electronic eigenstates play a significant role in reactions initiated by light rather than thermal energy. Describing excited states on a quantum computer is not significantly more complex than describing ground states, while classical simulation algorithms often exploit ground-state-specific properties. Photochemical reactions also involve the conversion of light energy into phonons, which can lead to the breakdown of the Born-Oppenheimer approximation, especially near electronic structure spectrum degeneracy points known as *conical intersections* (detailed in Chapter 6). The numerical study of chemistry beyond the Born-Oppenheimer approximation is very limited due to the complex correlations between electrons and nuclei, and the difficulty of discretizing space in the absence of clear atomic orbitals. A concurrent simulation of electrons and nuclei could be in principle achieved through first-quantized quantum simulation methods [76]. The size of the quantum computers required to implement these methods is beyond the foreseeable future, but in the long term such quantum computers could be achieved.

Another example of shift in the problem focus includes the recent proposal of using Hamiltonian learning to reconstruct NMR spectra [77, 78]. This problem also pertains the field of computational quantum chemistry, but it completely bypasses the electronic structure Hamiltonian, focusing on the interactions between nuclear spins instead. Finally, quantum algorithms could be used to assist known classical methods in innovative ways. Finally, quantum algorithms can complement existing classical methods in innovative ways, as seen in the *quantum computing quantum Monte Carlo* (QC-QMC) proposal. There, sampling on a variational quantum state provides information about correlations in the electronic structure wavefunction, while high-precision energy calculations are performed classically.

1.6 Outline of this thesis

This thesis introduces a number of quantum algorithms tailored to the simulation of physical systems. Chapter 2 introduces a category of algorithms aiming to prepare ground states of natural systems through simulated cooling. Chapter 3 presents a novel error mitigation approach, *echo verification*, and investigates its application in single-control phase estimation algorithms. The two following chapters explore applications of echo verification in different contexts: Chapter 4 develops techniques for estimating expectation values within the measurement model induced by echo verification; Chapter 5 applies echo verification to mitigate the effect of non-adiabatic transitions in the adiabatic algorithm. Finally, Chapter 5 explores the challenge of detecting conical intersections in molecular models – an under-explored problem in quantum chemistry, well-suited to quantum simulation – and proposes a resilient quantum algorithm to solve it.

Chapter 2: Quantum Digital Cooling

In chapter 2, we explore the idea of simulating cooling by coupling the system to a single-qubit "fridge". This auxiliary qubit is reset periodically to its low-energy state, allowing to extract energy and entropy from the system. The use of a single-qubit fridge differs from the natural thermalization process, where systems cool by releasing energy into extensive, cold, and ergodic baths. While simulating such baths is theoretically possible, it comes with a significant computational burden. Our investigation of single-qubit fridges leads us to introduce a category of algorithms designed for preparing the ground states of simulated Hamiltonians, which we name

quantum digital cooling. We establish two approaches to quantum digital cooling: a efficient strong-coupling approach which coarsely approximates the ground state, and a more expensive approach that can achieve arbitrary accurate approximations. We then assess the performance of both methods through numerical benchmarks.

Chapter 3: Error mitigation via verified phase estimation

Chapter 3 introduces a novel error mitigation technique that leverages the quantum information remaining in the device's state following a Hadamard-test-based measurement. This method involves checking if, after the measurement, the system state can be projected back onto the input state. This *verification* step has a high probability of failure in the presence of errors, enabling the rejection of erroneous results, and it does not introduce any bias to the measurement outcome when errors are absent. Our technique finds successful application within the framework of the single-control quantum phase estimation algorithm (introduced in Section 1.4.5), leading to the development of the *verified phase estimation (VPE)* algorithm. We demonstrate the application of VPE to both phase estimation and expectation value estimation problems for various models, showing improvements of several orders of magnitude over unmitigated estimation at near-term error rates.

Additionally, this chapter introduces a variant of VPE, known as *control-free VPE*, which eliminates the need for a control qubit. This simplifies the control circuitry significantly, bringing near-term implementations of our technique within reach.

In further literature, this error mitigation technique gained recognition under the name of *echo verification* [69, 79], a terms that draws parallel with the Loschmidt echo [80]. Control-free VPE has also been tested experimentally in a superconducting quantum processor, realizing simulations of systems of up to 10 qubits, and thus leading to the implementation of the largest variational algorithm for a correlated chemistry system to date [69].

Chapter 4: Optimizing the information extracted by a single-qubit measurement

In chapter 4, we study expectation value estimation (see Section 1.3.3) in a restricted model of quantum computation, where we are only allowed to extract a single bit of information per each n-qubit quantum state preparation. This restriction is motivated by echo verification, where all

but one qubit are used in the verification step to detect errors. We indeed show that, in echo-verification-like schemes, extracting more than one bit of information is counterproductive towards estimating expectation values.

Within this restricted model, we optimize expectation value estimation by decomposing the target observable into a sum of bitwise-measurable terms. We construct optimal decompositions analytically, and we propose a set of rules to improve on a given decomposition which can sometimes be applied even in presence of experimental constraints. We find the optimal decomposition of a fast-forwardable operator, and show a numerical improvement over a simple Pauli decomposition by a factor $n^{0.7}$.

Chapter 5: Virtual mitigation of coherent non-adiabatic transitions by echo verification

In chapter 5, we develop an extension of echo verification tailored to applications to the adiabatic state preparation algorithm (see Section 1.4.3). This technique, which we call *adiabatic echo verification*, mitigates both coherent and incoherent errors arising, respectively, from non-adiabatic transitions and hardware noise. This is an unconventional application of error mitigation, which is typically applied to hardware errors only. Even in the absence of hardware noise, the estimator bias of the observable is reduced when compared to standard adiabatic preparation, achieving up to a quadratic improvement.

Our method is based on two quasi-adiabatic evolutions with mirrored schedules implementing the echo. These are interleaved by a dephasing step by random-time evolution, and by the Hadamard-test-based measurement. The dephasing promotes coherent errors from non-adiabatic transitions into incoherent errors, making them amenable to verification. Our method requires positive-time dynamics only, making it more suitable to application in analog quantum simulators.

Chapter 6: A hybrid quantum algorithm to detect conical intersections

In chapter 6, we tackle a problem of interest in photochemistry: the detection of conical intersections in molecular models. The chapter introduces a quantum simulation algorithm tailored to the NISQ era to solve this problem.

Conical intersections are significant points in the geometry of a molecule, where the electronic potential energy surfaces cross in a topologically protected manner. Close to a conical intersection the Born-Oppenheimer

1

approximation breaks down, with crucial implications for chemical processes like photoisomerization and non-radiative relaxation. One prominent example is their role in the vision process, where a conical intersection facilitates the isomerization of retinal after absorbing a photon, which in turn triggers a cascade of chemical signal that result in the perception of light.

Conical intersections are characterized by a non-zero Berry phase, a topological invariant defined on a closed path in atomic coordinate space. The berry phase assumes the value of π when the path encircles the intersection manifold, and 0 otherwise. The algorithm we propose tracks the approximate ground stare along the chosen path, using a parametrized quantum circuit ansatz updated by a fixed number of Newton-Raphson steps. At the end of the algorithm, a Hadamard test is used to measure a single bit of information, which determines whether the Berry phase is π or 0. Since the final result is discrete, our procedure succeeds even for a cumulative error bounded by a constant; this allows us to bound (analytically) the total sampling cost and to readily verify the success of the procedure. The application of our algorithm is demonstrated numerically on a small toy model of the formaldimine molecule ($H_2C{=}NH$).

Quantum digital cooling

2.1 Introduction

Ground state preparation is an essential algorithm in the quantum computing toolbox. Any polynomial-time quantum algorithm can be mapped to the problem of estimating the ground state energy of an artificial Hamiltonian given an approximation to its ground state [81], and without such additional input this problem is known to be QMA-hard for even 2-local Hamiltonians [30]. Digital quantum simulation of problems in materials science and chemistry, one of the 'killer apps' of a quantum computer, is most often concerned with properties of ground states of the simulated systems [15, 21], and many problems in optimization may be mapped to ground state finding problems [49, 55]. This has led to a wide range of schemes for digital ground state approximation, via adiabatic evolution [49], variational methods [19, 55, 82], phase estimation [83], amplitude amplification [84–86] and approximate imaginary time evolution and other Hamiltonian function techniques [28, 87–89]. However, these algorithms suffer from large computational costs or approximation errors, making designing better schemes an active area of interest.

In nature, ground states are achieved by coupling to a large cold reservoir, which takes energy from the system in keeping with the second law of thermodynamics. Simulating an entire bath would require an impractically

large quantum register, however it has long been suggested that this may be mimicked by coupling to a single qubit which may be reset to its ground state with sufficient frequency [21]. This idea has been since studied in digital quantum computing for the initialization of quantum devices [90, 91] and as an inspiration of an algorithm based on resonant transitions and postselection [92]. This idea was also explored in analog simulation settings, for the preparation of physical [93] and artificial [94, 95] ground states. However, cooling an artificial system in the digital quantum setting provides a set of unique challenges — the system being studied may be completely different from the physical quantum hardware, and the digitized Hamiltonian may be only an approximation to the target of interest. Furthermore, the periodic non-unitary reset may break the unitary evolution in short time-scale chunks which do not conserve energy, implying that one may artificially reheat the system without clever protocol design. This is of critical importance in near-term devices, where limited coherence times compete against the desire for slower cooling cycles.

In this chapter, we detail how one may prepare ground states of an artificial Hamiltonian on a digital quantum computer via quantum digital cooling (QDC). We first present an analytic study of the cooling of a two-level system, from which two different approaches may be outlined to de-excite to the ground state whilst preventing reheating. We investigate the behaviour of both methods in the digitized setting, and find they continue to be robust. The protocols deriving from these two principles are tested in the one-qubit black-box Hamiltonian setting, where the energy gap and matrix elements are unknown. We extend these protocols to N-qubit systems, and investigate their ability to cool small-scale simulations of the transverse-field Ising model numerically. Our LogSweep protocol, based on the weak-coupling approach, is demonstrated to converge to the ground state of all three phases of the transverse-field Ising model. It further shows a relative energy error constant in the system size at a fixed protocol length for the weakly-coupled and critical phases of this model, which corresponds to a requirement to simulate time evolution for $O(N^2)$ and $O(N^3)$ Trotter steps respectively. By contrast, the stong-coupling BangBang protocol shows the ability to prepare low-cost ground-state approximations of the same model in the paramagnetic and ferromagnetic regime, but seems to perform much worse close to the critical point, where the system spectrum shows a less-ordered structure. The small number of calls to the system evolution operator needed to realize this protocol makes it attractive for near term application.

2.2 Cooling a system with a single fridge qubit

In nature, gapped physical systems cool down to a state with high overlap to the ground state when interacting with a bath that is cold and large, under the condition of ergodicity. By *cold*, we mean that temperature T_{B} of the bath is small compared to the ground state gap Δ_{S} of the system to be cooled: $k_{\mathrm{B}} T_{\mathrm{B}} \ll \Delta_{\mathrm{S}}$ (with k_{B} Boltzmann's constant). By *large*, we mean that the bath has a sufficiently large Hilbert space that the above condition is unchanged by the addition of the energy from the system. By *ergodic*, we mean the system must not have symmetries that prevent excitations to be transferred from the system to the bath, or that reduce the effective size of the accessible bath Hilbert space. Given a system with Hamiltonian H_{S} and eigenstates $H_{\mathrm{S}}|E_j\rangle = E_j|E_j\rangle$, energy conservation implies that the bath must have states at energies $E_j - E_0$ to allow de-excitation of the eigenstates E_j. This is typically achieved by considering a bath with a continuous or near-continuous low-energy spectrum [Fig. 2.1(a)]. However, the bath need not cool all states immediately to the ground state. Instead, a bath typically absorbs single quanta of energy $\epsilon = E_i - E_f$ that correspond to local excitations of the system $|E_i\rangle \rightarrow |E_f\rangle$, at a rate given by Fermi's golden rule:

$$\frac{dP_{\mathrm{i}\rightarrow\mathrm{f}}}{dt} = \frac{2}{\hbar} \int_0^\infty d\epsilon \, |\langle E_f, \epsilon | H_{\mathrm{C}} | E_i, 0 \rangle|^2 \, \rho_B(\epsilon)$$
$$\times \lim_{t\rightarrow\infty} \frac{\sin[(E_i - E_f - \epsilon)\,t]}{E_i - E_f - \epsilon} \tag{2.1}$$
$$= \frac{2\pi}{\hbar} \, |\langle E_f, \epsilon | H_{\mathrm{C}} | E_i, 0 \rangle|^2 \, \rho_B(E_i - E_f), \tag{2.2}$$

where H_{C} is the coupling between the system and the bath, and ρ_B is the density of states of the bath[1]. This requires the bath to be large enough to prevent re-excitation of these states as the system continues cooling. In other words, the bath must have a large Hilbert space compared to the one of the system. This ensures that, at equilibrium, most of the entropy is distributed in the bath.

To represent such a large bath with an auxiliary register on a quantum device in order to cool a system register would be impractically costly. We propose approximating the presence of a much larger bath B with a single

---

[1] In the rest of this chapter we drop $\hbar$ factors, assuming a choice of units for which $\hbar = 1$.

auxiliary qubit F [Fig. 2.1(b)], with bath Hamiltonian $H_B \rightarrow H_F = \epsilon\, Z_F/2$. This can be considered a simplified form of a collisional model [96] that does not allow for non-Markovian effects (that would be, in our case, unwanted). The coupling between the bath and the system takes the form $H_C = \gamma X_F \otimes V_S/2$, where γ is the coupling strength, and V_S a *coupling term* that acts on the system alone. A key advantage of the digital approach is that we are free to choose V_S as desired to optimize the cooling protocols. The Hamiltonian of the entire system and bath then takes the form

$$H = H_S + H_F + H_C. \tag{2.3}$$

This has the immediate problem that the bath can only absorb a single quantum of energy ϵ; howvever, we may circumvent this by periodically resetting the auxiliary qubit to $|0\rangle$. The non-unitary reset effectively extracts energy and entropy from the auxiliary qubit to a much larger external bath (i.e. the experimenter's environment). For this reason, we call this auxiliary qubit a *fridge qubit* (hence F). The non-unitarity introduced in the process is necessary to dissipate entropy, allowing to prepare the ground state from an arbitrary starting state. As the time between resets is finite, the $t \rightarrow \infty$ limit in Eq. (2.1) is no longer justified and energy is no longer conserved. This is both a blessing and a curse — we need not precisely guess the energy gap $\Delta = E_i - E_f$ of the transition that we need to de-excite, but we run the risk of reheating the system at each cooling round. Minimizing re-heating while maximizing the range of targeted de-excitations is key to the successful design of QDC protocols. In a realistic experiment, qubit re-heating would be effectively increased by reset infidelity on the auxiliary qubit, making the protocol less effective.

2.3 De-exciting a single transition: the 1+1 model

In order to design some basic protocols for QDC, we turn to a toy '1+1' model. We take a single-qubit system with Hamiltonian $H_S = \Delta\, Z_S/2$, and couple it to a single fridge qubit with coupling term $V_S = X_S$. Although this model is simple, it can represent a pair of levels being targetted for cooling in a larger quantum system. We will make use of this interpretation when extending these cooling protocols in section 2.4.
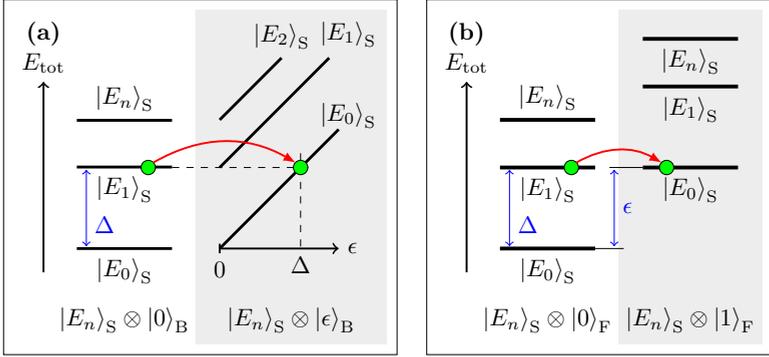
Figure 2.1: The de-excitation of the system transition $|E_1\rangle_S \to |E_0\rangle_S$ mediated by: **(a)** a continuous-spectrum natural bath B, where an excitation $|\epsilon\rangle_B$ at energy ϵ is produced, and **(b)** a single-qubit digital fridge F, which can be excited if $\epsilon = \Delta$.

2.3.1 Elementary approaches to digital cooling: strong and weak coupling

Let us first assume Δ is known, in which case resonant cooling ($\epsilon = \Delta$) is the most effective choice of ϵ. With this fixed, the transition probabilities after time t may be calculated exactly to be

$$P_{1\to 0} = \sin^2\left(\frac{\gamma}{2}t\right), \; P_{0\to 1} = \frac{\gamma^2 \sin^2(t\Omega)}{4\Omega^2}, \qquad (2.4)$$

where $\Omega = \sqrt{\gamma^2/4 + \epsilon^2}$. We wish to maximise the cooling probability $P_{1\to 0}$ while minimizing the reheating probability $P_{0\to 1}$ by optimizing the remaining free parameters: the coupling strength γ and the cooling time t. To maximize the cooling rate $P_{1\to 0} = 1$, we must set

$$t = \pi\gamma^{-1}. \qquad (2.5)$$

We assume this constraint throughout this chapter. Note that this choice is beyond the perturbative regime $\gamma t \ll 1$, for which the Fermi golden rule Eq. (2.1) is formulated.

To minimize reheating, we can then choose either of two very different approaches, based on strong or weak coupling. The weak-coupling approach is based on the observation that the off-resonant transition probability $P_{0\to 1}$ is bounded by $\gamma^2/4\Omega^2$. As such, we may suppress reheating to

an arbitrary level by choosing sufficiently small γ. The time-cost for Hamiltonian simulation of e^{iHt} scales at best linearly in t [97], so this implies one may obtain the ground state with failure probability p in time $O(p^{-1})$, regardless of the initial state of the qubit. The strong-coupling approach consists of tuning γ so that $\Omega t = \pi$, which is achieved when

$$\gamma = \frac{2}{\sqrt{3}}\epsilon. \tag{2.6}$$

This fixes the reheating exactly to 0, guaranteeing the qubit to be in the ground state in the shortest possible time, but at the cost of requiring fine-tuning.

Unlike in analog quantum simulation, digital devices cannot exactly implement the dynamics of the Hamiltonian in Eq. (2.3), and must approximate it digitally instead. A common approach to such digitization is that of the Suzuki-Trotter expansion [22, 23], which we now explore for the two cooling paradigms. We apply the (second-order) expansion of the coupled system-bath evolution with Trotter number M,

$$e^{-i(H_\mathrm{S}+H_\mathrm{F}+H_\mathrm{C})\,t} \sim \left[e^{-iH_\mathrm{C}\frac{t}{2M}} e^{-i(H_\mathrm{S}+H_\mathrm{F})\frac{t}{M}} e^{-iH_\mathrm{C}\frac{t}{2M}} \right]^M. \tag{2.7}$$

Note that, when we later approach larger systems, we will practically realize $e^{-iH_\mathrm{S}t/M}$ as a single second-order Trotter step, effectively implementing a second-order Trotter simulation of the coupled Hamiltonian $H_\mathrm{S}+H_\mathrm{F}+H_\mathrm{C}$ with trotter number M. If we restrict to the subspace containing the states involved in the cooling transition $|10\rangle_\mathrm{SF} \to |01\rangle_\mathrm{SF}$, at resonant cooling we have $H_\mathrm{S} + H_\mathrm{F} \propto \mathbb{1}$ (specifically, in this model $|01\rangle$ and $|10\rangle$ generate a zero-eigenvalue subspace of $H_\mathrm{S} + H_\mathrm{F}$). Thus, the Trotterized evolution behaves exactly like the continuous one with regards to the cooling transition. We study reheating probabilities as a function of t for different values of M in the weak-coupling regime. We observe (Fig. 2.2) that the digitized evolution approximates well the behavior of the continuum limit whenever $t\Omega/\pi \lesssim M$ (i.e. for the first M Rabi oscillations with pulse Ω). For longer times $t\Omega\pi \gtrsim M$, the second-order Trotter approximation fails, leading to reheating rates far larger than in the continuum limit. This allows us to define a practical choice of M to avoid reheating due to the Trotter approximation – we require

$$M \gg \sqrt{1 + \epsilon^2/\gamma^2}, \tag{2.8}$$

which sets the working point $t = \pi\gamma^{-1}$ before the $M/2$ Rabi oscillation.
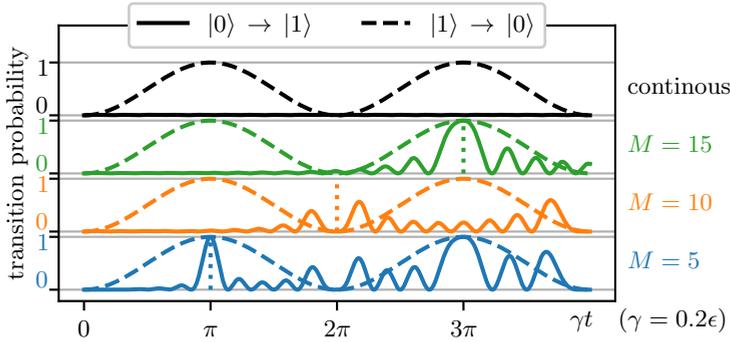
Figure 2.2: Effects of Trotterization on cooling and reheating probabilities as a function of the coupling time t, for different numbers of Trotter steps M per cooling cycle. Vertical dotted lines indicate the M-th reheating oscillation, at which point the Trotter approximation fails.

In the strong-coupling case $t\,\Omega/\pi = \sqrt{3}$, which implies that a single step is sufficient to satisfy the condition above. Indeed, digitized cooling with probability 1 and no reheating can be realized by a *bang-bang* approach (inspired by similar approach in variational methods [98, 99]). This consists in defining the evolution as in Eq. (2.7) with $M = 1$, while adjusting the coupling strength to $\gamma = 2\epsilon$. With this choice, the digitized evolution implements resonant Ramsey interference on the cooling transition $|10\rangle_{\mathrm{SF}} \to |01\rangle_{\mathrm{SF}}$ and anti-resonant Ramsey interference on the reheating transition $|00\rangle_{\mathrm{SF}} \to |11\rangle_{\mathrm{SF}}$.

A key difference between the two approaches to digital cooling is in their behavior off-resonance, i.e. when the energy gap is mis-targeted or not precisely known. For the bang-bang approach, detuning reduces the cooling efficiency while symmetrically boosting reheating [Fig. 2.3(a)]. The wide resonance peak around zero detuning makes this approach ideal to quickly cool transitions which energy is known up to a small error. In the weak-coupling approach the resonance peak becomes sharper and the reheating gets more suppressed as the coupling is made smaller [Fig. 2.3(b)], approaching the energy conservation limit. Detuning makes cooling inefficient, but thanks to the low reheating probability this weak-coupling cooling can be iterated with different values of ϵ to try to match the transition energy, without destroying the cooling effect.
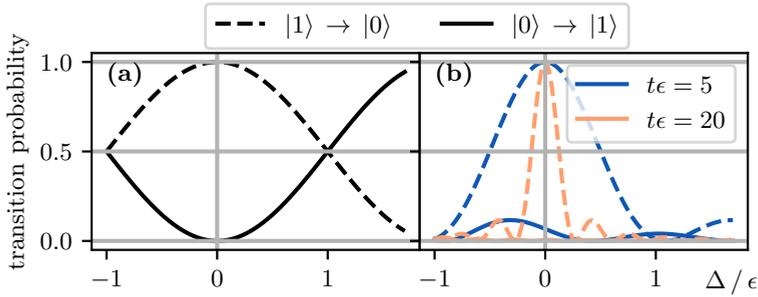
Figure 2.3:   Effect of fridge-system detuning $\delta = \Delta - \epsilon$ on the cooling (dashed lines) and reheating (solid lines) probabilities for **(a)** the bang-bang cooling approach, and **(b)** the weak-coupling cooling approach, where colors indicate different couping strengths.

2.3.2 Common symmetries and the coupling alternation method

In large systems of interest, we do not know the Hamiltonian's eigenstates, making it more challenging to couple between them. This is required for cooling, as can be seen by the direct dependence of the cooling rate $\frac{dP_{i \to f}}{dt}$ on the overlap $|\langle E_f, \epsilon | H_{\mathrm{C}} | E_i, 0 \rangle|^2$ in Eq. 2.2. This overlap dependence will prohibit cooling if the system Hamiltonian H_{S} and the coupling operator V_{S} share a common symmetry S (i.e. $[S, H_{\mathrm{S}}] = [S, V_{\mathrm{S}}] = 0$). When this is the case, the Hamiltonian may be simultaneously diagonalized with H_{S}, and a state with some overlap to any eigenspace of S that does not contain the ground state cannot be cooled to the ground state by coupling with V_{S}. Note that this is a strictly stronger condition than just requiring $[H_{\mathrm{S}}, V_{\mathrm{S}}] \neq 0$. A simple solution is to alternate over a set of couplings $\{V_{\mathrm{S}}^i\}$ as we cool. Then, a symmetry S of H would need to commute with *each* V_{S}^i in order to guarantee that a state starting with overlap in a high-energy eigenspace will remain there. Sets of coupling terms $\{V_{\mathrm{S}}^i\}$ on N qubits need only be size $O(N)$ to break commutation with all non-trivial operators (for example, the set of all single-qubit Pauli operators), so although symmetries need to be taken into account, they will not destroy the feasibility of QDC protocols.

This issue may be demonstrated on the protoype $1 + 1$ qubit model by considering the system Hamiltonian $H_{\mathrm{S}} = h\,\vec{n} \cdot \vec{\sigma}$, where $\vec{n}$ is a 3-dimensional unit vector (so H_{S} points in a random direction on the Bloch

$\overline{P_{1\to0}} = 94.7 \pm 13.9\%$ $\overline{P_{1\to0}} = 98.5 \pm 0.8\%$ $\overline{P_{1\to0}} = 98.4 \pm 0.5\%$
$\min(P_{1\to0}) = 0\%$ $\min(P_{1\to0}) = 95\%$ $\min(P_{1\to0}) = 97\%$
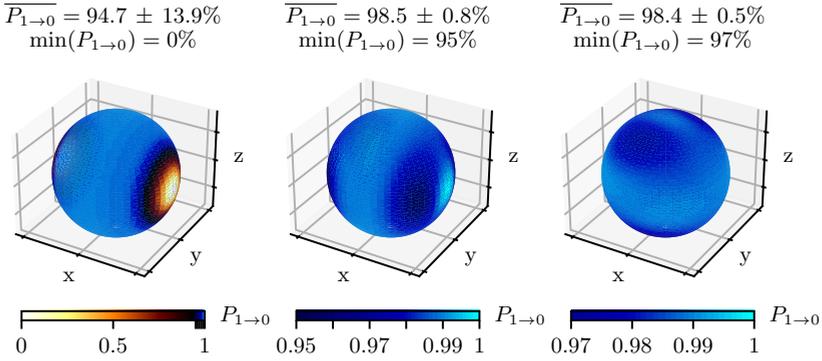
Figure 2.4:   Probabilities $P_{1\to0}$ of transitioning from $|1\rangle$ to $|0\rangle$ after three iterations of the weak-coupling ($t\,\epsilon = 10$) cooling procedure, with coupling potentials $V_S^i = X, X, X$ (left), $V_S^i = X, Y, X$ (center), $V_S^i = X, Y, Z$ (right), on a system qubit with Hamiltonian $H_S = h\,\vec{n}\cdot\vec{\sigma}$ and known energy splitting h. The orientation of the unit vector $\vec{n}$ is represented on spherical surfaces. The average, standard deviation and minimum of $P_{1\to0}$ are shown above each panel.

sphere), $2h$ is a fixed energy splitting, and σ is the vector of Pauli-matrices. For a fixed coupling operator V_S, there is a risk that $[H_S, V_S] \approx 0$. When this is the case, the off-diagonal elements of V_S in the system eigenbasis are zero, preventing cooling. We may prevent this by alternating between different coupling terms during the cooling protocol, such that no non-trivial Hamiltonian can commute with all such coupling terms. This may be achieved for the $1 + 1$ model by iterating over $V_S^i \in \{X_S, Y_S, Z_S\}$, or alternatively over $V_S^i \in \{X_S, Z_S\}$. The effectiveness of this scheme is studied in Fig. 2.4 for resonant coupling. We see the probability $P_{1\to0}$ of successful cooling of the weak coupling approach ($t\,\epsilon = 10$) increased to $\min(P_{1\to0}) = 97\%$ for all choices of $\vec{n}$ when iterating over $V_S^i = X_S, Y_S, Z_S$, and above 95% when iterating $V_S^i = X_S, Z_S, X_S$, compared to the possibility for complete cooling failure $[\min(P_{1\to0}) = 0]$ when V_S^i is held constant. Similar results ar observed for off-resonant coupling.

2.4 Scalable QDC protocols

We now look to use the insight obtained for cooling in the 1+1 toy model to develop QDC schemes for larger systems. The sub-additivity of entropy places a rough lower bound on the number of cooling steps required to cool an N-qubit system. This limits the entropy ΔS_{S} that the system can transfer to the fridge qubit before the non-unitary reset to $\Delta S_{\mathrm{S}} \geq -\Delta S_B \geq -1$ bit. If we demand the ability to cool an arbitrary state, a QDC protocol must also be able to cool the maximally-mixed state, which has entropy $S_{\mathrm{S}} = N$. We then require N repetitions of an optimal coupling-and-reset step to reach the pure ground state (which has entropy $S_{\mathrm{S}} = 0$). This can be obtained in the simple example of cooling N non-interacting qubits with known energies, by simply repeating the protocols of the $1+1$ model. However, this cannot be generalised to arbitrary strongly-correlated systems, as cooling is complicated by irregular and unknown energy gaps and coupling terms between eigenstates. This is to be expected, as preparing ground states of arbitrary Hamiltonians is a known QMA-hard problem [30]. However, as cooling is a physically-motivated process, we hope QDC to be able to achieve polynomial scalings for systems of physical interest, i.e. models of systems that are found in low-temperature equilibrium states in nature. We focus for the rest of this chapter on exploring this thesis.

In the rest of this text, we introduce two scalable QDC protocols for N-qubit systems: the strong-coupling-based BangBang protocol and the weak-coupling-based LogSweep protocol. These extend and generalize the two approaches we established for the 1+1 toy model of section 2.3.1. Each protocol iterates over a sequence of *cooling steps*, each of which consists of coupling the fridge qubit to part of the system for a short time evolution, and then resetting the fridge qubit to its ground state. The protocols differ in the choice of coupling strengths γ_i, coupling terms V_S^i and fridge energies ϵ_i at each i-th cooling step. [The coupling time for each cooling step is fixed by Eq. (2.5)].

2.4.1 The BangBang protocol

We now develop a protocol to extend the strong-coupling approach (Eq. 2.6) to a larger system. This motiviation is in line with recently proposed bang-bang approaches to adiabatic state preparation [98, 99], which are known to outperform initial theoretical expectations stemming from a naive Trotter error estimate. We are thus optimistic that this 'BangBang' protocol may provide a low-cost, near-term method for QDC. However,

such a protocol needs to associate each fridge-system coupling with a single fridge energy, that should match the transitions that this coupling triggers. An appropriate choice of fridge energy can be estimated via a perturbation theory approximation. To derive this approximation, we note that the rate of a transition between eigenstates $|E_i\rangle \to |E_j\rangle$ depends on the matrix element of the coupling V_S:

$$V_{(ij)} := \langle E_i| V_S |E_j\rangle = \frac{\langle E_i| [H_{\rm S}, V_S] |E_j\rangle}{E_i - E_j}. \tag{2.9}$$

If V_S is local and bounded, $[H_{\rm S}, V_S]$ is as well, so the matrix element $V_{(ij)}$ will be bounded proportionally to $(E_i - E_j)^{-1}$. The matrix element is additionally bounded by $\|V\|$; this second bound becomes dominant when $E_i - E_j/\|V\|$ falls below the maximum off-diagonal element of $[H, V]$ in any basis, which we define with the notation $\|[H, V]\|_\perp$:

$$\|O\|_\perp = \max_{\langle\phi|\psi\rangle=0} |\langle\phi| O |\psi\rangle| = \max_{|\Phi\rangle,|\Psi\rangle} \frac{\langle\Phi| O |\Phi\rangle - \langle\Psi| O |\Psi\rangle}{2}, \tag{2.10}$$

where O is Hermitian and the maxima are taken over all possible states $|\psi\rangle, |\phi\rangle$ and $|\Psi\rangle, |\Phi\rangle$. A simple proof is given in Appendix 2.A. We use this energy scale to set the fridge energy:

$$\epsilon_i = \|[V_{\rm S}^i, H_{\rm S}]\|_\perp \tag{2.11}$$

for any coupling potential $V_{\rm S}^i$. This way, the maximum-energy transitions accessible by V_S are on resonance, while smaller energy ones (which are less important for cooling) still have a higher probability of cooling than of reheating [see Fig. 2.3(a)]. This defines the BangBang protocol: we iterate over coupling to each qubit, with ϵ_i fixed by Eq. (2.11). As this protocol does not attempt to suppress reheating, we choose a single coupling $V_S = Y_n$ for the n-th qubit, instead of iterating over $V_S = X_n, Y_n, Z_n$ (as was suggested in Sec. 2.3.2). In general, the best choice of V_S will depend on the system that we want to cool, and the couplings should be picked to enable as many transitions as possible. We repeat the coupling to each qubit R times, resulting in a total of RN cooling steps. Each cooling step contains two first-order Trotter steps simulating $e^{-iH_{\rm C}t/2}$ (of depth $d_{H_{\rm C}}^{(1)}$), a single second-order Trotter step for $e^{-iH_{\rm S}t}$ (of depth $d_{H_{\rm S}}^{(2)}$), and a reset gate, resulting in a total circuit depth

$$d_{\rm BangBang} = RN(2d_{H_{\rm C}}^{(1)} + d_{H_{\rm S}}^{(2)} + 1). \tag{2.12}$$
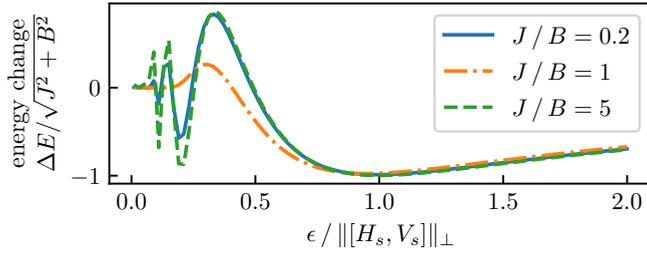
Figure 2.5:   Change in energy expectation value for the application of a single cooling step to the maximally mixed state of a $N = 8$ qubit transverse field Ising chain Eq. (2.13), depending on the fridge energy ϵ. The coupling potential is $V_{\mathrm{S}} = Y_3$, the Pauli Y on the third qubit. The relation $B^2 + J^2 = 1$ fixes the energy scale.

To test the BangBang protocol, we study the cooling of a N-qubit transverse-field Ising chain

$$H_{\mathrm{S}} = \sum_{i=0}^{N} B X_i + \sum_{i=0}^{N-1} J Z_i Z_{i+1}, \qquad (2.13)$$

where B represents the transverse magnetic field Zeeman splitting and J is the Ising coupling strength. The relative coupling strength J/B dictates whether the system is in the paramagnetic ($J/B \ll 1$), ferromagnetic ($J/B \gg 1$), or critical ($J/B \sim 1$) phases. This ability to simply tune between three phases of matter with significantly different physical properties make the TFIM a good benchmark model to investigate the ability of different QDC schemes in various scenarios.

We first demonstrate that our choice for the fridge energy Eq. (2.11) is appropriate. In Fig. 2.5, we plot the effect of a single cooling step on the maximally-mixed state. We observe that cooling is maximized for fridge energies around the point defined by Eq. (2.11), for all phases of the TFIM. We find this behaviour to hold for all other (local) choices of coupling potential V_S tested in our work, as predicted.

We next turn to the ability of the BangBang protocol to prepare an approximation ρ of the ground state, starting from a maximally-mixed (i.e. infinite temperature) initial state. We benchmark by the final state

with two figures of merit: the ground state fidelity

$$F = \mathrm{Tr}\left[\,|E_0\rangle\langle E_0|\,\rho\,\right], \qquad (2.14)$$

and the energy relative to the ground state energy $Tr[H_\mathrm{S}\rho]/|E_\mathrm{GS}|$. This last property is local in local system, and represents an energy density in TFIM. To verify convergence, we compare cooling results to a *reheating limit*, obtained by running the protocol with the ground state as initial state. We observe that all TFIM phases converge after $R \approx N$ repetitions (with the weakly-coupled phase system converging already at the first repetition). In Fig. 2.6 we plot the energy density of the cooled state, as well as the reheating limit, as a function of the number of sites in the system. This shows that convergence is indeed achieved for $R = N$ independently on the system phase and size, and that the final energy density stays approximately constant, without showing any other trend. The BangBang protocol achieves a final energy density close to 90% and 95% of $\|H_\mathrm{S}\|_\perp$ for the ferromagnetic and paramagnetic regime respectively, while performing significantly worse in the critical regime. This is to be expected, as in this regime the spectrum is no longer banded, and excitation energies are not as uniform as in the paramagnetic or ferromagnetic regimes. Following Eq. (2.12), the protocol's circuit depth is $7NR$ for a gate-set containing single- and double-qubit rotations (and the reset gate). Given the low cost of the protocol, we suggest that this is of particular interest for near-term experiments, and may be further refined by other cooling protocols, or methods such as quantum phase estimation, in the long term.

2.4.2 The LogSweep protocol

Refrigeration at weak-coupling suppresses reheating, but only allows for the cooling of transitions within a narrow energy band [as shown in Fig. 2.3(a)]. We may take advantage of this in a larger system, where a wide range of energy gaps are present, by sweeping the fridge energy ϵ_k from high to low as we iterate over cooling steps. (As low-energy transitions are more susceptible to re-heating than high-energy transitions, this will in general be more efficient than sweeping from low to high.)

To construct a full protocol, we further need to fix the set of fridge energies ϵ_k and linewidths $\delta_k = t_k^{-1} = \pi\,\gamma_k$ that we plan to use for each cooling step. We will be guided by two principles. First, the target band of fridge energies $(E_\mathrm{min}, E_\mathrm{max})$ should be tightly covered by the cooling lines $\epsilon_k \pm \delta_k$. Second, the reheating should be suppressed to a certain degree throughout the protocol. As by Eq. (2.4) the reheating suppression
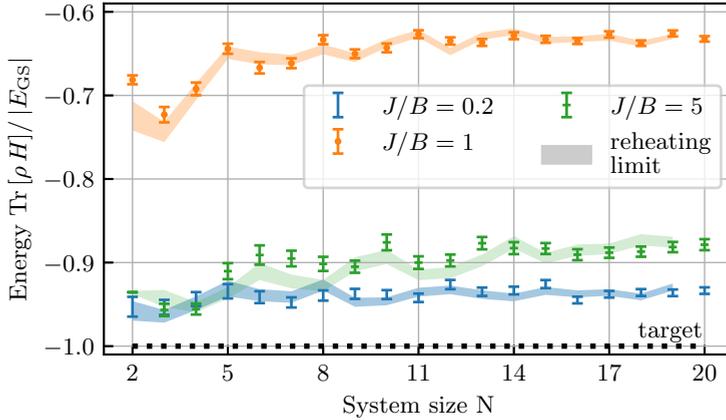
Figure 2.6: Performance of the BangBang protocol as a function of the system size N for the three different phases of the transverse-field Ising model (detailed in legend). The coupling potentials are $V_S^i = Y_i$. Dots correspond to result when the protocol is applied to the maximally-mixed state, shaded regions corresponds to result when protocol is applied to the true ground state (which gives a bound on protocol re-heating). Data generated by Trotterized wave-function simulations of the protocol, and random sampling of the initial mixed state and of nonunitary operations (details in App. 2.F). All points are run with 200 samples, and average results are plotted with the sampling error.

depends on γ_k/ϵ_k, we fix this value to a small constant throughout the protocol (i.e. we choose $\gamma_k \propto \epsilon_k$). Thus we define the LogSweep protocol, where the fridge energy ϵ_k is swept over $(E_{\min}, E_{\max})$ in a logarithmic gradation. Specifically, given the *gradation number* K, we index each cooling step $k = 1, \ldots, K$, and we define

$$\epsilon_k = E_{\min}^{\frac{k-1}{K-1}} E_{\max}^{1-\frac{k-1}{K-1}}, \tag{2.15}$$

and choose δ_k to fix $\epsilon_{k+1}+\delta_{k+1}/\zeta = \epsilon_k-\delta_k/\zeta$, with ζ a constant (potentially dependent on K). In App. 2.B, we prove that such a scheme will cool a single transition in the range $(E_{\min}, E_{\max})$ with probability 1 as $K \to \infty$, and in App. 2.C we demonstrate that the logarithmic gradation is optimal for such a scheme for a choice of $\zeta(K) \sim \log(K)$. To make sure all system excitations have a chance of being dissipated, we further iterate

the couplings V_S over a set of local couplings $\{V_S^i\}$ throughout the system: for the considered spin systems we choose $\{V_S^i\} \equiv \{X_n, Y_n, Z_n\}$ for each qubit n (see Sec.2.3.2), for a total of $3NK$ cooling steps. The number of Trotter steps M_k for each cooling step k is chosen to prevent re-heating. This follows Eq. 2.8, but as transition energies between system eigenstates may be as large as the Hamiltonian spread $2\|H_S\|_\perp$, we set

$$M_k = 2\sqrt{1 + \frac{2\|H_S\|_\perp^2}{\gamma_k^2}}. \tag{2.16}$$

The choice of the fridge energy range $[E_{\min}, E_{\max}]$ will generally depend on heuristics on the system. $E_{\max}$ should be greater or equal than the largest energy of the transitions that we are able to de-excite with the chosen couplings V_S (for local Hamiltonians we can estimate this with the techniques described in 2.4.1). For ground state cooling, $E_{\min}$ should be close to the system ground state gap Δ_{GS}, as no transition with an energy lower than Δ_{GS} can lead from an excited state to the ground state.

We first test the LogSweep protocol as applied to the 1+1 model defined in Sec. 2.3.1, with the system gap Δ now taking an unknown value between $E_{\min}$ and $E_{\max}$ (Fig. 2.7). At each step $k = 1, \ldots, K$ we want to maximise cooling of transitions $\Delta \sim \epsilon_k$, while minimizing reheating of previously-cooled transitions $\Delta \sim \epsilon_{k'}$, $k' < k$. As demonstrated by the black dashed curve in Fig. 2.7, when $E_{\max}/E_{\min} = 5$ this can be achieved well with only $K \approx E_{\max}/E_{\min}$ steps. Note that, to maintain a constant relative linewidth (and thus constant maximum reheating per step), we should scale $K \sim E_{\max}/E_{\min}$. This implies $K \to \infty$ as $E_{\min} \to 0$, in line with the third law of thermodynamics.

In a larger system, the situation is more complex than in the model above. Instead of a single transition from the excited state $|E_1\rangle \to |E_0\rangle$ which occurs with unit probability when $\epsilon = \Delta = E_1 - E_0$, our system may transition to one of many eigenstates $|E_j\rangle$, to each with a transition probability $A_{i,j}$ (assuming a initial state $|E_i\rangle$). As there are many possible target states, the maximum transition probability might be very small ($\max_j A_{i,j} \ll 1$). If we restrict to a single transition $|E_i\rangle \to |E_j\rangle$ with the above reduced cooling rate, one may show that the LogSweep protocol still cools that transition with unit probability as $K \to \infty$, albeit at a rate that scales exponentially in $A_{i,j}$. Luckily, we do not need to ensure any specific transition occurs, instead we may cool sequentially

$$|E_i\rangle \to |E_{j_0}\rangle \to |E_{j_1}\rangle \to \ldots \to |E_0\rangle, \tag{2.17}$$
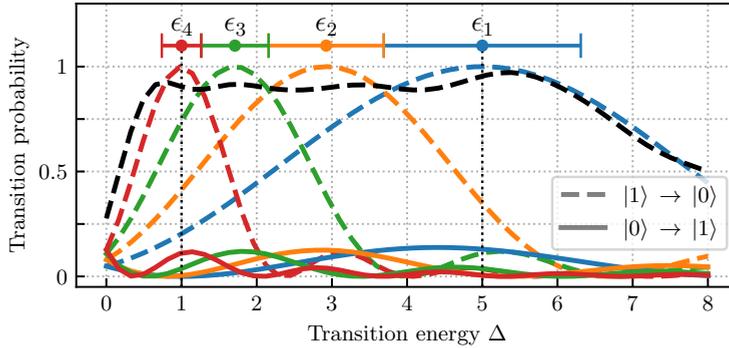
Figure 2.7: Choices of energies ϵ_k and linewidths δ_k (bars at the top of the graph showing $\epsilon_k \pm \delta_k$) for a $K = 4$ LogSweep protocol applied to the model introduced in section 2.3.1 with an unknown $\Delta \in (E_{\min} = 1, E_{\max} = 5)$. Colored lines show cooling (dashed) and reheating (solid lines) probabilities for each j-th step alone, the dashed black line shows the cooling probability after sequential application of the 4 steps.

with a growing number of possible cooling paths as the system grows and the transition probabilities spread over more eigenstates. A good choice of the fridge energy interval $[E_{\min}, E_{\max}]$ and of the coupling potentials $\{V_S^i\}$ allows all eigenstates to be connected to the ground state by sequences of transitions $|E_{j_l}\rangle \rightarrow |E_{j_{l+1}}\rangle$ that have unit probability of being de-excited for $K \rightarrow \infty$. However, a single transition probability approaches 1 only over the entire LogSweep protocol. In particular, if the transition $|E_{j_l}\rangle \rightarrow |E_{j_{l+1}}\rangle$ during step k^* of the protocol corresponds to an energy loss $E_{j_l} - E_{j_{l+1}} \gg \epsilon_{k^*}$, this transition will be off-resonance for the entire remaining duration of the protocol (as $\epsilon_k < \epsilon_{k^*}$ for $k > k^*$), making it unlikely to occur. This can cause convergence issues especially when cooling systems with banded spectra. For such systems , as we set $E_{\min} \approx \Delta_{\mathrm{GS}}$ as detailed above, there may be a point k^* in the protocol after which ϵ_k will become smaller than the average interband gap, but never as small as the spread of a single band. After this point, states at the bottom of a band might transition to states in the lower band, but states at the top of each band never have any resonant transitions to lower energy states, thus becoming absorbing states. This effect is clearly shown in Fig. 2.8, representing the LogSweep-cooled states of the transverse-field Ising model in different regimes. We start with the maximally-mixed state, and plot

the resultant distribution over the eigenstate energies. In the banded regimes (side panels), we observe that the distribution of energies in any given band is tilted towards the higher-energy states in that band (i.e. the aforementioned absorbing states), by some orders of magnitude. This dead-ends ultimately hinder sequential cooling, and prevent the LogSweep cooling from converging to the same state independently on the initial state. The effect worsens as K is increased, as transition linewidths δ_k become smaller making off-resonant transitions less and less probable. This issue can be fixed in practice by using an initial state with fewer high-energy excitation (e.g. a classical approximation of a low-energy state). We solve the issue in principle, by constructing an *iterative LogSweep* protocols, where the LogSweep cooling is repeated with growing K. The early, lower-cost iterations cool the highest energy excitations, while the larger K iterations grant vanishing reheating, and probabilities approaching unity for the cooling transitions allowed by symmetries. Thus, adding iterations with larger and larger K, will make the whole protocol converge to the system ground state (unless symmetries forbid all paths from some states to the ground state). Note that this adjustment is not required for systems with a continuous spectra (i.e. critical systems), as in such a system there will be on-resonance transitions for any state with an energy $E_{\min}$ or more above the ground state.

We now investigate the performance of the (iterative) LogSweep protocol on different phases of the transverse-field Ising model. In Fig. 2.9, we plot the ground state infidelity of the prepared state ρ [$1 - F$ with F as in Eq. (2.14)], as a function of K. The protocol consists in $K - 1$ sweeps of a LogSweep QDC protocol, each sweep having gradation number $g_l = 2, \ldots, K$. The Hamiltonian simulation is performed by second-order Trotter approximation. We investigate the protocol effect on two initial states ρ_0: the maximally-mixed state $\rho_0 = \mathbb{1}/2^n$ to check for cooling capabilities (dots), and the ground state $\rho_0 = |E_0\rangle \langle E_0|$ (crosses) to show the lower bound originated by reheating. We observe polynomial convergence to the ground state in all three phases of the model, attaining an infidelity of $\varepsilon = 1 - F$ in approximately $K \sim O(\varepsilon^{-1/\beta})$ energy gradation steps for $\beta \approx 0.4$-0.8. Additionally, we verify that the protocol converges to the reheating limit for the critical and strongly-coupled regimes. In the weakly-coupled regime instead, although the cooling is far more efficient because of the local nature of the system excitations, the reheating bound is not saturated. We attribute this to the very small linewidths $\{\delta_k\}$, consequence of the well-defined transition energies, together with the strong banding of the system spectrum.

The number of Trotter steps for a single iteration of the LogSweep pro-
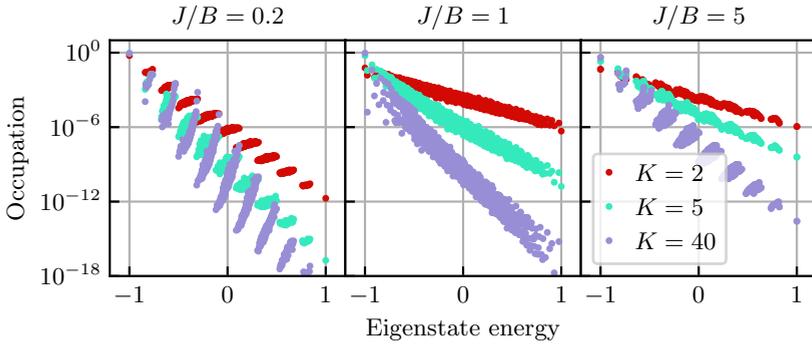
Figure 2.8: Effect of banding on single LogSweep iterations. A maximally mixed state in the three different phases of the 7-qubit TFIM spin chain is evolved by the LogSweep protocol for three different values of K. We plot the distribution of the result here over the system's eigenstates (indexed by energy), at three different values of K. We see that while the critical system demonstrates an approximate thermal or exponential distribution, the weak and strongly-coupled systems demonstrate an inversion in the population of the system within each band, which increases with K. Data generated by continuous-evolution density-matrix simulation (details in App. 2.F).

tocol with gradation number g_l on a system of N spins with Hamiltonian H_S scales as $O\left(\|H_S\|_\perp \Delta_{\mathrm{GS}}^{-1} N g_l^2 \log(g_l)^{-1}\right)$. Thus, the iterative implementation required to deal with the banded cases needs a total number of Trotter steps

$$M_{\mathrm{tot}} \sim O\left(\|H_S\|_\perp \Delta_{\mathrm{GS}}^{-1} N K^3 \log(K)^{-1}\right) \tag{2.18}$$

The gate complexity required to attain an error (infidelity) ε for the models studied scales thus as $O(\varepsilon^{-3})$ - $O(\varepsilon^{-8})$.

We next investigate the scaling of the LogSweep protocol as a function of the system size. In Fig. 2.10 we plot the relative error in the ground state energy as a function of the system size for a single (not iterated) LogSweep with gradation number $K = 5$. We see a constant error in the ground state energy as a function of the system size for the weakly-coupled and critical systems. Thus, here we expect no need to scale K with N for the protocol to be accurate. Let us also note that the gap
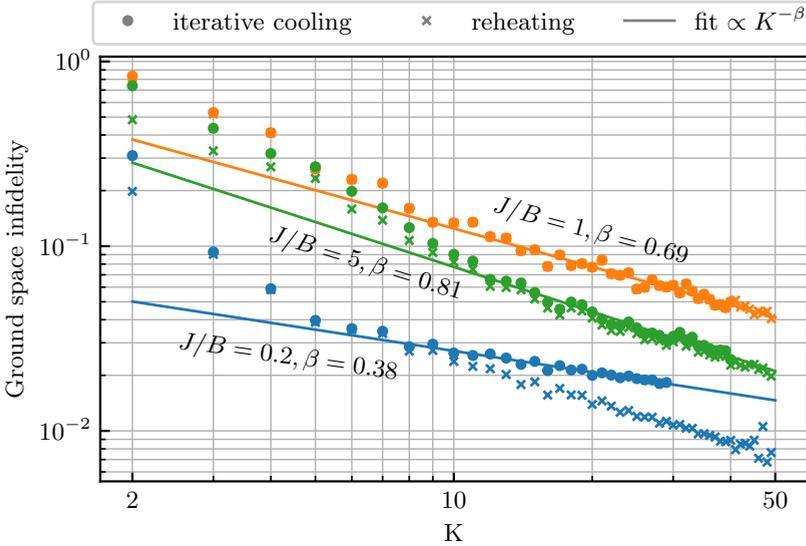
Figure 2.9: Convergence of the LogSweep protocol to the ground state as a function of the gradation number K, starting from the maximally mixed state (dots) and the ground state (crosses), for three phases of the transverse-field Ising model (detailed in legend). Data was generated by deterministic density-matrix simulations of the iterative LogSweep protocol, with second-order Trotter Hamiltonian simulation (details in App. 2.F).

in these two cases shrinks as $\Delta_{\mathrm{GS}}/\|H\| \sim N^{-1}$ and $\Delta_{\mathrm{GS}}/\|H\| \sim N^{-2}$ respectively. Using the above arguments and the estimate (2.18), one can find how the circuit length (in terms of time evolution steps), required to obtain a constant energy error, scales with N. We obtain $O(N^2)$ for the weakly-coupled and $O(N^3)$ for the critical case. From this analysis, we expect that the QDC protocol may be asymptotically competitive with methods such as adiabatic state preparation, whose runtime naively scales as $O(1/\Delta_{\mathrm{GS}}^2)$ [15, 49]. In the strongly-correlated phase, we do not see such positive results; the energy error increases with the system size, though the relative error remains beneath 10% for up to 14 spins. This may be explained by the relative growth of the extension of excitations within the strongly correlated phase, while cooling is performed with local couplings. Due to the error in the simulation, we are unable to reliably extract an
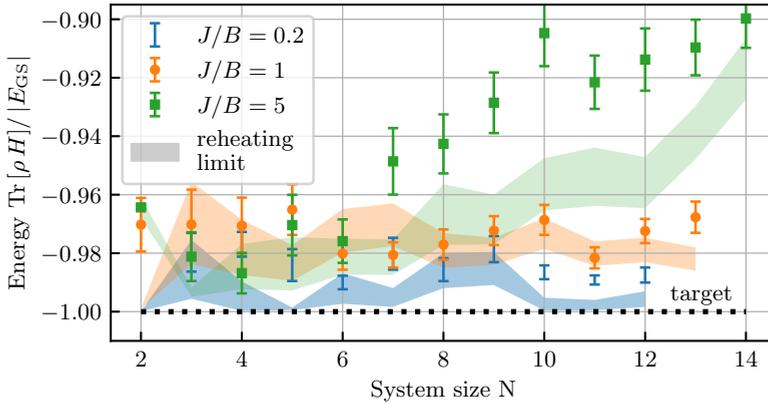
Figure 2.10:   Performance of the LogSweep protocol as a function of the system size for the three different phases of the transverse-field Ising model (detailed in legend), with fixed $K = 5$. Dots correspond to result when protocol is applied to the maximally-mixed state, shaded region corresponds to result when protocol is applied to the true ground state (which gives a bound on protocol re-heating). Data generated by Trotterized wave-function simulations of the protocol, and random sampling of the initial mixed state and of nonunitary operations (details in App. 2.F). All points are run with 100 samples, and average results are plotted with the sampling error.

estimate of the computational cost in the same way as for the critical and weakly-coupled systems. Future work may explore whether this error may be improved on by adjusting the form of the coupling terms $\{V_S^i\}$ based on heuristics on the considered system.

2.5 Conclusion

In this chapter, we investigated how cooling can be simulated on a digital quantum computer, and demonstrated that this can be exploited for the design of scalable algorithms for preparing ground states of N-qubit systems. We identified how one can meet many of the fundamental challenges that the digital approach to cooling raises and use the leverage offered exclusively by digital quantum hardware, namely the freedom of choice in the coupling strength and fridge energy. We laid out a general

approach of simulating a cold bath with a single auxiliary qubit, which is iteratively coupled to various locations in the system and reset periodically to extract entropy and energy. We studied how to digitize the system-fridge coupling simulation without causing additional reheating, and how to avoid symmetries which produce non-ergodic behavior that hinders cooling. By tuning coupling parameters beyond the perturbative regime described by Fermi's golden rule, efficient cooling of targeted transitions can be realized. Following these principles we proposed two protocols for preparing approximate ground states of N-qubit systems — the BangBang protocol and the LogSweep protocol. We studied numerically how these protocols perform on the three phases of the 1D transverse-field Ising model. We found that the BangBang protocol quickly cools the system near to the ground state in the paramagnetic and in the ferromagnetic regime, but has difficulty in the critical regime. The LogSweep protocol is observed to cool all three phases to the ground state at a polynomial cost in the overlap error. In the weakly-coupled and critical phases, the LogSweep protocol further demonstrates a constant energy error as a function of the system size (for fixed gradation number), making it a competitive state preparation method.

Thanks to the low number of steps required, we believe the BangBang protocol has the potential of finding a near-term application, specially if integrated with projective or variational methods to improve its performance. The requirement of precise Hamiltonian simulation and multiple cooling steps makes the LogSweep algorithm in its current form unsuitable for near-term implementation. Nevertheless, the scaling arguments discussed at the end of section 2.4.2 show that our algorithm can be competitive with other non-NISQ methods such as projective quantum phase estimation (QPE) and adiabatic state preparation. QPE in its standard form requires multiple auxiliary qubits and an initial state with a finite ground state overlap, while as our has no requirement on the initial state. Adiabatic state preparation requires an integrable Hamiltonian which can be adiabatically connected to the required Hamiltonian, and requires time scaling as $O(1/\Delta_{\min}^2)$ [15, 49] where $\Delta_{\min} < \Delta_{\mathrm{GS}}$ is the minimum gap along the adiabatic path.

The introduction of quantum digital cooling opens future research directions related to the characterization of proposed protocols, their optimization, and their extension beyond ground state preparation. A study of the effect of noise on currently proposed QDC protocols, and the optimization of such protocols for noise resilience, are in order to establish their applicability on near-term devices. Applying QDC to more complex physical systems, in areas such as quantum spin liquids, many-body

localization and quantum chemistry, would bring new challenges to the protocol construction. A thorough study of the role in the cooling process played by the symmetries and locality of coupling could lead to the design of more optimized protocols. Furthermore, various extensions to the QDC protocols we proposed can be suggested. In a parallelized version of QDC, the use of multiple fridge qubits coupled to various locations in the system might allow to trade space complexity for time complexity. A variationally-optimized QDC protocol might be devised, that can efficiently prepare a state in the ground state manifold of some Hamiltonian starting from an arbitrary initial state — differently from the variational quantum eigensolver [19] which requires the preparation of a fiducial state at every iteration. The principles of QDC might inspire a new class of efficient non-unitary quantum algorithms, where non-unitary operations are mediated by a single auxiliary qubit, with possible application e.g. in the simulation of open quantum system dynamics.

One application of particular future interest for QDC protocols is in the preparation of Gibbs thermal states, which are useful e.g. for semi-definite programming [100]. This seems especially promising given the near-thermal distribution in Fig. 2.8 of the critical system under the evolution of the LogSweep scheme. However, it is as of yet unclear how to overcome the finite width of the distribution, and how well these protocols behave in the banded case (or for more general systems). Adjustment of the LogSweep protocol to produce robust thermal state preparation schemes is an obvious target for future research.

2.A Proof of Eq. (2.10)

To prove Eq. (2.10) we first show that

$$|\langle\phi|\,O\,|\psi\rangle| \le \max_{|\Phi\rangle,|\Psi\rangle} \frac{\langle\Phi|\,O\,|\Phi\rangle - \langle\Psi|\,O\,|\Psi\rangle}{2}, \qquad (2.19)$$

for all $|\psi\rangle\,,|\phi\rangle : \langle\psi|\phi\rangle = 0$. We can assume without loss of generality $\langle\phi|\,O\,|\psi\rangle$ is real and nonnegative (if it's not, we can multiply one state by an irrelevant global phase), and drop the absolute value. As $\langle\psi|\phi\rangle = 0$ we

can define the states $|\pm\rangle = \frac{|\phi\rangle \pm |\psi\rangle}{\sqrt{2}}$ we can then write

$$
\begin{aligned}
\langle\phi|\,O\,|\psi\rangle &= \frac{1}{2}(\langle\phi|\,O\,|\psi\rangle + \langle\psi|\,O\,|\phi\rangle) \\
&= \frac{\langle+|\,O\,|+\rangle - \langle-|\,O\,|-\rangle}{2}
\end{aligned}
$$

immediately proving Eq. (2.19). The opposite inequality is proven by noticing that the $|\Psi\rangle$ and $|\Phi\rangle$ that maximize the right of Eq. (2.10) have to be eigenvalues (by the variational principle). With these, we can redefine the states $|\pm\rangle = \frac{|\Phi\rangle \pm |\Psi\rangle}{\sqrt{2}}$ which are also granted to be orthogonal, thus

$$
\begin{aligned}
\frac{\langle\Phi|\,O\,|\Phi\rangle - \langle\Psi|\,O\,|\Psi\rangle}{2} &= \mathrm{Re}[\langle+|\,O\,|-\rangle] \\
&\leq |\langle+|\,O\,|-\rangle| \\
&\leq \max_{\langle\phi|\psi\rangle=0} |\langle\phi|\,O\,|\psi\rangle|
\end{aligned}
$$

which combined with Eq. (2.19) proves Eq. (2.10).

2.B Asymptotic reheating and cooling probabilities for QDC protocols

Let us consider a two-state subsystem of a larger Hilbert space with a gap energy E, evolving under a QDC protocol on the kth step via a coupling term that does not mix the $\{|01\rangle, |10\rangle\}$ and $\{|00\rangle, |11\rangle\}$ subspaces (where the second index denotes the fridge). Under this assumption, the evolution of the system within this space is a Markov process. Following the main text, let the fridge energy on the kth step be ϵ_k, the coupling strength be γ_k, and the time evolved for in the cooling protocol t_k. Additionally, let the spacing of the fridge energies to be

$$
(\epsilon_k - \epsilon_{k+1}) = \zeta(\delta_k + \delta_{k+1}) = \frac{\alpha}{2}(\gamma_k + \gamma_{k+1}),
$$

for some K-dependent $\alpha = \alpha(K) = \frac{2}{\pi\zeta(K)}$. We may calculate the transition matrix for the Markov process, $p^{(k)}(E)$ (defined by $p_{a,b}^{(k)}(E) = P(|a\rangle \to |b\rangle)$

in a single cooling step) as

$$p^{(k)}(E) = \begin{pmatrix} 1 - \sin^2(\Omega_k t_k/2)\frac{\gamma_k^2}{\Omega_k^2} & \sin^2(\omega_k t_k/2)\frac{\gamma_k^2}{\omega_k^2} \\ \sin^2(\Omega_k t_k/2)\frac{\gamma_k^2}{\Omega_k^2} & 1 - \sin^2(\omega_k t_k/2)\frac{\gamma_k^2}{\omega_k^2} \end{pmatrix}, \qquad (2.20)$$

where

$$\omega_k = \sqrt{(E - \epsilon_k)^2 + \gamma_k^2} \qquad (2.21)$$

$$\Omega_k = \sqrt{(E + \epsilon_k)^2 + \gamma_k^2}. \qquad (2.22)$$

Assuming no additional cooling or heating to the rest of the system during the protocol, the transition matrix for the $k_0 \to k_1$ block takes the form

$$P_{k_0,k_1}(E) = \prod_{k=k_0}^{k_1} p^{(k)}(E), \qquad (2.23)$$

and the transition matrix for the entire process may be written $P(E) = P_{1,K}(E)$.

Exact analytic evaluation of this expression in the large K limit is quite difficult. Instead, we aim for a conservative estimate, bounding the final cooling probability $p_c = [P(E)]_{01}$ from below. For this, given the energy E, we first lower bound the 'initial' cooling around the resonant step k_c, i.e. such k_c that $|\epsilon_{k_c} - E|$ is minimal. Then we give an upper bound on reheating during the following protocol steps $k = k_c, ..K$. Given the estimated cooling probability $p_c^{(k_c)}$ and reheating probability $p_{rh}^{(k_c;K)}$, we can obtain a lower bound for p_c:

$$p_c > (1 - p_{rh}^{(k_c;K)})p_c^{(k_c)} \qquad (2.24)$$

The value of $p_c^{(k_c)}$ can be conservatively estimated from the formula:

$$1 - p_c^{(k_c)} < \prod_{k=1}^{K}(1 - \sin^2(\omega_k t_k/2)\frac{\gamma_k^2}{\omega_k^2}) \qquad (2.25)$$

$$< \prod_{k, \frac{|E-\epsilon_k|}{\gamma_k}<1}((E - \epsilon_k)^2/\gamma_k^2), \qquad (2.26)$$

where the second line follows from the inequality $\sin\left(\frac{\pi\sqrt{1+x^2}}{2}\right)/(1 + x^2) \geq$

min$(0, 1-x^2)$ applied to each term in the product. In the perfect resonance scenario, $|E - \epsilon_{k_c}| = 0$ and the cooling probability is exactly 1. The worst case scenario is when E is right between the two neighbouring ϵ_k's, thus $|E - \epsilon_{k_c}| = \frac{\alpha}{2}\gamma_k$. In this case, we can calculate the logarithm of (2.26) in the leading order of K^{-1}, α:

$$2 \sum_{k=k_c^{(-)}}^{k_c^{(+)}} \log \left| \frac{\epsilon_k - E}{\gamma_k} \right| = 2 \int_{\epsilon^{(-)}}^{\epsilon^{(+)}} \log \left| \frac{\epsilon - E}{\gamma(\epsilon)} \right| d\epsilon \frac{dk}{d\epsilon} \tag{2.27}$$

$$= \frac{2}{\alpha} \int_{\epsilon^{(-)}}^{\epsilon^{(+)}} \log \left| \frac{\epsilon - E}{\gamma(\epsilon)} \right| \frac{d\epsilon}{\gamma(\epsilon)}. \tag{2.28}$$

Here, we used the fact that $\gamma\alpha$ defines energy spacing (and so $\frac{d\epsilon}{dk} = \alpha\gamma(\epsilon)$), and introduced summation limits $k_c^{(\pm)}$, $\epsilon^{(\pm)}$ as the points where $\frac{\epsilon - E}{\gamma} = \pm 1$. As this implies scaling $\epsilon^{(\pm)} = E + O(\gamma)$, (2.28) should scale as $O(1/\alpha)$. The calculation can be completed for the LogSweep gradation ϵ_k, γ_k, which implies $\epsilon_k' \propto \gamma(\epsilon) \propto \epsilon$. In particular, if $x = \frac{\epsilon - E}{\gamma}$ then $dx = \frac{E d\epsilon}{\epsilon\gamma} = \frac{d\epsilon}{\gamma}(1 + O(1/K))$, and we have:

$$\frac{2}{\alpha} \int_{\epsilon^{(-)}}^{\epsilon^{(+)}} \log \left| \frac{\epsilon - E}{\gamma(\epsilon)} \right| \frac{d\epsilon}{\gamma(\epsilon)} = \frac{4}{\alpha} \int_0^1 \log x \, dx = -\frac{4}{\alpha}. \tag{2.29}$$

Substituting into Eq. 2.26, we find the initial cooling probability bounded by

$$p_c^{(k_c)} \gtrsim 1 - \exp(-4/\alpha(K)). \tag{2.30}$$

The reheating accumulated between steps k_c and K, $p_{rh}^{(k_c;K)}$, can be upper bounded as:

$$p_{rh}^{(k_c;K)} \leq 1 - \prod_{k=k_c}^{K} \left(1 - \sin^2(\Omega_k t_k/2) \frac{\gamma_k^2}{\Omega_k^2} \right) \tag{2.31}$$

The product in Eq. (2.31) can be further estimated as:

$$\prod_{k=k_c}^{K} \left(1 - \sin^2(\Omega_k t_k/2) \frac{\gamma_k^2}{\Omega_k^2} \right) \geq \prod_{k=k_c}^{K} \left(1 - \frac{\gamma_k^2}{\Omega_k^2} \right) \tag{2.32}$$

$$\geq \prod_{k=k_c}^{K} \left(1 - \frac{\gamma_k^2}{(E + \epsilon_k)^2} \right) \tag{2.33}$$

$$\simeq \exp\left(- \sum_{k=k_c}^{K} \frac{\gamma_k^2}{(E + \epsilon_k)^2} \right), \tag{2.34}$$

where in the last line we assumed that $\gamma_k \ll E + \epsilon_k$ for all k. As we are most concerned about the large K asymptotics of the total cooling probability, let us now analyze how the expression (2.34) behaves in this limit. Since γ_k^2 scales as $O(1/K^2)$ and we have K terms in the sum, we generally expect $O(1/K)$ scaling for the sum. Such scaling would imply a rapidly vanishing reheating for a large-K protocol. In the specific case of the LogSweep protocol, to the leading order in $1/K$ one indeed obtains:

$$p_{rh}^{(k_c;K)} \lesssim \sum_{k=k_c}^{K} \frac{\gamma_k^2}{(E + \epsilon_k)^2} \approx \frac{1}{\alpha(K)} \int_{E_{\min}}^{E} \frac{\gamma(\epsilon)}{(E + \epsilon)^2} d\epsilon \tag{2.35}$$

$$\approx \frac{\log \frac{E_{\max}}{E_{\min}}}{\alpha^2(K)K} \left(\frac{1}{2} - \frac{E}{E + E_{\min}} + \log\left(\frac{2E}{E + E_{\min}} \right) \right) \tag{2.36}$$

$$\equiv \frac{R(E_{\min}, E_{\max}, E)}{\alpha^2(K)K}. \tag{2.37}$$

Here, we used Eq. (2.15) and the fact that $\alpha(K)\gamma_k$ defines energy spacing $|\epsilon_{k+1} - \epsilon_k|$. Finally, combining Eqq. (2.24) - (2.37), we obtain an asymptotic lower bound to the final cooling probability:

$$p_c = \left(1 - \exp\left(-\frac{4}{\alpha(K)} \right) \right) \cdot \left(1 - \frac{R(E_{\min}, E_{\max}, E)}{\alpha^2(K)K} \right). \tag{2.38}$$

This estimate implies $p_c \to 1$ for large K, provided that both $e^{-4\alpha^{-1}(K)} \to 0$ and $\frac{1}{K\alpha^2(K)} \to 0$.

To ensure that the infidelity is minimized and thus $\alpha(K)$ is optimal, we solve the extremum condition $\partial_\alpha(e^{-4\alpha^{-1}} + \frac{R}{\alpha^2 K}) = 0$ for α. The solution can be expressed in terms of the product logarithm function W, $\alpha(K) = 4\,W^{-1}\,(8K/R)$. For large K, at the leading order we obtain simply:

$\alpha(K) = 4 \, \log^{-1}(8K/R)$. The infidelity then scales down almost linearly with K: $1 - p_c = \frac{\log^2(8K/R)}{16K}$. This asymptotically optimal $\alpha(K)$ yields the choice $\zeta(K) = \frac{1}{2\pi} \log(8K/R)$, which we use in all our simulations.

2.C Optimizing energy spacing in LogSweep protocol

In Sec. 2.4.2, we argued that the energy spacing of the LogSweep protocol is optimal for the protocol precision for a K-step protocol. This was based on the reheating estimate taken from the cooling step k_c only. One may ask, if this persists when one includes the total reheating into account. In the large K limit, we can use the estimate (2.34) for this check. Fixing the constraint $\gamma_k = \frac{|\epsilon_{k+1} - \epsilon_k|}{\alpha}$, we proceed by means of variational calculus:

$$\frac{\delta}{\delta \epsilon_k} \sum_{k=k_c}^{K} \frac{\gamma_k^2}{(E + \epsilon_k)^2} = 0 \tag{2.39}$$

$$\Rightarrow \frac{\delta}{\delta \epsilon(k)} \int_{k_c}^{K} \frac{(\epsilon'(k))^2}{(E + \epsilon(k))^2} dk = 0 \tag{2.40}$$

$$\Rightarrow \epsilon''(k) \cdot (E + \epsilon(k)) = (\epsilon'(k))^2. \tag{2.41}$$

The solution to Eq. (2.41) that satisfies boundary conditions $\epsilon(k_c) = E$, $\epsilon(K) = E_{\min}$, is as follows:

$$\epsilon_k = (2E)^{\frac{K-k}{K-k_c}} (E + E_{\min})^{\frac{k-K}{K-k_c}+1} - E. \tag{2.42}$$

This shows that the logarithmic character of the optimal spacing persists when we consider total reheating (cf. Eq. (2.15)). However, we cannot directly use the embelished result (2.42) for our cooling protocol. That is because this formula uses the targeted energy E as a reference, whereas we are targetting a continuum of energies. Therefore, we keep using the simpler and more practical formula Eq. (2.15) for the LogSweep protocol.

2.D Cooling rate for LogSweep protocol in a large system

In a large system, the above analysis is complicated by the presence of multiple transitions from every energy level. We now give a simplified

61

analysis that focuses on a pair of states $|E_i\rangle$, $|E_j\rangle$, in a spirit similar to Appendix 2.B. This means we formulate the protocol as a Markov process equivalent to (Eq. 2.20), where the transitions to levels other than i and j are ignored. Note that in the perturbative limit, this is a good approximation of the actual Markov process as restricted onto the subspace $|E_i\rangle$, $|E_j\rangle$. Specifically, even though we ignore the indirect transitions between i and j via other levels, this is justified at the first order of pertubation theory. Unlike in the $1 + 1$ model however, the transitions here are imperfect. If our total coupling has strength γ (i.e. $\|H_C\| = 2^N \gamma$), following the analysis in Sec. 2.4.1 the coupling between states $|E_i\rangle$ and $|E_j\rangle$ will take the form $\gamma \sqrt{A_{i,j}}$ with $\sqrt{A_{i,j}}$ scaling down as $O((E_i - E_j)^{-2})$. This has the effect of scaling both the cooling and re-heating rates by $A_{i,j}$, recasting the Markov process (Eq. 2.20) as

$$
p_{i,j}^{(k)} = \begin{pmatrix} 1 - A_{i,j} \sin^2(\frac{\Omega_k t_k}{2})\frac{\gamma_k^2}{\Omega_k^2} & A_{i,j} \sin^2(\frac{\omega_k t_k}{2})\frac{\gamma_k^2}{\omega_k^2} \\ A_{i,j} \sin^2(\frac{\Omega_k t_k}{2})\frac{\gamma_k^2}{\Omega_k^2} & 1 - A_{i,j} \sin^2(\frac{\omega_k t_k}{2})\frac{\gamma_k^2}{\omega_k^2} \end{pmatrix}.
$$

As this only reduces both the heating and cooling rates, our claim that reheating in the LogSweep protocol tends to 0 as $K \to \infty$ still holds. However, we need to repeat the analysis of App. 2.B to bound the cooling rate $p_c^{(k_c)}$ below and check that it continues to tend to 1. For the sake of generality, we drop the i, j indices, and consider a cooling probability restricted by a k-independent factor A.

With this adjustment, we may recast Eq.2.26 when $A << 1$ as

$$
1 - p_c^{k_c} < \prod_{k,\, \frac{|E-\epsilon_k|}{\gamma_k} < 1} \left[\left(1 - \frac{A\pi^2}{4} \right) + \frac{A\pi^4}{48} \frac{(E-\epsilon_k)^2}{\gamma_k^2} \right]. \tag{2.43}
$$

Then, taking the log and converting again to an integral, we obtain

$$
\log\left(1 - p_c^{k_c} \right) < \frac{1}{\alpha} \int_{\epsilon^-}^{\epsilon^+} \log\left[B + A' \frac{(E-\epsilon)^2}{\gamma(E)^2} \right] \frac{d\epsilon}{\gamma(\epsilon)}, \tag{2.44}
$$

where $A' = \frac{A\pi^4}{48} \sim 2A$, and $B = 1 - \frac{A\pi^2}{4} < 1$. Next, setting $x = \frac{E-\epsilon}{\gamma(\epsilon)}$, and using the fact that for the LogSweep protocol $\gamma(\epsilon) \sim \epsilon$, we find

$$
\log\left(1 - p_c^{k_c} \right) < \frac{2}{\alpha} \int_{-1}^{+1} \log\left(B + A' x^2 \right) dx. \tag{2.45}
$$

This may be evaluated by integrating by parts, giving

$$\log\left(1 - p_c^{k_c}\right) < -\frac{2}{\alpha} \int_{-1}^{+1} \frac{x^2}{BA'^{-1} + x^2} \tag{2.46}$$

$$= \frac{-4}{\alpha} \left[1 - BA'^{-1} \tan^{-1}\left(A'B^{-1}\right)\right] \tag{2.47}$$

$$\sim -\frac{4}{3\alpha} A'^2 B^{-2} + O(A^4). \tag{2.48}$$

Using the optimal scaling $\alpha(K) = 4\log^{-1}(K)$ we identified in Appendix 2.B, this adjusts our bound in the cooling rate to

$$p_c^{k_c} \gtrsim 1 - K^{-\frac{1}{3}A'^2 B^{-2}}, \tag{2.49}$$

which continues to tend to 1 as $K \to \infty$, albiet at a rate reduced proportional to A.

This result requires some consideration in a large system — if our coupling Γ from a state $|E_i\rangle$ is spread over transitions to J states $|E_j\rangle$, we have $A_{i,j} \sim J^{-1}$, and the probability of any transition being cooled can be found to be

$$\prod_j (1 - p_{c,j}^{k_c}) \sim e^{-\sum_{j=1}^{J} \frac{1}{3\alpha} A_j^2 (1 - A_j^2)} \sim e^{-\frac{1}{3\alpha J}}. \tag{2.50}$$

This implies that we require $\alpha \sim J^{-1}$ in order to maintain a constant cooling rate, which in turn may require adjustments to the optimal scaling identified in Appendix 2.B. As such adjustments are highly system-dependent, we do not investigate them further here.

2.E Effect of banding on QDC protocols

In this appendix we demonstrate the effect of banding on single sweeps of the LogSweep protocol. In Fig. 2.11, we plot the infidelity of a single shot of the LogSweep protocol with gradation number K acting on the maximally-mixed state, as a function of K (triangular markers). We see that in the critical case, the system continues to tend to the ground state polynomially in K. However, for the TFIM chain in the weakly- and strongly-coupled phases, we find that the protocol fails to converge as a function of K, due to the banding issue described above. This lack of convergence is rectified in the series marked by dots (same data as in Fig. 2.9) by repeating the LogSweep protocol as a function of K. We note
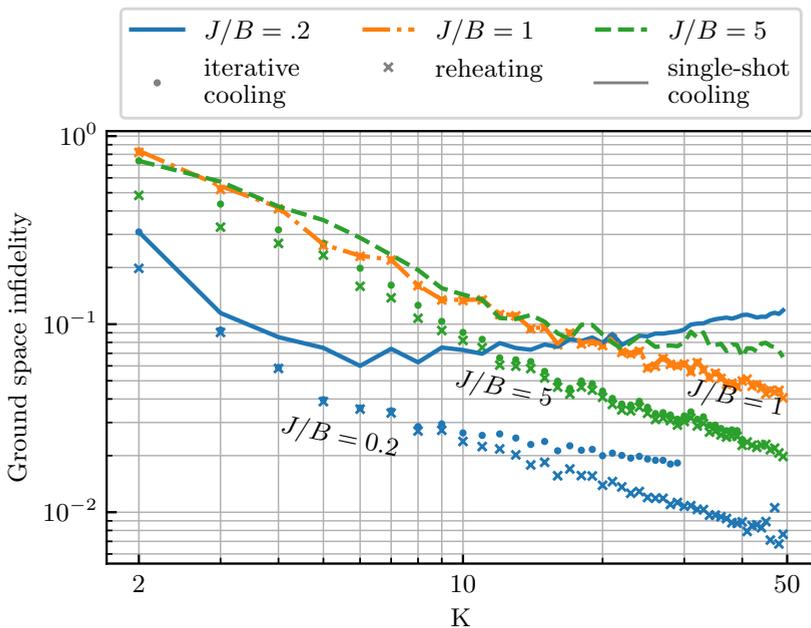
Figure 2.11:   Difference between cooling by applying the a single LogSweep protocol with gradation number K (round markers), and iterating LogSweep for all $g_l = 2, ..., K$ (solid lines). The iterative and reheating data are the same as in Fig. 2.9, the same context and simulation techniques apply.

that the failure in the strong-coupling case is not of the same degree as in the weak-coupling case, which we ascribe to the fact that the banding is not as strongly pronounced in Fig. 2.8, and so the result has not yet presented itself.

2.F Details on numerical methods

In this appendix we report the methods used to simulate QDC protocols on many-qubit systems. The python code is packaged and available on github, at https://github.com/aQaLeiden/QuantumDigitalCooling [101], and makes cirq [102] to build and simulate the required quantum circuits. All the data reported in this chapter, along with more simulations results that

are left out in the interest of space and clarity, can be found in the same repository.

The simulations for the BangBang protocol energy validation Fig. 2.5, as well as the study of LogSweep performance with increasing K (Figures 2.9 and 2.11) were performed using cirq density matrix simulator. This stores the system's state in a density matrix, to which are applied sparse unitaries (representing the circuit's unitary gates) and the eventual quantum channel representing the reset gate. For BangBang, as prescribed by the protocol, the unitary circuit applied before each fridge reset gate is defined by second-order Trotter expansion of the coupled system-fridge Hamiltonian with a single step (Trotter number $M = 1$). This corresponds to Eq. 2.7 where also e^{-iH_S} is substituted by its $M = 1$ second-order Trotter expansion.

To push to a larger number of qubits the results on scaling of both protocols, data for in Figures 2.6 and 2.10 were generated with cirq state vector simulator. The non-unitary reset required by QDC protocols and the initially-mixed state used to benchmark cooling cannot be represented deterministically in state vector simulations. These are instead implemented by random sampling. Each sample is constructed by choosing at random an initial computational basis state (these are enough to sample the maximally mixed state, because of the density matrix equivalence class). For each non-unitary reset gate, the outcome is sampled probabilistically. This process is repeated for 100 samples for each data point. The mean value of the quantity of interest is plotted, together with an interval representing the standard deviation of the mean.

In both the density matrix and the wavefunction simuations performed with cirq, the numerical error causes the final state to be often non normalized. In the worst cases, the deviations from unit L2 norm (for wavefunctions) and trace norm (for density matrices) are up to few parts per thousand and percent respectively. This is attributed to the large number of short-time Trotter steps required to produce LogSweep data in Fig. 2.9, which translate to sparse unitaries with small entries and a building up of numerical error. As all operations are linear, the first-order error can be dealt with by forcing normalization on the final state. This technique was used for results reported in figures 2.9, 2.10 and 2.11

The numerical error of Trotterized sparse-unitary simulations still is too large to show the final state occupations in Fig. 2.8, which range over more than 18 orders of magnitude. For this reason, these simulations were performed by constructing the continuous evolution operator $e^{-i(H_\mathrm{S}+H_\mathrm{F}+H_\mathrm{C})t}$ for each unitary evolution step. These results were validated by comparing with the Trotterized approach the results for large occupations and the energy expectation values (which are less sensitive to the numerical error).

Error mitigation via verified phase estimation

3.1 Introduction

Error mitigation is likely essential for near-term quantum computations to realize valuable applications. State-of-the-art technology in superconducting qubits has recently pushed quantum computers beyond the capability of their classical counterparts [3] and enabled intermediate scale demonstrations of quantum algorithms for optimization [103, 104], quantum chemistry [59, 103, 105], and machine learning [106], with tens of qubits and hundreds of quantum gates. However, these experiments clearly reveal a noise barrier that needs to be overcome if such applications will ever scale to the classically intractable regime. In the long-term, a path towards this goal is known through quantum error correction [107–109]. Yet, the requirements to successfully error correct large-scale quantum applications [14, 71, 73, 110, 111] are still a few orders of magnitude above the current state-of-the-art, and will likely require many years to achieve. In the meantime, quantum applications research has focused on finding the elusive beyond-classical NISQ (noisy, intermediate-scale quantum) application [18], with the hope to accelerate the path to practical quantum computing. However without the resources to correct errors, one must develop strategies to mitigate the aforementioned noise barrier. Otherwise, the output of NISQ devices will be corrupted beyond usefulness for

algorithms significantly more complex than those already attempted.

Much of the attention in the NISQ era has been directed towards variational algorithms, with applications in optimization [55], chemistry and materials science [82], and machine learning [112, 113]. These shift much of the complexity of the algorithm to a classical outer loop involving many circuit repetitions, leaving the quantum computer with the task only of preparing quantum states and estimating expectation values of operators on said states. However, preparation circuits need to have significant depth to avoid being classically simulated [114]. Errors accumulated over this circuit quickly distort the prepared state to one different than was targeted. This has meant that most quantum experiments to date have had difficulty achieving standard accuracy benchmarks prior to applying error mitigation techniques [59, 103, 103, 105, 115]. However, accuracy improvements of orders of magnitude have been achieved with error mitigation in these experiments, suggesting there may yet be hope for NISQ.

The zoo of error mitigation techniques is large and varied. One may first attempt to design algorithms that are naturally noise-robust. For example, the optimization procedure in a variational algorithm makes the algorithm robust against control errors (e.g. over- or under-rotations when gates are applied) [82]. Also, subspace expansions of the variational quantum eigensolver (VQE) in materials science and chemistry correct errors that keep one within the desired subspace considered [116] or more generally through by approximate symmetry projection [117]. Given the ability to artificially introduce additional noise into a device, one can extrapolate from multiple experiments at different noise levels to a hypothetical noiseless experiment [118], which has shown promising results on real devices [119]. One may alternatively probabilistically compile circuits by inserting additional gates to average out or cancel out noise, given sufficient knowledge of the error model of the device [118, 120]. When classically post-processing partial state tomography data from an experiment, one may attempt to regularize the obtained results using reduced density matrix constraints [121]. Finally, one may mitigate errors that take a state outside of a symmetry-conserving subspace of a quantum problem, either by direct post-selection, or artificial projection of the estimated density matrix in post-processing, producing a 'symmetry-verified' state [29, 66, 117, 122]. Recent efforts have extended this protocol by introducing symmetries into problems to increase the range of errors that may be detected [123], which is analogous to the way quantum error correcting codes introduce engineered symmetries.

Ideally, we would prefer to go beyond verifying that a system's state remains within a target subspace and instead directly verify that the system's

state is the one we desire. This would result in reaching the information theoretic optimal limit of post-selected error mitigation in which one could completely mitigate the effect of all errors by repeating the experiment a number of times scaling inversely with the circuit fidelity (equivalent to the ability to perfectly detect errors). The fact that the circuit fidelity is expected to decrease exponentially in the gate complexity indicates that eventually we will still need error-correction; however, moving closer to this limit is certain to enable more powerful NISQ experiments.

In this work we develop a method for error mitigation of quantum phase estimation experiments, by verifying that the system returns to its initial state after the phase estimation step. We show that the set of experiments that pass this condition contain all the necessary information to perform quantum phase estimation. This yields a powerful error mitigation technique, as in most cases errors will not return the system to this initial state. Our techniques apply to variants of phase estimation which might involve post-processing on a single control qubit [124, 125], or when performing recently-developed control-free variants [27, 126]. We further develop it into a simple scheme for verified expectation value estimation by dividing a target Hamiltonian into a sum of fast-forwardable terms. This yields a simple, low-cost scheme for the measurement of expectation values, which may be immediately incorporated into the quantum step of a variational quantum algorithm. We study the mitigation power of this protocol in numerical simulations of small-scale experiments of free-fermion, transverse Ising, and electronic structure Hamiltonians. Verification is observed to mitigate all single (and even all double) errors throughout many of these simulations, as evidenced by a clear second (or third)-order sensitivity in our results to the underlying gate error rate. We observe in the best-case scenario case an up to $10,000$-fold suppression of error at physical error rates; this is not achieved for all systems studied, but verification is found to improve experimental error in all simulations performed. We find the error mitigation power to be highly system-, circuit-, and noise model-dependent. Finally, we study the measurement cost of this protocol in the presence of sampling noise, finding that it is comparable to standard partial state tomography techniques for energy estimation.

The outline of this paper is as follows. In Sec. 3.2, we give a pedagogical example of how one might verify the estimation of expectation values of an arbitrary Hamiltonian, by writing it as a sum of Pauli operators and performing (fast-forwarded) verified phase estimation on each individual term. In Sec. 3.3 we then derive the theory behind verified phase estimation itself, outline how it can mitigate errors, give algorithms for performing verified phase estimation with a single control qubit, or with access to a

reference state, and study the increased sampling noise cost. In Sec. 3.4, we extend these ideas to give algorithms for verified expectation value estimation, and derive the conditions under which one may perform verified estimation of multiple expectation values in parallel (i.e. using the same system register). In Sec. 3.5, we then implement these ideas, studying the mitigation power of verified expectation value estimation in a variety of systems and implementations developed earlier in the text ounder various noise models, and testing the convergence of the protocol under sampling noise.

3.2 Pedagogical example of verification protocol for expectation value estimation

In this section we outline a simple implementation of verified expectation value estimation of a target operator H on a state $|\psi\rangle$, as a practical example of the more complicated methods to be found later in the text. The idea behind all verification protocols is to prepare $|\psi\rangle = U_{\mathrm{p}}|0\rangle$, indirectly estimate $\langle H \rangle$ via phase estimation, and then verify that we remain in $|\psi\rangle$ by uncomputing $|0\rangle = U_{\mathrm{p}}^{\dagger}|\psi\rangle$ and measuring in the computational basis. If $|\psi\rangle$ is not an eigenstate of H, the system may by shifted away from this state by the QPE unitary — i.e. even in the absence of error we do not expect the system to always pass verification. However, as we will show later in this work, the data required for phase estimation is contained entirely within the set of experiments that pass verification; we may effectively ignore any experiments that fail. This in turn allows us to ignore any errors that knock the system away from $|\psi\rangle$, making this a potent error mitigation scheme. We have constructed various implementations of this idea, which we will expand on in Sec. 3.3 and Sec. 3.4, and compare in Sec. 3.5. However, the most general protocols require relatively complicated circuits and classical post-processing. For clarity of exposition, in this section we focus on stepping through a simple protocol for the verification of expectation values, which avoids complex signal processing and circuity requirements. The protocol we describe will work for arbitrary H and $|\psi\rangle$, and may often be a desirable choice for a real experiment. However, depending on the choice of H and $|\psi\rangle$ and the noise model, other protocols described later in the text may be more optimal in terms of their mitigation power.

A process diagram for a simplified verified phase estimation protocol is given in Fig. 3.1. To begin, we write H as a sum of fast-forwardable
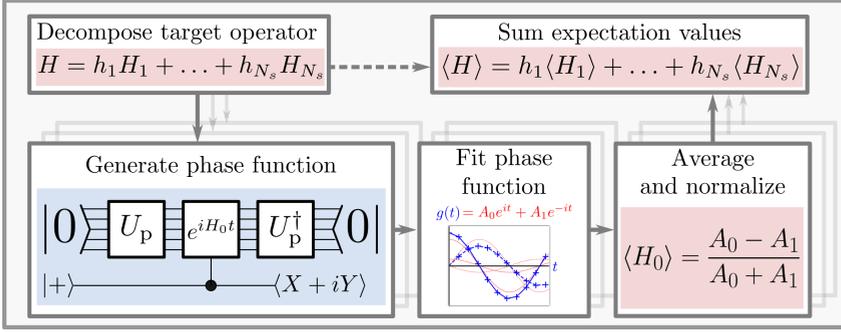
Figure 3.1: Process diagram of the protocol for verified estimation of the expectation value of a Hamiltonian on a state $|\psi\rangle = U_\mathrm{p}|\vec{0}\rangle$. Blue denotes circuits to be executed or data to be extracted from a quantum computer, red denotes signal details to be estimated via classical post-processing. The protocol proceeds as follows: (top-left) a complex Hamiltonian H is split into a number of fast-forwardable summands H_s. The spectral function $g(t)$ of $|\Psi\rangle$ under time evolution of each piece is obtained (bottom-left) via verified, fast-forwarded phase estimation. In this example, a control qubit is used to extract the phase function via phase kickback. The resulting data is a weighted sum of oscillations with frequencies equal to the eigenvalues $E_j^{(s)}$ of the corresponding factor (bottom middle). This may be decomposed in a variety of classical post processing techniques to obtain estimations of the expectation values $\langle H_s\rangle$ depending on the type of H_s chosen (bottom-right). Regardless of the method used, the expectation values must be normalized to obey Eq. 3.29, the last step in the verification process. As the expectation value is linear, the verified estimates of $\langle H_s\rangle$ obtained may be immediately summed together to give a verified estimate for $\langle H\rangle$ (top-right).

terms H_s (multiplied by coefficients h_s)

$$H = \sum_{s=1}^{N_s} h_s H_s. \tag{3.1}$$

Here, by fast-forwardable, we mean that each H_s is chosen such that time evolution $e^{iH_s t}$ may be implemented on a quantum register with the same number of gates for each value of t. Although fast-forwarding is forbidden for arbitrary H [97], decomposition of any sparse, row-computable H into

a linear combination of polynomially many fast-forwardable Hamiltonians is always possible [127]. For example, the N-qubit Pauli operators $P_i \in \mathbb{P}^N = \{\mathbb{1}, X, Y, Z\}^{\otimes N}$ form a basis for the set of all N-qubit operators and are themselves fast-forwardable; we take this decomposition for our simple example.

We then implement verified phase estimation (with a single control qubit) to estimate the expectation values $\langle \psi | H_s | \psi \rangle$. This involves evolving the system by H_s conditional on a control qubit. (Circuits to implement this are well-known, see e.g. Ref. [128].) The conditional evolution encodes a phase function on the the control qubit. That is, if we write X_c and Y_c for the X and Y Pauli operators on this control qubit, we have following the conditional evolution that

$$\langle X_c \rangle + i \langle Y_c \rangle = A_0 e^{it} + A_1 e^{-it} =: g(t). \tag{3.2}$$

Here, A_0 and A_1 are the squared amplitudes of $|\psi\rangle$ in the eigenbasis of H_s (which has known eigenvalues ± 1). The expectation value $\langle X_c \rangle$ may be estimated by measuring the control qubit M times in the x-basis, counting the number of times $m_{x,0}$ or $m_{x,1}$ a 0 or 1 was seen, and approximating

$$\langle X_c \rangle \approx \frac{m_{x,0} - m_{x,1}}{M}. \tag{3.3}$$

(A similar procedure may be performed for Y.) To verify this estimate, we uncompute the preparation of the system, and count the number $m_{x,0}^{(v)}$ ($m_{x,1}^{(v)}$) of measurements of 0 (1) on the control qubit when the uncomputed state on the system is returned to the initial $|0\rangle$ state. We then replace our estimation by

$$\langle X_c \rangle \approx \frac{m_{x,0}^{(v)} - m_{x,1}^{(v)}}{M}. \tag{3.4}$$

(Note that we only replace the numerator, and not the denominator, of Eq. 3.3, which makes this not strictly post-selection — see Sec. 3.3.2 for more details.) The expectation value $\langle H_s \rangle$ is encoded within the phase function $g(t)$, and must be inferred from these estimates above. In our example protocol, this requires inferring the amplitudes A_0 and A_1 (as the eigenvalues ± 1 are already known). These may be simply estimated by a two-parameter fit of Eq. 3.2 to the extracted values of $g(t)$.

As we show later in the text, in the absence of error Eq. 3.3 and Eq. 3.4 yield the same result (in the large M limit). Errors tend to scatter the system into a state that fails verification. The primary effect this has on the estimator in Eq. 3.4 is to re-scale $g(t) \rightarrow p_{ne} g(t)$ (where p_{ne} is the

probability of no error occurring). However, the converse is not true; states may fail verification due to the relative dephasing betwen the $|0\rangle$ and $|1\rangle$ eigenstates of H_s, and we cannot infer the value of p_{ne} from a single point $g(t)$. Instead, we can infer the value of p_{ne} from the normalization of the starting state $|\psi\rangle$. As our circuit is fast-forwarded, under reasonable noise assumptions p_{ne} is independent of t, and this propagates immediately through the fit of Eq. 3.2: $A_0, A_1 \to p_{\mathrm{ne}}A_0 =: \tilde{A}_0, p_{\mathrm{ne}}A_1 =: \tilde{A}_1$. The normalization of $|\psi\rangle$ requires $A_0 + A_1 = 1$, and we may correct for this by estimating

$$\langle H_s \rangle = \frac{\tilde{A}_0 - \tilde{A}_1}{\tilde{A}_0 + \tilde{A}_1}. \tag{3.5}$$

Finally, as expectation values are linear, after repeating this procedure for all H_s in Eq. 3.1, we may sum the result;

$$\langle H \rangle = \sum_s h_s \langle H_s \rangle. \tag{3.6}$$

Note that each H_s will have different values of A_0, A_1, and $g(t)$ (we have avoided explicitly labeling the above for simplicity). In practice, the number of samples for estimation of each $\langle H_s \rangle$ should be varied to minimize the error in the final estimation of $\langle H \rangle$ (i.e. importance sampling on the h_s coefficients).

3.3 Schemes for verified phase estimation

3.3.1 Review of single-control quantum phase estimation

Quantum phase estimation (QPE) refers to a family of protocols to learn eigenphases $e^{i\phi_j}$ of a unitary operator U. Equivalently, quantum phase estimation may be used to learn eigenvalues E_j of a Hermitian operator H, as each such operator generates a unitary via exponentiation: $U = e^{iHt}$ [62]. (Such estimation requires limiting the size of t to prevent aliasing - $e^{iE_j t} = e^{iE'_j t}$ if $E_j t = E'_j t + 2n\pi$, which makes estimation ambiguous.) The eigenvalues of H and the eigenphases of U are related by the same exponentiation and correspond to the same eigenstates $|E_j\rangle$ — if $H|E_j\rangle = E_j|E_j\rangle$, $U|E_j\rangle = e^{i\phi_j}|E_j\rangle$, and $\phi_j = E_j t$.

In the single-control variant of QPE, the phases ϕ_j are learnt by imprinting them on a control qubit — a process known as phase kickback. Any unitary U may be implemented as a (perhaps approximate) quantum

circuit on a quantum 'system' register, but quantum mechanics tells us that $e^{i\phi}|\psi\rangle \equiv |\psi\rangle$ for all pure states $|\psi\rangle$ and numbers $\phi \in \mathbb{R}$. This implies that if the system register were prepared in the pure state $|E_j\rangle$ and U applied, we would not be able to infer the phase ϕ_j from the resulting state $e^{i\phi_j}|E_j\rangle \equiv |E_j\rangle$. However, a relative phase ϕ between two states, $\frac{1}{\sqrt{2}}(|\psi_1\rangle + e^{i\phi}|\psi_2\rangle)$, is a physical observable that may be detected. Such detection may be achieved by acting the unitary U conditional on the control qubit being in the state $|1\rangle$ (and doing nothing when the control qubit is in the state $|0\rangle$). This is commonly written as the 'controlled' unitary $C - U$. When $C - U$ acts on a system register prepared in an eigenstate $|E_j\rangle$ and a control qubit prepared in the state $(|0\rangle + |1\rangle)/\sqrt{2}$, the global state evolves to

$$C - U \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|E_j\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\phi_j}|1\rangle)|E_j\rangle. \qquad (3.7)$$

We see that the eigenphase $e^{i\phi_j}$ from the system register is kicked back onto the control qubit, while the system register itself remains unchanged. We may estimate this eigenphase $e^{i\phi_j}$ by repeatedly performing the QPE protocol, measuring the control qubit in the X or the Y basis, and recording the number of single-shot readouts of 1 and 0. In the Hamiltonian case, from this estimate one may immediately infer $\frac{1}{it}\text{Arg}(e^{i\phi_j}) = E_j \mod 2\pi t$. The error in the estimation of E_j decreases with t; asymptotically optimal protocols need to balance this against the ambiguity modulo $2\pi t$ by repeating the estimation at multiple values of t [129–131]. In terms of estimating the eigenphases $e^{i\phi_j}$ of a unitary U, this optimization requires repeating the above procedure for $C - U^k$ at varying points k.

Often, one does not prepare an eigenstate $|E_j\rangle$, but instead prepares a starting state

$$|\psi_s\rangle = \sum_j a_j |E_j\rangle. \qquad (3.8)$$

Applying $C - U^k$ to such a state no longer leaves it unchanged, but instead entangles it with the control qubit. This produces the combined state (on the system+control register)

$$|\Psi(k)\rangle = C - U^k \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|\psi_s\rangle \qquad (3.9)$$

$$= \sum_j \frac{a_j}{\sqrt{2}}(|0\rangle + e^{ik\phi_j}|1\rangle)|E_j\rangle. \qquad (3.10)$$

When one has instead performed controlled time evolution (via the unitary $C - e^{iHt}$), one may instead write

$$|\Psi(t)\rangle = C - e^{iHt} \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) |\psi_{\rm s}\rangle \tag{3.11}$$

$$= \sum_j a_j \frac{1}{\sqrt{2}} (|0\rangle + e^{iE_j t}|1\rangle)|E_j\rangle. \tag{3.12}$$

The sum over j in the above equation looks problematic, but it turns out that the eigenphases ϕ_j (or eigenvalues E_j) remain encoded on the control qubit, in a sum weighted by the norm square $A_j := |a_j|^2$ of the initial amplitudes a_j. To be precise, one may trace over the system register to obtain the reduced density matrix of the control qubit

$$\rho_{\rm c}(t) = \text{Trace}_{\rm sys}\left[|\Psi(t)\rangle\langle\Psi(t)|\right] \tag{3.13}$$

$$= \frac{1}{2} \begin{pmatrix} 1 & g(t) \\ g^*(t) & 1 \end{pmatrix}, \tag{3.14}$$

with $g(t)$ the phase function of $|\psi_{\rm s}\rangle$ under H

$$g(t) = \sum_j A_j e^{iE_j t}. \tag{3.15}$$

Estimates of $g(t)$ may be obtained as an expectation value

$$g(t) = 2\, \text{Trace}_{\rm c}\left[\rho_{\rm c}(t)|0\rangle\langle 1|\right] \tag{3.16}$$

$$= \text{Trace}_{\rm c}\left[\rho_{\rm c}(t)X\right] + i\text{Trace}_{\rm c}\left[\rho_{\rm c}(t)Y\right], \tag{3.17}$$

of the Pauli operators X and Y. Measuring these expectation values requires rotating the control qubit into the x- or y-basis, reading it out, and averaging the output over many repetitions (or shots) of the experiment.

For a unitary operator U one may obtain an equivalent phase function

$$g(k) = \sum_j A_j e^{ik\phi_j}, \tag{3.18}$$

by estimating

$$g(k) = 2\,\mathrm{Trace}_c\big[\rho_c(k)|0\rangle\langle 1|\big] \tag{3.19}$$

$$= \mathrm{Trace}_c\big[\rho_c(k)X\big] + i\,\mathrm{Trace}_c\big[\rho_c(k)Y\big], \tag{3.20}$$

$$\rho_c(k) = \mathrm{Trace}_{\mathrm{sys}}\big[|\Psi(k)\rangle\langle\Psi(k)|\big], \tag{3.21}$$

with $|\Psi(k)\rangle$ defined in Eq. 3.10. The tomography to extract these expectation values is the same as described in the previous paragraph.

Information about the eigenvalues E_j and amplitudes $A_j = |a_j|^2$ may be inferred classically from estimates of $g(t)$ at multiple values of t. When these are estimated sufficiently well, the expectation value of the Hamiltonian may be calculated

$$\langle H \rangle = \sum_j A_j E_j. \tag{3.22}$$

Inference of the amplitudes A_j from $g(t)$ to error ϵ takes asymptotically time $\Omega(\epsilon^{-2})$ on a quantum device, even when the eigenvalues E_j are already known [1]. By propagating variances, this implies equivalent convergence in the estimation of expectation values via Eq. 3.22. One need not resolve all 2^N eigenvalues of an N-qubit operator in order to evaluate Eq. 3.22. Time-series analysis methods [125] or integral methods [132] produce a coarse-grained approximation to the spectrum that may be averaged over to obtain expectation values with similar convergence rates. Alternatively, for simple operators with highly-degenerate spectrum (e.g. Pauli operators), curve fitting will be sufficient to extract the required data (as described in Sec. 3.2)[2].

3.3.2 Verifying a phase estimation experiment

As the data from single-control quantum phase estimation is accumulated entirely on the control qubit, one would be tempted to throw the system register away (or rather, reset the register and begin anew). In the absence of error correction this temptation grows larger; noise levels in near term devices are high enough that coherent states of more than a few qubits degrade over the course of any reasonably-sized algorithm to within a few percent fidelity to the target state — if not less [103]. However, even when

---

[1] This may be calculated via Cramer-Rao bounds as the derivative $\frac{\partial g(t)}{\partial A_j}$ is bounded as a function of t, which is not true for the derivative $\frac{\partial g(t)}{\partial E_j}$.

[2] The minimum number of points on the curve that require fitting is determined by the number of eigenvalues and amplitudes that need fitting.

corrupted, the information contained within the system register is valuable, as one can use this information to diagnose potential errors in the data to be read from the control qubit. For instance, in the presence of global symmetries of the Hamiltonian, one could imagine mitigating errors that do not commute with this symmetry via symmetry verification [29, 122, 123]. In verifying these symmetries, we are in effect projecting the system into a subspace of the global Hilbert space which contains the information we desire. One could imagine constructing ever-smaller Hilbert spaces, which trades circuit complexity for error-detection power. It turns out that the limit of this construction is achievable: instead of measuring one or more symmetries on the system register, we can instead verify that it has returned to its initial state $|\psi_{\mathrm{s}}\rangle$. (This is similar to the echo-type measurements made in randomized benchmarking [133] or quantum Hamiltonian learning [134].)

Assuming $|\psi_{\mathrm{s}}\rangle$ is prepared from the computational basis state $|0\rangle$ by a preparation unitary U_{p}, this measurement may be achieved by applying $U_{\mathrm{p}}^{\dagger}$, and reading out each qubit in the computational basis. One would expect such a measurement to distort the phase function $g(t)$, but this is not so, as we may expand the trace in Eq. 3.13 to show that

$$\mathrm{Trace_c}\big[\rho_{\mathrm{c}}(t)|0\rangle\langle 1|\big]$$
$$= \mathrm{Trace_c}\Big[\langle\psi_{\mathrm{s}}|\Psi(t)\rangle\langle\Psi(t)|\psi_{\mathrm{s}}\rangle|0\rangle\langle 1|\Big]. \qquad (3.23)$$

Here, the left-hand side of the equation is the expectation value of $\rho_{\mathrm{c}}(t)$ regardless of the state of the control register, and the right-hand side is the (non-normalized) expectation value of $\rho_{\mathrm{c}}(t)$ on verified experiments only. The lack of normalization means this is not a post-selection technique; instead one assumes that the contribution of states which fail verification to the final estimation of $g(t)$ is zero. (By contrast, states that pass verification either contribute $+1$ or -1 to the estimation of $g(t)$.)

We can make a physical argument why Eq. 3.23 holds and verification should not affect the estimation of $g(t)$ in the absence of noise. Let us decompose the reduced density matrix on the control qubit

$$\rho_{\mathrm{c}} = \rho_{\mathrm{c}}^{(\mathrm{v})} + \rho_{\mathrm{c}}^{(\mathrm{f})}, \qquad (3.24)$$

into the ensemble of states $\rho_{\mathrm{c}}^{(\mathrm{v})}$ that have passed verification, and those that have failed. When the control qubit is in the $|0\rangle$ state, the system register is not evolved, so in the absence of noise the state will pass verification every time. This implies that a verification failure in the absence of noise projects

the control qubit into the $|1\rangle$ state; $\rho_c^{(f)} = |1\rangle\langle1|$. As $\mathrm{Trace}[|1\rangle\langle1|0\rangle\langle1|] = 0$, this fraction of states on average contributes nothing to the estimate of $g(t)$. In other words,

$$\mathrm{Trace}[\rho_c|0\rangle\langle1|] = \mathrm{Trace}[\rho_c^{(v)}|0\rangle\langle1|] = g(t). \tag{3.25}$$

Note that post-selecting (i.e. keeping only the experimental data where verification was passed) would instead prepare the state $\rho_c^{(v)}/\mathrm{Trace}[\rho_c^{(v)}]$. This will not yield the desired result, as

$$\frac{\mathrm{Trace}[\rho_c^{(v)}(X+iY)]}{\mathrm{Trace}[\rho_c^{(v)}]} = \frac{g(t)}{1+|g(t)|^2} \tag{3.26}$$

which is not equal to $g(t)$ unless $|\psi_s\rangle$ is an eigenstate of e^{iHt} (in which case $\rho_c^{(v)} = \rho_c$). (Moreover, this rescaling can be up to a factor 2 in the absence of noise, and the spectrum of this new function is significantly different to the original.) To give some intuition, one can imagine phase estimation on a mixed state in two steps: performing phase estimation on individual states to generate a set of signal functions e^{iE_jt}, and then summing and returning the weighted result $g(t)$. The set of states that fail verification, $\rho^{(f)}$, captures the relative dephasing between these states, which cannot be ignored when attempting to recover this result. Instead, an explicit protocol for the measurement of a single $g(t)$ within verified single-control phase estimation takes the form of Algorithm 3.1. We consider the increased sampling cost in the presence of error in Sec. 3.3.3.

3.3.3 Why verification mitigates errors

The mitigation power from verification is based on the relative size of the Hilbert spaces in which the states which have passed verification and states which have failed verification, $\rho = \rho^{(v)} + \rho^{(f)}$, live. If we define the Hilbert spaces in which the two ensembles live $\mathcal{H}^{(v)}$ and $\mathcal{H}^{(f)}$ respectively, we have $\dim[\mathcal{H}^{(v)}] = 2$, while $\dim[\mathcal{H}^{(f)}] = 2^{N+1} - 2$. An error that occurs during the circuit is then likely to scatter the system into the set of rejected states. As an extreme example, the probability that a completely random error (i.e. an error that scatters all states to a random state) at any point in the circuit will yield a state in $\mathcal{H}^{(v)}$ can be immediately calculated to be $2/(2^{N+1} - 2) \sim 2^{-N}$. This includes errors during preparation of $|\psi_s\rangle$ by the unitary U_p and the inversion of $U_p^\dagger$ to perform the verification itself. As we are not post-selecting on the verification output $g(t)$ is still affected by this shift, but the distortion may be accounted for in classical

---

Algorithm 3.1: Single-control VPE

Input: Circuits to implement $U_\mathrm{p}, U_\mathrm{p}^\dagger$ and controlled time evolution e^{iHt}.

Number of repetitions M of measurement in the x and y basis.

Output: An estimate of $g(t)$ with variance $O(\frac{1}{M})$ in both the real and imaginary part.

1 Prepare classical initial variables $g^x = 0$, $g^y = 0$;
2 Prepare the system register in a starting state $|\psi_\mathrm{s}\rangle = U_\mathrm{p}|0\rangle$ and the control qubit in the state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$;
3 Simulate time evolution e^{iHt} conditional on the control qubit;
4 Apply the inverse circuit $U_\mathrm{p}^\dagger$ to the system register;
5 Rotate the control qubit into the X or Y basis and measure it to obtain a number $m \in [0, 1]$;
6 If all qubits in the system register read 0, increment the relevant variable g^x or g^y by $(-1)^m$;
7 Repeat steps 2-6 M times in the X basis and M times in the Y basis, and estimate $g(t)$ by $\frac{g^x}{M} + i\frac{g^y}{M}$

---

post-processing. In this simple noise model the effect of noise is then to replace the estimate of $g(t)$ by

$$g_\mathrm{err}(t) = p_\mathrm{ne}(t)g(t) + \mathcal{O}(2^{-N}p_\mathrm{err}(t)), \tag{3.27}$$

where $p_\mathrm{ne}(t)$ and $p_\mathrm{err}(t)$ are the probabilities of no error or some error occurring, respectively. (In App. 3.A we derive the specific requirements for this to be the case.) Assuming that errors occur at a constant rate as a function of the circuit depth, and all scatter the system outside $\mathcal{H}^{(\mathrm{v})}$, for fast-forwardable Hamiltonians $p_\mathrm{ne}(t) = p_\mathrm{ne}$, and

$$g_\mathrm{err}(t) = p_\mathrm{ne}g(t) = \sum_j (p_\mathrm{ne}A_j)e^{iE_j t}. \tag{3.28}$$

This can be seen as a uniform damping of each squared amplitude A_j to $A_j' = p_\mathrm{ne}A_j$. Such damping may be corrected for classically as we know $|\psi_\mathrm{s}\rangle$ is normalized

$$\sum_j A_j = 1, \tag{3.29}$$

and so we may estimate

$$A_j = \frac{A'_j}{\sum_j A'_j}.$$

(3.30)

Depending on the classical signal processing method used, one may not obtain estimates of all A'_j and E_j, but may instead directly calculate $\sum_j A'_j E_j$ and $\sum_j A'_j$. For example, one could use $g_{\mathrm{err}}(0) = \sum_j A'_j$ as such a reference point. For non-fast-forwardable Hamiltonian, assuming again that errors occur at a constant rate throughout the circuit and that all scatter the system outside $\mathcal{H}^{(\mathrm{v})}$, we have

$$g_{\mathrm{err}}(t) = e^{-t/\tau_{\mathrm{err}}} g(t) = \sum_j A_j e^{i(E_j + i/\tau_{\mathrm{err}})t}.$$

(3.31)

This can be seen to be an imaginary shift to the eigenvalues $E_j \to E_j + i\tau_{\mathrm{err}}$. It can be corrected for in signal processing of the phase function by taking only the real parts of the E_j eigenvalues.

The above analysis is not necessarily true for simulation of an arbitrary Hamiltonian under a realistic noise model. In particular, if the instantaneous state during simulation is a near-eigenstate of the error model, then the correction in Eq. 3.27 may be as large as $\mathcal{O}(1)$ instead of $\mathcal{O}(2^{-N})$. In App. 3.A we study this in more detail, and specify the conditions under which errors will distort the results of verified phase estimation.

Sampling costs

The error mitigation from verification comes at the cost of increasing the number of samples require to estimate $g(t)$. Assuming all errors fall outside the verified subspace, estimating $g(t)$ to precision ϵ requires estimating $g_{\mathrm{err}}(t)$ to precision $p_{\mathrm{ne}}\epsilon$. To obtain g^x in Alg. 3.1 (and equivalently for g^y) we average over a set of M experimnetal outputs that may take the values $\{-1, 0, 1\}$. Let us define the ith experimental output g^x_i, and we have

$$P(g^x_i = \pm 1) = \frac{1}{2} p_{\mathrm{ne}} (1 \pm g^x),$$

(3.32)

$$P(g^x_i = 0) = 1 - p_{\mathrm{ne}}.$$

(3.33)

Our estimate of the noisy $g_{\mathrm{err}}(t)$ is then given by

$$\mathrm{Re}[g_{\mathrm{err}}(t)] = P(g^x_i = 1) - P(g^x_i = -1).$$

(3.34)

As each experiment is IID, the variance on our estimates of these probabilities is

$$\mathrm{Var}[P(g_i^x = \pm 1)] = \frac{1}{M} \frac{1}{2} p_{\mathrm{ne}}(1 \pm g^x)$$

$$\times \left(1 - \frac{1}{2} p_{\mathrm{ne}}(1 \pm g^x)\right), \qquad (3.35)$$

$$\mathrm{Cov}[P(g_i^x = 1), P(g_i^x = -1)] \qquad (3.36)$$

$$= -\frac{1}{4M} p_{\mathrm{ne}}^2 \left(1 - [g^x]^2\right). \qquad (3.37)$$

Propagating variances obtains

$$\mathrm{Var}\left[\mathrm{Re}\left[g_{\mathrm{err}}(t)\right]\right] = \frac{1}{M} p_{\mathrm{ne}} - \frac{1}{M} p_{\mathrm{ne}}^2 [g^x]^2. \qquad (3.38)$$

We may then bound the requirements to estimate $g_{\mathrm{err}}(t)$ to variance $\epsilon^{-2} p_{\mathrm{ne}}^{-2}$ by

$$M \geq \epsilon^{-2} p_{\mathrm{ne}}^{-1}. \qquad (3.39)$$

This is exactly what one would expect from an actual post-selection technique (i.e. where $M p_{\mathrm{ne}}$ samples were used to estimate $g(t)$). We remind the reader that p_{ne} here is the probability of no error occurring over the entire circuit. As one should expect for an error mitigation technique, this in turn grows exponentially with the size of the circuit required to implement e^{iHt} or U_{p}. In a simple model, if the error per qubit per moment is p (i.e. assuming qubit decay is more dominant than gate noise in the model), an N-qubit circuit of depth d would have

$$p_{\mathrm{ne}} = (1 - p)^{Nd}, \qquad (3.40)$$

and thus the number of shots required to estimate (the real or imaginary part) of $g(t)$ would scale as

$$M \sim (1 - p)^{-Nd} \epsilon^{-2}. \qquad (3.41)$$

This is not to be ignored; verification requires at least doubling the size of the circuit, which if $p_{\mathrm{ne}} = 0.01$ (as has been reported [3] and mitigated successfully [103] in previous experiments) will increase the measurement count by a factor of 100. Some of the methods presented in this work involve increasing the circuit depth by factors of up to 14, which will be impractical for large experiments without further circuit optimization.

Control noise

An important realistic error to consider in QPE is error on the control qubit. This keeps the system within the verified subspace, and so is not captured by the above analysis. However the effect of many common error channels may still be mitigated by verification. For example, let us assume that the circuit decomposition of $C - U$ involves the control qubit performing only single-qubit gates and controlled operations on the rest of the circuit (which is typically the case). In this case, one may show that the effect of a depolarizing channel of strength λ

$$R_{\text{depol}}[\rho] = (1 - \frac{3\lambda}{4})\rho + \frac{\lambda}{4}(X\rho X + Y\rho Y + Z\rho Z), \qquad (3.42)$$

acting on the control qubit at any point in the circuit, sends the final state of the system to

$$(1 - \lambda)\rho_{\text{ne}} + \lambda\rho_{\text{err}}, \qquad (3.43)$$

where ρ_{ne} is the state in the absence of error, and

$$\text{Trace}[\langle \psi_s | \rho_{\text{err}} | \psi_s \rangle |0\rangle\langle 1|] = 0. \qquad (3.44)$$

In this case, the (noisy) estimate of $g(t)$ is sent to $(1 - \lambda)g(t)$, and expectation values and eigenvalues may be recovered via the same analysis as in Sec. 3.3.3. However, the above analysis will not hold for a more general noise model, and schemes such as randomized compiling [135] may be required to unbias the estimate of $g(t)$. An example of this biasing effect is if an amplitude-damping channel

$$
\begin{aligned}
R_{\text{ampdamp}}[\rho] =& (1 - \lambda)\rho + \frac{\lambda}{2}(Z + I)\rho(Z + I) \\
&+ \frac{\lambda}{2}(X + iY)\rho(X - iY),
\end{aligned}
\qquad (3.45)
$$

is present on the control qubit between the final measurement pre-rotation and readout in the computational basis. Left unchecked, this will shift the estimate of $g(t)$ to

$$g_{\text{err}}(t) = (1 - \lambda)g(t) + \lambda. \qquad (3.46)$$

In addition to damping the true signal $g(t)$, this additive signal presents as a 0-energy eigenvalue in the spectrum of $g(t)$. This will not be accounted for by naive renormalization of $\langle H \rangle$ as outlined in Alg. 3.3; the estimation protocol will instead estimate $(1 - \lambda)\langle H \rangle$. Though this could be corrected in post-processing, we suggest that a more stable mitigation is to flip the

$|0\rangle$ and $|1\rangle$ states on the control qubit for 50% of experiments. This may be compiled into the final pre-rotation, and does not increase the total sampling cost of the experiment (only half as many samples need to be taken at each pre-rotation setting for the same accuracy). We observe similar biases on bitflip noise channels which tend to decay the real and imaginary parts of $g(t)$ asymmetrically. This may be compensated for in turn by compiling a $\frac{\pi}{4}$ Z-rotation on the intial control qubit state, and uncompiling it in the final prerotation. (One can see that this commutes with all gates in the circuit). For the noise models studied numerically in this text we have found either one or both of the above compilation schemes sufficient to mitigate control error. More complicated noise models may required more complicated compilation schemes; extending the above will be an interesting task for future work. In particular, the above analysis does not apply to correlated two-qubit noise during operations between the control qubit and the rest of the system.

3.3.4 Verified control-free phase-estimation

As was recently demonstrated in Ref. [27], the control qubit may be removed from a QPE experiment if we have the ability to prepare an alternative reference eigenstate $|\psi_r\rangle$ of the Hamiltonian H (with $\langle\psi_s|\psi_r\rangle = 0$). For example, in the electronic structure problem in quantum chemistry the number-conserving Hamiltonian has the vacuum as a potential reference state. (A similar situation was considered in Ref. [136] for the purposes of random gap estimation, but estimating single eigenvalues E_j from this class of experiments is somewhat awkward.) This was also recently considered as an extension to the well-known robust QPE scheme [137], requiring both $|\psi_r\rangle$ and $|\psi_s\rangle$ to be eigenstates of the system [126]. Note that $|\psi_r\rangle$ need not necessarily be a zero-energy eigenstate of H, though the corresponding eigenenergy E_r should be known to high accuracy. In this case, one need prepare the correlated state $\frac{1}{\sqrt{2}}(|\psi_s\rangle + |\psi_r\rangle)$, and perform uncontrolled time evolution, and finally measure the off-diagonal element $|\psi_s\rangle\langle\psi_r|$. This is shown in the circuit (Fig. 3.2). Evaluating the circuit provides an estimate of

$$\text{Trace}\Big[U(|\psi_r\rangle + |\psi_s\rangle)(\langle\psi_r| + \langle\psi_s|)U^\dagger|\psi_r\rangle\langle\psi_s|\Big]$$
$$= e^{-iE_r t}g(t), \qquad (3.47)$$

and the additional phase may be subtracted in post-processing.

The protocol for verified control-free phase estimation does not differ

$$\vec{0}\rangle \boxed{\boxed{\mathbb{H}}}\boxed{U_{\mathrm{p}}}\boxed{e^{iHt}}\boxed{U_{\mathrm{p}}^{\dagger}}\frac{|0\rangle\langle 1|}{\langle \vec{0}|}$$
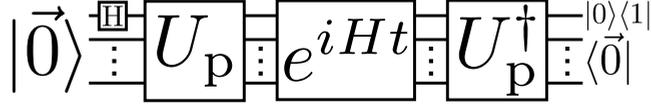
Figure 3.2: Quantum circuit for control-free verified phase estimation. The preparation unitary U_{p} is defined in Eq. 3.48. The first gate in the circuit is a Hadamard gate (Roman H) on the top-most qubit (labeled the target qubit in text), which should not be confused with the Hamiltonian H.

significantly from the single-control case. Besides the loss of the control qubit and removal of control from the time evolution circuit, we also now require our preparation circuit to prepare the starting state $\frac{1}{\sqrt{2}}(|\psi_{\mathrm{s}}\rangle + |\psi_{\mathrm{r}}\rangle)$. We assume that this is achieved by first applying a Hadamard gate to a single target qubit in the system register, placing the system in the state $\frac{1}{\sqrt{2}}(|0\rangle + |\vec{1}_{\mathrm{T}}\rangle)$. (Here we use the notation $|\vec{1}_{\mathrm{T}}\rangle$ for the basis state where the target qubit is in the $|1\rangle$ state and all other qubits are in $|0\rangle$.) Then, the desired preparation may be achieved by a preparation unitary U_{p} which performs the mapping

$$U_{\mathrm{p}}|0\rangle \mapsto |\psi_{\mathrm{r}}\rangle, \qquad U_{\mathrm{p}}|\vec{1}_{\mathrm{T}}\rangle \mapsto |\psi_{\mathrm{s}}\rangle. \qquad (3.48)$$

(We use the same notation as for the single-control unitary on purpose, as under the association $|0\rangle|\psi_{\mathrm{s}}\rangle \leftrightarrow |\psi_{\mathrm{r}}\rangle$ and $|1\rangle \leftrightarrow |\psi_{\mathrm{s}}\rangle$ one may see the two are equivalent.) With this definition, estimation of $|\psi_{\mathrm{r}}\rangle\langle\psi_{\mathrm{s}}|$ may be achieved by inverting U_{p}, as

$$|\psi_{\mathrm{r}}\rangle\langle\psi_{\mathrm{s}}| = U_{\mathrm{p}}|0\rangle\langle\vec{1}_{\mathrm{T}}|U_{\mathrm{p}}^{\dagger}. \qquad (3.49)$$

In particular, after inversion, the reduced density matrix of the target qubit contains the desired phase function $g(t)$, and the verification consists of checking whether all other qubits are measured into 0. The resulting control-free protocol is fully written out in Algorithm 3.2. The analysis of Sec. 3.3.3 is identical for the control-free case, with the absence of the issue of control noise, as is the analysis of Sec. 3.3.3. However, we note that at the beginning and the end of any experiment, single-qubit noise on the target qubit behaves similarly to control qubit noise. This necessitates averaging over multiple initial and final rotations of the target qubit to prevent bias in the estimation of $g(t)$.

The above analysis implies that the algorithms studied in Ref. [27, 136] should be amenable to verification immediately as well. It also provides

---

Algorithm 3.2: Control-free VPE

Input: Circuits to prepare a superposition of $|\psi_s\rangle$ and $|\psi_r\rangle$, invert
      the preparation, and implement time evolution e^{iHt}.
      Number of repetitions M and M of measurement in the x
      and y basis.
      The reference eigenstate energy E_r

Output: An estimate of $g(t)$ (Eq. 3.47) with variance $O(\frac{1}{M})$ in
      both the real and imaginary part.

1 Prepare classical initial variables $g^x = 0$, $g^y = 0$;
2 Prepare the system register in a starting state
    $\frac{1}{\sqrt{2}}(|\psi_s\rangle + |\psi_r\rangle) = U_p \frac{1}{\sqrt{2}}(|0\rangle + |\vec{1}_T\rangle)$;
3 Apply the unitary U^k, (or equivalently simulate time evolution e^{iHt})
    ;
4 Apply the inverse circuit $U_p^\dagger$ to the system register ;
5 Rotate the target qubit into the X or Y basis and measure it to
    obtain a number $m \in 0, 1$;
6 Measure all other qubits, and if they all read out 0, increment the
    relevant variable g^x or g^y by $(-1)^m$;
7 Repeat steps 2-6 M times in the X basis and M times in the Y
    basis, and estimate $g(t)$ by $e^{iE_r t}(\frac{g^x}{M} + i\frac{g^y}{M})$.

---

some additional explanation for the error-robustness observed in the robust
phase estimation of Ref. [126].

3.4 Verified expectation value estimation

In many circumstances, one wishes not to know the eigenvalues of a
Hermitian operator H, but instead its expectation value $\langle H \rangle$ under a
specified state $|\Psi\rangle$. For instance, in a variational quantum eigensolver [82],
one prepares a state $|\Psi(\vec{\theta})\rangle = U(\vec{\theta})|0\rangle$ dependent on a set of classical input
parameters $\vec{\theta}$, then measures the expectation value $E(\vec{\theta}) = \langle \Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle$.
This is then optimized over $\vec{\theta}$ in a classical outer loop, with the optimized
state $|\Psi(\vec{\theta}_{opt})\rangle$ hopefully a good approximation of the true ground state
$|E_0\rangle$. In quantum variational algorithms it is typical that $\langle \Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle$
is estimated by means of partial state tomography [66, 67, 138]. However,
noise in the preparation unitary $U(\vec{\theta})$ causes an errant state $\rho_{err}(\vec{\theta}) \neq$

Algorithm 3.3: Verified expectation value estimation

Input: (Noisy) circuits to implement U_{p}, $U_{\mathrm{p}}^{\dagger}$ and controlled time
evolution e^{iHt}.
A set of t values.
Number of repetitions M of measurement in the x and y
basis (that can be t-dependent).
A method for classical signal processing (e.g. a curve fitting
algorithm).

Output: An estimate of $\langle H \rangle$.

---

1 Estimate $g_{\mathrm{err}}(t)$ for all given points t using Alg. 3.1 to the chosen
precision;
2 Obtain estimates for individual E_j and A'_j values via classical signal
processing;
3 Estimate $\langle H \rangle$ as

$$\langle H \rangle = \frac{\sum_j A'_j E_j}{\sum_j A'_j}. \tag{3.51}$$

$|\Psi(\vec{\theta})\rangle\langle\Psi(\vec{\theta})|$ to be prepared and tomographed, propagating the preparation
error directly to a final estimation error. The noise analysis in Sec. 3.3.3
extends to both the preparation and mitigation unitaries, so if verified
phase estimation is used to provide estimates of eigenvalues and amplitudes,
one may reconstruct

$$\langle\Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle = \sum_j |a_j|^2 E_j, \tag{3.50}$$

and inherit the mitigation power of the verification protocol. This has the
added advantage that control errors in the preparation circuit (which, being
a repeated error, are not mitigated against) are able to be compensated for
during the outer optimization loop of the VQE, as is well-known [82, 103].
Quantum phase estimation has previously been suggested as an alternative
to partial state tomography for expectation value estimation, both to
improve the rate of estimation [139], and to provide a witness for the
presence of eigenstates of the Hamiltonian [140]. The verification protocols
described in this work should be applicable to these methods as well. A
general algorithm for verified expectation value estimation takes the form
of Algorithm 3.3

One might worry that the sum in Eq. 3.50 is over an exponentially large number of eigenstates $|E_j\rangle$. However one need not resolve all eigenvalues E_j in order to accurately estimate the expectation value $\langle\Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle$; if eigenvalues within δ of each other are binned, the resulting expectation value will be accurate to within δ. We may formalize this by considering the spectral function g_S of $|\psi_s\rangle$ under H,

$$g_S(E) = \sum_j A_j \delta(E - E_j). \tag{3.52}$$

This can be seen to be the Fourier transform of the phase function $g(t)$ [strictly, $g(t)$ is the inverse Fourier transform of $g_S(E/2\pi)$], and a coarse-grained approximation may be obtained via time-series methods [125] or integral methods [132] with rigorous bounds on each. Numerically, we find signal processing methods such as Prony's method [124] also perform acceptably (see Sec. 3.5.4). For fast-forwardable Hamiltonians (such as Pauli operators), one often already knows the target eigenvalues of the problem. Furthermore, the eigenspectrum of these Hamiltonians is often highly degenerate, making simple curve fitting a practical (and attractive) alternative.

Instead of analysing the phase function at many points as described above, one may expand

$$\text{Im}[g(t)] = \sum_j |a_j|^2 \sin(E_j t) \tag{3.53}$$

$$= t \sum_j |a_j|^2 E_j + \frac{1}{3} t^3 \sum_j |a_j|^2 E_j^3 + \mathcal{O}(t^5) \tag{3.54}$$

$$\frac{1}{t}\text{Im}[g(t)] = \langle\Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle + \mathcal{O}(t^2), \tag{3.55}$$

and simply estimate $\text{Im}(g(t))$ for short times t. This is similar to the manner in which eigenphases are estimated in the WAVES protocol [140] (sans verification). In this case, the normalisation of the resulting amplitudes (Eq. 3.30) must be achieved by the condition that $g(0) = \sum_j A_j$, yielding

$$\langle H \rangle = \frac{\text{Im}[g_{\text{err}}(t)]}{t|g_{\text{err}}(0)|} + \mathcal{O}(t^2). \tag{3.56}$$

3.4.1 Fast-forwarded and parallelized Hamiltonian decompositions

As expectation values are linear, we may estimate $\langle H \rangle$ by splitting it into multiple terms, estimating the expectation values of each term individually, and re-summing;

$$H = \sum_s H_s \rightarrow \langle H \rangle = \sum_s \langle H_s \rangle. \qquad (3.57)$$

If individual H_s may be simulated at lower circuit depth, this can reduce the accumulation of unmitigated errors, at the cost of requiring more simulation. This ability becomes especially useful if one chooses the H_s to be fast-forwardable. Here, we define a fast-forwardable Hamiltonian H_s as one for which a circuit implementation of $e^{iH_s t}$ has constant depth in t. The circuit depth required to simulate e^{iHt} for arbitrary H is bounded below as $\mathcal{O}(t)$ [97], but for certain operators this may be improved on [141]. For example, as the Pauli operators $\{\mathbb{1}, X, Y, Z\}^{\otimes N}$ are both fast-forwardable and form a basis for the set of N-qubit Hermitian operators, a set of H_s terms may be taken from these to decompose an arbitrary Hamiltonian. As another example, given an instance of the electronic structure problem, one may attempt a low-rank factorization of the interaction operator into a sum of $\mathcal{O}(N)$ diagonalizable (and thus fast-forwardable) terms [142].

In order to speed up estimation of expectation values of multiple terms H_s in a decomposed Hamiltonian $H = \sum_s H_s$, it may be possible perform the verified phase estimation step of each H_s in parallel. For example, we can perform time evolution of L multiple summands, each controlled by a different control qubit, in between the preparation and verification steps of a single instance. In the absence of verification, such parallelization will not affect the outcome of quantum phase estimation of any individual H_s, so long as all terms estimated in parallel commute. This follows immediately from the fact that the time evolution for one such term does not evolve the system between eigenspaces of another. This is complicated by the addition of verification, as the additional circuitry means that the system may evolve away from $|\psi_s\rangle$ despite a specific control qubit being in $|0\rangle$. In App. 3.B, we show that this gives rise to a set of spurious signals in the estimated phase function $g^{(s)}(t)$:

$$g_q^{(s)}(t) = \sum_{v,j,j'} B_{j,j'}^{(s)} e^{iF_{v,j,j'}^{(s)} t}. \qquad (3.58)$$

Here, the ghost eigenvalues are

$$F_{v,j,j'}^{(s)} = E_j^{(s)} + \sum_{s' \neq s} v_{s'} \left(E_j^{(s)} - E_{j'}^{(s')} \right),\qquad(3.59)$$

where $E_j^{(s')}$ are the true eigenvalues of the Hamiltonians $H_{s'}$, and v is a L-bit vector written in binary (i.e. $v_s \in 0, 1$). The corresponding, v-independent amplitudes are

$$B_{j,j'} = \frac{1}{2^L} A_j A_{j'}.\qquad(3.60)$$

Although this is a far more complicated signal than the standard phase function $g(t)$, we calculate in App. 3.B that it yields the same expectation value; i.e.

$$\sum_{v,j,j'} B_{j,j'} F_{v,j,j'}^{(s)} = \langle H_s \rangle.\qquad(3.61)$$

This implies that verified parallel phase estimation may proceed in much the same way as the series protocol.

3.4.2 Comparison to other methods of error mitigation

Error mitigation techniques differ vastly, both in their cost to implement and their effectiveness against different forms of noise. This implies that care needs to be taken in a real experiment to choose the best mitigation technique (or combination of mitigation techniques) for the job. Though a comparison between multiple techniques in a realistic setting lies outside the scope of this work, we give some predictions here on how VPE might compare in performance to other mitigation techniques, and whether it might be possible to compare to different techniques. We can classify all error mitigation techniques that the authors know of into the following broad categories

- *Circuit design* — many forms of noise may be mitigated by careful design of a circuit to e.g. minimize crosstalk between simultaneous gates [103], cancel out Z over- or under-rotation (e.g. via echo pulses [143]), or optimize a circuit variationally to cancel out control parameter drift on a long timescale [82, 144]. (Whether or not this counts as error mitigation or calibration of the underlying quantum device is left to the reader to decide.) Depending on the source of noise these techniques may significantly reduce or even nullify its effect, which may be far more effective than VPE. On the other hand, noise sources such as T1 error
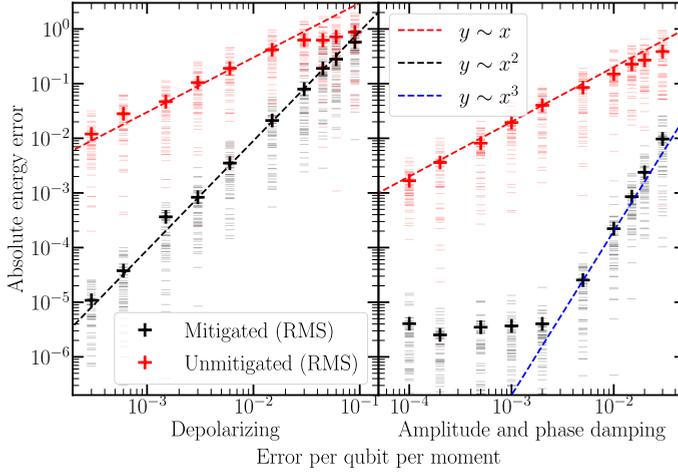
Figure 3.3: Mitigation of a 4-qubit Givens rotation circuit via verified phase estimation. (left) Error in estimation of random states in a free-fermion system (Eq. 3.63) under a uniform depolarizing channel. (right) Error in the same estimation, but this time under an amplitude and phase damping model. In both plots, the RMS error (crosses) is calculated over 50 different estimations for each error rate using either standard partial state tomography (red) or using verified control-free phase estimation. Individual data points (dashes) are additionally shown. For reference, dashed lines showing linear (red), quadratic (black), and cubic (blue) dependence on the gate error rate are plotted.

cannot be easily calibrated away (due to the associated photon loss); in these situations (where VPE performs quite well) these methods will have little effect. VPE is clearly compatible with any such techniques, as these consist of adjustments to the implementation of a given circuit rather than an algorithmic overhead.

- *Post-selection or verification techniques* — this class of techniques uses knowledge of the problem to restrict the state of the quantum device to within a small region of the N-qubit Hilbert space, often by leveraging symmetries of the Hamiltonian of the problem to be solved. VPE itself falls into this category, alongside symmetry verification [29, 67], and quantum subspace expansion techniques [116, 117]. The performance of these techniques is dependent on their ability to catch errors outside the allowed Hilbert space, so as the dimension of the Hilbert space for VPE

is only 2, we expect it to have greater mitigation power in general than these other techniques. (This can be observed in App. 3.G, where VPE shows an asymptotic improvement over symmetry verification in a small numerical simulation.) However, as the circuit depth of VPE is typically far longer than that of other post-selection or verification techniques (which can be achieved in some cases without any additional circuitry), the requirements on the number of measurements to overcome sampling noise will be significantly worse. As these techniques overlap in their effect on the quantum state, it is not particularly possible to combine them; instead one should choose the best trade-off between mitigation power and number of measurements.

- *Error extrapolation techniques* — assuming that one can artificially introduce noise into a system, these techniques rely on parameterizing the output f of a quantum circuit as a function of a 'noise parameter' $f = f(\lambda)$, fitting a functional form, and extrapolating to $\lambda = 0$. The noise parameter can either be adjusted experimentally (e.g. by adjusting the wait time or detuning of an underlying gate) [118, 120] or algorithmically (e.g. by inverting noisy gates [145]). The mitigation power of such a technique depends on how well the noise can be tuned as a function of this single parameter, and how well one can pin down a functional form for $f(\lambda)$. This is not easily comparable to VPE, as the physical source of the mitigation is qualitatively significantly different. We expect that the relative performance will depend on the experiment and the hardware itself. In theory these methods could be combined with VPE (either by extrapolating the phase function or the VPE result). However, it is unclear whether the output of VPE will be more challenging to fit, reducing the effectiveness of the extrapolation.

- *Result extrapolation techniques* — instead of fitting the output f of a quantum circuit to an artificial noise term, one can consider comparing the output of similar quantum circuits tailored to efficient classical simulation. This technique has been demonstrated experimentally in Refs. [103, 146], and proposed within a VQE setting (by tuning the parameters to points where the solution is known) [147]. In some sense VPE can be considered to be similar to these methods, with the $|\psi_{\mathrm{s}}\rangle|0\rangle$ or $|\psi_{\mathrm{r}}\rangle$ states providing an entangled reference state for the target evolution. However, this relationship is not completely clear, as VPE strictly relies on the coherence between the two states. Understanding this similarity is a clear avenue for future research. Regardless, VPE should be able to be combined with at least some of these techniques to provide yet more mitigation power.

- *Probabilistic cancellation techniques* — given knowledge of the true process maps of the gates being performed on a quantum device, one can in principle construct families of quantum circuits that when combined, yield a target noiseless result [118, 120]. However, these methods require much additional characterization of the device, which is a problem in systems with large amounts of drift. In principle given sufficient knowledge of the noise this method works perfectly, but at a greatly increased measurement cost, making it difficult to make a fair comparison in a theoretical setting. Testing this method against VPE in a real experiment would be an interesting target for future research.

- *Purification techniques* — as the output of a quantum algorithm is often ideally pure, these techniques attempt to reduce errors by mapping a noisy impure state to a purer one. This may be achieved e.g. for free fermion states via McWeeny purification [103], or for more general states via virtual distillation [148]. For more complex states the McWeeny process cannot be used, but it has proven remarkably effective when available. Virtual distillation and VPE appear to be remarkably similar in their increased measurement cost and their mitigation performance, as well as their circuit structure. Understanding this similarity and comparing the two in more detail is a clear avenue for future research.

3.5 Numerical Experiments

To investigate the mitigation capability of verified phase estimation, we first use it for expectation value estimation. To prepare states, we take different variational ansatze with randomly-drawn parameters. We compare the performance of verified and unverified circuits across multiple target Hamiltonians, noise strengths and noise models, to attempt to identify trends in the method. All simulations were executed using the Cirq quantum software development framework [102] and simulators therein. Hamiltonians and complex circuits were further generated using code from the OpenFermion [117] libraries. Except for when mentioned, the Cirq noise models were chosen to be a constant error rate per qubit per moment, where a moment is a period of the circuit where gates occur. Equivalently, this can be thought of as an error rate per qubit per gate, but including error on idling gates as well. The noise models considered are not as complex as those typically observed in experiment (which are typically highly non-uniform, and can include crosstalk and non-Markovianity alongside other effects), but we expect our results should provide a suggestion of the mitigation power of this method in a real quantum device.

3.5.1 Givens rotation circuits for free-fermion Hamiltonians

We first test the mitigation ability of the verification protocol on an instance of a "Givens rotation circuit" of the form developed for implementing rotations of single-particle fermionic basis functions in [149]. This circuit takes the form

$$U(\vec{\theta}) = \exp\left(i\sum_{j,l} \vec{\theta}_{j,l} c_j^\dagger c_l\right), \tag{3.62}$$

where $c_j^\dagger$ and c_j are the creation and annihilation operators for a fermion on site j, and $\theta_{j,l} = \theta_{l,j}$. Such a circuit is classically simulatable, but it is a critical piece of infrastructure in quantum computing applications for quantum chemistry [66, 71, 73, 103, 142]. It is also low depth: it may be decomposed exactly by a sequence of matchgates [150], with optimal compilation in a circuit depth of exactly N. When acting on a N-qubit register prepared in the state $\prod_{n=0}^{N_f-1} X_n |0\rangle$, this may prepare an arbitrary ground state of a free-fermion Hamiltonian with N_f particles by an appropriate choice of $\vec{\theta}$. In this work, we take a simple free-fermion Hamiltonian as an example - namely a one-dimensional chain:

$$H = -t\sum_j c_j^\dagger c_{j+1} + \text{h.c.} \tag{3.63}$$

Such a Hamiltonian may be diagonalized,

$$H = V^\dagger \sum_\alpha \epsilon_\alpha c_\alpha^\dagger c_\alpha V, \tag{3.64}$$

where V here takes the same form as in Eq. 3.62. This decomposition allows immediately for the fast-forwarding of time evolution, as

$$e^{iHt} = V^\dagger e^{it\sum_\alpha \epsilon_\alpha c_\alpha^\dagger c_\alpha} V \tag{3.65}$$

$$= V^\dagger \prod_\alpha e^{it\epsilon_\alpha c_\alpha^\dagger c_\alpha} V. \tag{3.66}$$

As the Givens rotation circuits conserve particle number, the vacuum $|0\rangle$ may be used as a reference state for control-free verified estimation. A superposition of this reference state and starting state $U(\vec{\theta}) \prod_{n=1}^{N_f} X_n |0\rangle$
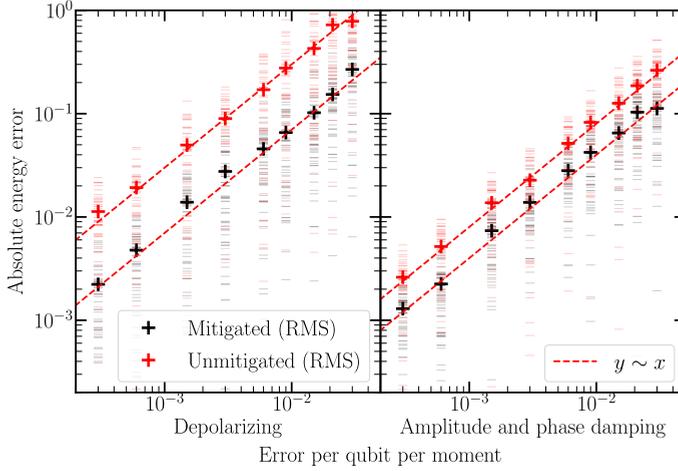
Figure 3.4: Mitigation of a 4-qubit VHA circuit via verified phase estimation. (left) Error in estimation of the energy of random states generated by the quantum approximate optimization ansatz in the critical phase of the transverse-field Ising model (Eq. 3.70) under a uniform depolarizing channel. (right) Error in the same estimation, but this time under an amplitude and phase damping model. In both plots, the RMS error (crosses) is calculated over 50 different estimations for each error rate (with randomly-chosen ansatz parameters) using either standard partial state tomography (red) or using verified control-free phase estimation. Individual data points (dashes) are additionally shown. For reference, dashed lines showing linear (red) dependence on the gate error rate are plotted.

may be prepared by acting the Givens rotation circuit on the GHZ state

$$|\text{GHZ}_{N_f}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + \prod_{n=1}^{N_f} X_n|0\rangle\right), \qquad (3.67)$$

which may itself be prepared by e.g. a chain of CNOT gates:

$$|\text{GHZ}_{N_f}\rangle = \prod_{j=N_f-1}^{1} \text{CNOT}_{j-1,j}\text{H}_0|0\rangle. \qquad (3.68)$$
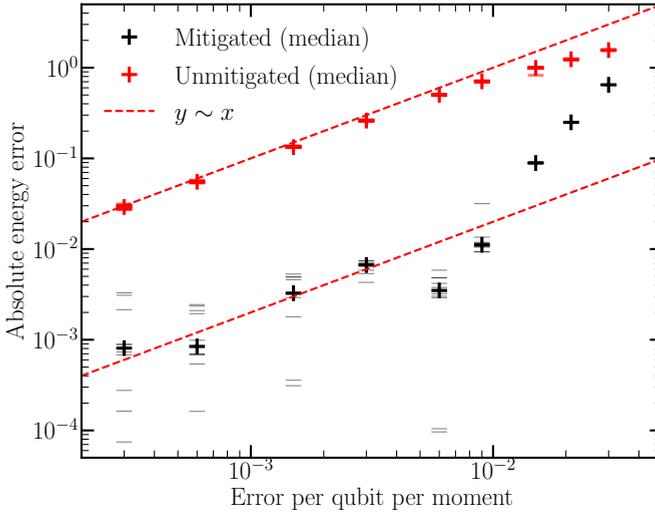
Figure 3.5: Error in estimating the ground state energy of a 4-site transverse-field Ising model (Eq. 3.70) by variational optimization of a VHA ansatz. The resulting expectation values are measured either by verified single-control phase estimation (black) or taken directly from the simulated state state (red). We plot the median (crosses) of the absolute energy error over 10 optimization attempts, each starting from a different initial point. Individual errors are plotted behind (faint dashes). Guide lines showing a linear dependence are additionally plotted (red dashed lines).

Note here the backwards product that runs left-to-right (i.e. the CNOT gate between qubit 1 and qubit 0 is executed first). Following the definitions in Sec. 3.3.4 for verified control-free phase estimation, we can write the complete preparation unitary as

$$U_{\mathrm{p}} = U(\vec{\theta}) \prod_{j=N_f-1}^{1} \mathrm{CNOT}_{j-1,j}. \tag{3.69}$$

Then, as the product of two Givens rotation circuits is itself a Givens rotation circuit [149], we may compile $VU(\vec{\theta}) = U(\vec{\theta}')$ and implement this in a single Givens rotation circuit.

The complete VPE circuit for this circuit consists of the GHZ preparation,

a single Givens rotation, a set of single-qubit z-rotations, uncomputing the Givens rotation, uncomputing the GHZ preparation, and measurement in the X or Y basis. The resulting circuit for verified phase estimation is more than twice the length of the circuit required for the unmitigated VQE. We assume here that the VQE tomography does not require any additional overhead, and directly estimate the expectation value from the simulated density matrix. For verified phase estimation, we extract the phase function from the simulated density matrix, and then process it to estimate expectation values using Prony's method. In order to not bias the final readout (which can lead to significant error in estimation), we average the rotation into the X and Y-bases over both $+\pi/2$ and $-\pi/2$ rotations 3.3.3. To simplify the analysis here, we do not include additional sampling noise. In Fig. 3.3, we plot the RMS error for two error models over a range of noise models and strengths. For each noise model and at each strength we sample 50 random choices for the initial parameters $\vec{\theta}$ (and set $t = 1$ in Eq. 3.63). In the presence of a uniform single-qubit depolarizing channel (Fig. 3.3, left), we see that the verified error displays a clear $\epsilon \sim p^2$ trend (where ϵ is the error in the final estimation, and p is the error per qubit per moment). This implies that the effect of all single errors in this noise model are suppressed by the error mitigation (or fortuitously cancel), but that pairs of errors near to each other in time may affect results. Under the effect of an amplitude and phase damping channel (Fig. 3.3, right), the suppression is even starker; we see a clear $\epsilon \sim p^3$ trend till the error drops to below 10^{-5}, providing up to four orders of magnitude gain in precision. Below 10^{-5} the error plateaus. This is due to numerical stability issues with Prony's method, and not a fundamental limit of the procedure [3]. This level of estimation error only becomes relevant after $> 10^{10} p_{\mathrm{err}}^{-2}$ individual shots have been taken (with p_{err} the probability of an error over the entire circuit). As such, we expect this to not be relevant for most experiments. The lower error rate makes some sense: amplitude damping errors can only ever reduce the number of excitations in the circuit, and so by themselves can never return to a state with non-zero overlap with $|\psi_{\mathrm{s}}\rangle$. However, the precise mode for the leading contribution to the error rate is still somewhat unclear.

---

[3]Proof of this can be found in e.g. Fig. 3.9, where phase fitting obtained a $\sim$ 100-fold reduction in this noise floor, which is typical for all simulations performed thus far.

3.5.2 The variational Hamiltonian ansatz for the transverse-field Ising model

We next attempt the verification of a completely different model and ansatz. The transverse-field Ising model (TFIM) is a well-known spin system, with Hamiltonian

$$H = J_z \sum_j Z_j + J_x \sum_j X_j X_{j+1}, \qquad (3.70)$$

where we take the sum $j+1$ modulo N (i.e. periodic boundary conditions). In one dimension, this model has a critical phase when $J_z = J_x$, making this a simple model to study interesting quantum phenomena. Exact ground states of this model may be found by the variational Hamiltonian ansatz (VHA) [54] for any values of J_x and J_z [151]. The VHA consists of alternating the Ising model and transverse field terms p times, with at each layer p the amount of time to be treated as a free variable:

$$U(\vec{\theta}) = \prod_p e^{i\theta_{p,z} \sum_j Z_j} e^{i\theta_{p,x} \sum_j X_j X_{j+1}}. \qquad (3.71)$$

(Note that for this given model the VHA is equivalent to the quantum alternating operator ansatz of Ref. [55].) The TFIM does not have any simple eigenstates, and nor does the VHA, so simple methods of control-free verified phase estimation are not available. Instead, we attempt single-control verified phase estimation. To lower the error incurred during the circuit, we perform VPE in series for every term in Eq. 3.70. Unfortunately, verification works significantly less well in this setting, as is shown in Fig. 3.4. For both noise models considered, we see a clear $\epsilon \sim p$ trend with ϵ the energy error in the final result, and p the error per qubit per moment. This suggests that errors that map the noiseless state into one with nontrivial overlap with the verified density matrix are dominant in this circuit. Regardless, we note that verification does provide an $\sim$ 8-fold improvement in error rate over the unmitigated circuit, despite the verification circuit requiring one additional qubit and being three times as long. This result is lessened in the presence of amplitude and phase damping noise, till the point where the mitigation only improves estimation by a factor of 2.

Variational optimization is well-known to mitigate certain types of coherent noise (e.g. coherent parameter drift) [82, 144]; it also appears to provide some mitigation of incoherent noise when in combination with

verified phase estimation. In Fig. 3.5, we perform a variational outer loop over the circuit studied in Fig. 3.4. Although the $\epsilon \sim p$ behaviour appears to roughly remain in the latter half of the optimization, the gain from error mitigation improves from $2 - 8\text{x}$ to around 50x, a significant improvement. We note that the optimization is no longer variationally bound - below about 10^{-2} error per qubit per moment, the results are scattered relatively evenly on either side of the true value. By contrast, in the absence of sampling noise partial state tomography results will always be variationally bound. We suspect this result may be due to the fact that slightly different circuits need to be run to measure different terms, yielding an 'effective state' that lies slightly outside the positive cone of allowed physical quantum states. Though this effect does not appear to be particularly severe in this case, further study may be needed to see it does not become an issue in larger experiments.

3.5.3 Fermionic swap networks for electronic structure Hamiltonians

As a final system for simulation, we move to studying the ability to verify molecular hydrogen on four qubits using a fermionic swap network. This ansatz was first studied in [149]; it consists of a network of two-qubit fermionic simulation gates, which take the form

$$
U_{\text{fsim}}(\theta, \phi) = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta) & i\sin(\theta) & 0 \\
0 & i\sin(\theta) & \cos(\theta) & 0 \\
0 & 0 & 0 & e^{i\phi}
\end{pmatrix}.
\tag{3.72}
$$

The parameters θ and ϕ are then left free to be optimized during the circuit. Molecular hydrogen is a simple example of the full electronic structure Hamiltonian, which takes the form

$$
H = \sum_{i,j} t_{i,j} c_i^\dagger c_j + \sum_{i,j,k,l} V_{i,j,k,l} c_i^\dagger c_j^\dagger c_k c_l.
\tag{3.73}
$$

Solving this Hamiltonian for mid-to-large system sizes ($\sim 60+$ qubits) with strong interactions is a key target application for quantum computers [64, 71, 73].

We study three different methods for verified expectation value estimation of the electronic structure Hamiltonian. Following a transformation from fermionic to qubit operators, Eq. 3.73 we first consider a decomposi-

tion over single Pauli operators for single-control VPE, as was performed for the transverse-field Ising model in Sec. 3.5.2. However, in order to perform control-free VPE on these terms, we require a reference state. Individual fermionic terms in Eq. 3.73 are number-conserving, so the fermionic vacuum is a good reference state for these, but this is not the case for individual Pauli terms. To circumvent this problem, we split Eq. 3.73 into fermionic terms (summed with their Hermitian conjugate), and decompose these into Pauli operators. (One can check that the resulting Pauli operators commute, and so their time evolution may be easily fast-forwarded.) The VPE circuits in both of the above methods are $3 - 4$ times the depth of the original VQE.

Alternatively, by performing a low-rank factorization of the Coulomb operator, we may write H in the form [142]

$$H = \sum_{i,j} t'_{i,j} c_i^\dagger c_j + \sum_l U_l^\dagger \left[\sum_\alpha E_\alpha^{(l)} c_\alpha^\dagger c_\alpha \right]^2 U_l \tag{3.74}$$

$$= H^{(0)} + \sum_l H^{(l)}, \tag{3.75}$$

Where the U_l are single-particle basis changes that may be implemented via Givens rotation circuits. Each such term in this factorization is fast-forwardable. $H^{(0)}$ is a free-fermion Hamiltonian and may be simulated via the methods discussed earlier in this section. The interacting factors $H^{(l)}$ may also be diagonalized by diagonalizing the single-particle $t_{i,j}^{(l)}$ matrices. One finds

$$e^{iU_l^\dagger H^{(l)} U_l t} = U_l^\dagger \prod_{\alpha\beta} E^{it E_\alpha E_\beta c_\alpha^\dagger c_\alpha c_\beta^\dagger c_\beta} U_l, \tag{3.76}$$

which may be easily implemented on superconducting hardware, as $e^{it E_\alpha E_\beta c_\alpha^\dagger c_\alpha c_\beta^\dagger c_\beta}$ is realised by a C-Phase gate. All of the above Hamiltonians, as well as the fermionic swap network itself, conserve particle number, and so we may again use the vacuum as a reference state for verified control-free quantum phase estimation. We do not consider the single-control version for comparison in this case. The resulting circuit is over ten times as long as the VQE itself, as we are unable to compile the final basis rotation into the ansatz.

The mitigation power of VPE differs vastly between the different choices of decomposition used, and the different noise models chosen. In Fig. 3.6, we plot the effect of mitigating depolarizing, and amplitude and phase

3
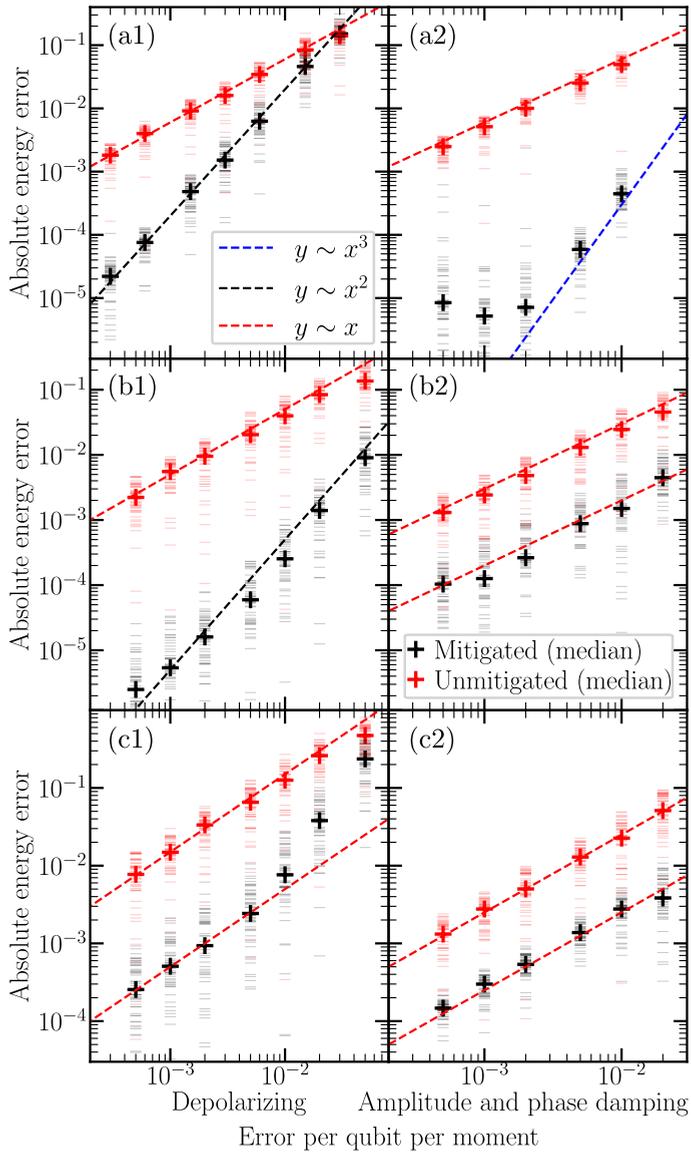
Figure 3.6: Mitigation of a 4-qubit fermionic swap network via verified phase estimation. (Continues on following page)

Figure 3.6: (continued) Three different VPE protocols are explored — a low-rank factorization (top row — a1, a2), a control-free number-conserving Pauli decomposition (middle row — b1, b2), and a single-control Pauli decomposition (bottom row — c1, c2). Details of all decompositions are given in the text. The low-rank factorization was studied for the H_2 Hamiltonian at the equilibrium bond distance with a swap network of depth 4, while the other two models were studied at a bond distance of 2 Angstrom with a swap network of depth 6. All protocols are tested under depolarizing (left column) and amplitude and phase damping (right column) noise models. In all plots, the median error (crosses) is calculated over 50 different estimations for each error rate using either standard partial state tomography (red) or using verified control-free phase estimation. Individual data points (dashes) are additionally shown. For reference, dashed lines showing linear (red), quadratic (black) and cubic (blue) dependence on the gate error rate are plotted.

3

damping channels, using the three decompositions described above. We see that control-free [Fig. 3.6(a1-b2)] VPE typically outperforms single-control VPE [Fig. 3.6(c1,c2)], despite the single-control VPE circuits being in all cases smaller (due to the lack of coherent state preparation). Under a depolarizing noise model, both control-free VPE implementations [Fig. 3.6(a1, b1)] demonstrate a second-order sensitivity to the physical qubit error rate, consistent with the previous results in Fig. 3.3. In this case, the Pauli decomposition clearly outperforms the low-rank factorization, which we attribute to the large reduction ($\sim 2-3\times$) in total circuit depth. However, although the low-rank factorization repeats the third-order sensitivity to amplitude and phase damping seen in Fig. 3.3 [Fig. 3.6(a2)], this is not observed in the Pauli decomposition case [Fig. 3.6(b2)]. We investigate this further in App. 3.F, and find that this first-order error can be traced back to the verified estimation of a single term — the two-body interaction term. We attribute this to the fact that the time evolution circuit for this term breaks number conservation (which is not the case for any other term in the sum), which makes it more susceptible to amplitude damping noise. Understanding this feature in detail, and determining whether better circuit optimizations exist, are clear targets for future research. In any case, all three implementations of VPE studied show at least an order of magnitude improvement compared to partial state tomography, and in some cases up to three orders of magnitude improvement, demonstrating the power of this technique.

3.5.4 Sampling costs

In a realistic experiment, direct estimation of any expectation value requires repeatedly re-preparing the target state and measuring in an appropriate basis to accumulate statistics on the probability of seeing a given 0 or 1 measurement. In verified phase estimation, this repetition must be performed instead on the control qubit (for single-control) or target qubit (for control-free) to accumulate the phase function. Re-preparation is necessary between subsequent measurements, as such a measurement collapses the global wavefunction, erasing the information about the probability to be estimated. This implies that each repetition carries substantial cost, and the rate of convergence of error estimation is a critical bottleneck in any variational algorithm. Although one might expect quantum phase estimation to speed up this estimation (which has been proposed previously [139]), this is only the case when one is estimating eigenvalues of the target Hamiltonian in a specific QPE instance. We wish to divide up our Hamiltonian for fast-forwarding purposes, and in most cases the resulting terms will not be simultaneously diagonalizable, so no set of mutual eigenstates will exist; instead, the results of Sec. 3.3.3 will hold. Furthermore, as our expectation value estimation requires to sum over multiple different amplitudes, we should not expect this to improve over the cost of partial state tomography (which requires non-commuting terms to be measured on separate preparations of the state). The error in expectation value estimation will further depend on the type of classical post-processing used.

In Fig. 3.7, we compare the convergence of two types of classical post-processing to that of standard partial state tomography. We perform this simulation on the 4-spin VHA-TFIM system studied in Fig. 3.4 and Fig. 3.5, on a representative point in the spectrum (the error-free variational minimum). We do not perform any measurement grouping or parallelization strategies for either method, and instead report our results as a function of the number of measurements per Pauli operator. The first method (green) assumes knowledge about the eigenvalues of the fast-forwarded Hamiltonians, in which case one need only fit the amplitudes, while the second (blue) first estimates the eigenvalues using Prony's method before fitting the amplitudes to the resulting signal. (We compensate for the presence of spurious phases in Prony's method by a slight adjustment described in App. 3.C.) All methods of estimation are seen to converge at a rate $\epsilon \sim M^{-1/2}$, where ϵ is the estimation error and M is the number of samples taken.

We see that using the prior knowledge of the phases gives a significant

102

advantage in convergence, with the resulting error rate being almost an order of magnitude worse when using Prony's method. This advantage persists in the presence of a depolarizing channel (1% error rate), although the convergence of all methods flattens as they approach the sampling-noise-free estimation value. We note that both classical post-processing methods converge to the same result here, as expected. It is unclear whether the good overlap between the unverified circuit and the phase fitting method is due to them both achieving a lower bound for convergence or just coincidence. Further investigation here would be a good target for future work. The addition of noise makes convergence more costly. This increase can be bounded below by removing the fraction of experiments where at least one error has occurred (as we are at best effectively removing these results). Confirming this trend would also be a good target for future work.

3.6 Conclusion

In this work, we presented a new method for error mitigation, based on verification of the system register in a single-control quantum phase estimation routine. We further extended this method to a scheme for verification of control-free quantum phase estimation. By writing a complex Hamiltonian as a sum of fast-forwardable parts and using this technique to estimate the expectation value of each part, this becomes a powerful error mitigation tool for near-term experiments such as variational algorithms. Errors that take the system away from the small verified subspace do not affect the mitigated QPE results (at the cost of requiring additional repetitions of the circuit). We performed numerical studies of this error mitigation capability of the verification protocol on three different systems, finding the suppression of all single depolarizing errors when a Givens rotation circuit or a fermionic swap network prepare random states of a small fermionic system. The suppression is further magnified in the presence of amplitude and phase damping, resulting in a gain of up to four orders of magnitude in accuracy. For a simulation of the transverse-field Ising model the error suppression is less pronounced. However, we find that variational optimization improves the error mitigation to a gain in accuracy of about 50-fold. We further demonstrated that the combination of variational optimization and verification mitigates against constant control error (which is not naturally mitigated by the verification itself). However, we found that the choice of post-processing technique in the classical post-processing may affect the estimation error by a factor of 10
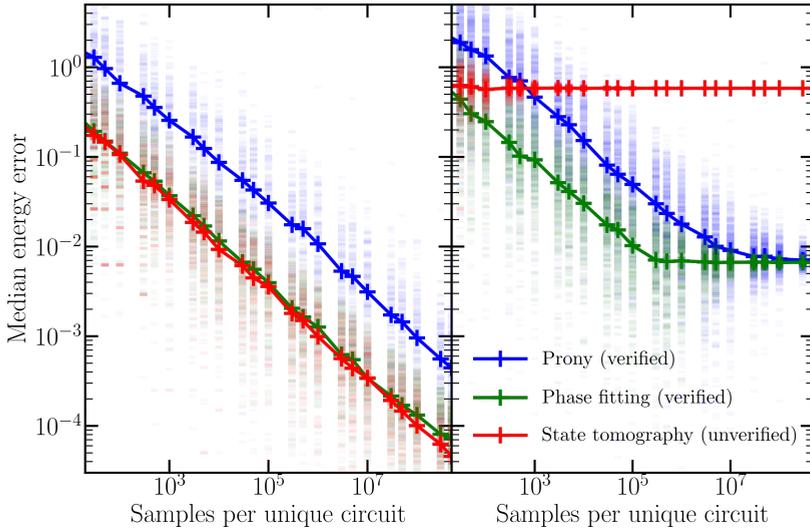
Figure 3.7: Convergence of the estimation of a single point in a 4-site transverse-field Ising model with the number of samples taken, using verified phase estimation processed either with Prony's method (blue) or by fitting known phases to the phase function (green), or standard partial state tomography (red) on individual Pauli terms. (Left) convergence in the absence of error. (Right) convergence in the presence of 1% depolarizing error per qubit per moment. In each subfigure we plot the median energy error (crosses and lines) over 200 simulations, which are plotted themselves behind (faint dashes).

in the presence of sampling noise.

Though verified phase estimation as presented already appears to be one of the most powerful error mitigation techniques available to NISQ-era quantum computing, further avenues for optimization exist. The wide range of possible options for verification, how to divide the Hamiltonian, and classical post-processing method all provide metaparameters which we have not yet determined how to optimize for any specific problem. Furthermore, circuits which quickly scramble errors would appear to make verification more reliable. Whether this observation can be used for meaningful optimization is a clear target for future work. Similarly, as errors need to have the instantaneous state as a near-eigenstate to not fail verification, the errors that verified phase estimation is most-susceptible to

must commute, and could potentially be corrected with a classical error correcting code. As these codes require much less overhead than full-blown QEC, this may be a practical method to ensure universal suppression of single-qubit errors. Future work could also investigate whether verified phase estimation may be combined efficiently with other error mitigation techniques. More generally, it would be timely to benchmark the zoo of error mitigation techniques against one another, and determine which combination of techniques works best in a range of situations.

3.A Error analysis

Let us formalize the ideas outlined in Sec. III.C of the main text by considering how the verified and unverified Hilbert spaces $\mathcal{H}^{(v)}$ and $\mathcal{H}^{(f)}$, and the verified and unverified ensembles $\rho^{(v)}$ and $\rho^{(f)}$ within them, evolve over the course of a noisy quantum circuit. (We remind the reader here that $\rho^{(v)}$ and $\rho^{(f)}$ are not normalized, hence our use of the word 'ensemble' rather than 'state'.) We will then attempt to provide some mechanisms for the observed scaling laws in Sec. V of the main text. At the end of the VPE circuit, the verified Hilbert space $\mathcal{H}^{(f)}$ is spanned by the two verified basis states. In single-control VPE, these are $|0\rangle|0\rangle$ and $|1\rangle|0\rangle$, while in control-free VPE these are $|0\rangle$ and $|\vec{1}_T\rangle$. Let us label these $|0_v\rangle$ and $|1_v\rangle$ respectively, and then we may define the verified Hilbert space as

$$\mathcal{H}^{(v)} = \mathrm{Span}\{|0_v\rangle, |1_v\rangle\}, \tag{3.77}$$

and the verified ensemble as

$$\rho^{(v)}(t) = P_v \rho(t) P_v, \quad P_v = |0_v\rangle\langle 0_v| + |1_v\rangle\langle 1_v|. \tag{3.78}$$

The system state ρ here is the state at the end of the VPE circuit, let us now consider how the system evolves to get here. This evolution is not a function of the simulated time t, as we may use entirely different circuits to estimate the phase function $g(t)$ and $g(t')$. Instead, we must frame the evolution of the state on the quantum device over the course of the VPE circuit in terms of the device time τ. That is, let us fix t, and assume that the circuit that implements $U = e^{iHt}$ is split into a set of discrete moments $U(\tau)$ (with the last moment occuring at time $\tau_{\max}$),

$$U = \prod_{\tau=0}^{\tau_{\max}} U(\tau), \tag{3.79}$$

where each moment consists of a set of gates acting in parallel

$$U(\tau) = \prod_i U(\tau)_i. \tag{3.80}$$

This is how circuits are represented in the cirq quantum programming framework [102], and is a good way of approximating the behaviour of a real quantum circuit.

To best understand how noise and verification work together, we must move to the interaction picture, or rather a rotating reference frame. In the Schrödinger picture, the system begins entirely within $\mathcal{H}^{(\mathrm{v})}$, as in all cases it is initialized in $|0_{\mathrm{v}}\rangle = |0\rangle$ and immediately rotated to $\frac{1}{\sqrt{2}}(|0_{\mathrm{v}}\rangle + |1_{\mathrm{v}}\rangle)$. It then evolves out of $\mathcal{H}^{(\mathrm{v})}$ as we prepare, evolve, and un-prepare the system, even in the absence of error. However, for us it is more helpful to consider the states that will be rotated into $\mathcal{H}^{(\mathrm{v})}$ at the end of the circuit. This may be achieved by re-defining the verified basis states in the reference frame

$$|0_{\mathrm{v}}\rangle \to \left(\prod_{\tau' > \tau} U(\tau') \right)^{-1} |0_{\mathrm{v}}\rangle \tag{3.81}$$

$$|1_{\mathrm{v}}\rangle \to \left(\prod_{\tau' > \tau} U(\tau') \right)^{-1} |1_{\mathrm{v}}\rangle. \tag{3.82}$$

(This a slightly non-standard choice of reference frame, as we are shifting backwards in time from the final state, rather than forwards in time from the initial state, but it makes our error analysis far easier.) In the absence of error, this is the Heisenberg picture: our system remains in the state

$$\rho = |\rho\rangle\langle\rho| \tag{3.83}$$

$$|\rho\rangle = \frac{1}{\sqrt{2}}|0_{\mathrm{v}}\rangle + \frac{g(t)}{\sqrt{2}}|1_{\mathrm{v}}\rangle + \sqrt{\frac{1 - |g(t)|^2}{2}}|\rho^{(\mathrm{f})}\rangle, \tag{3.84}$$

throughout the entire circuit. Here $|\rho^{(\mathrm{f})}\rangle$ is the fraction of the state that will eventually fail verification —

$$\rho^{(\mathrm{f})} = |\rho^{(\mathrm{f})}\rangle\langle\rho^{(\mathrm{f})}|. \tag{3.85}$$

(Recall here that $|\rho^{(f)}\rangle \neq 0$ even in the absence of noise, but this fraction of the state does not contribute to the phase function — $\langle\rho^{(f)}|X_c + iY_c|\rho^{(f)}\rangle =$

0.) We may project our system at any device time τ into the verified Hilbert space via Eq. 3.78, but with the basis states in their rotating reference frame (Eq. 3.81 and Eq. 3.82).

Noise may be added to the above by treating it as a perturbation and switching to the interaction picture. Without loss of generality, we may say the effect of noise is to shift the unitary $U(\tau)$ at each moment

$$U(\tau) \rightarrow R_{U(\tau)}(1 - p_\tau + p_\tau E_\tau), \tag{3.86}$$

where R_U is the process map associated with a unitary U

$$R_U[\rho] = U^\dagger \rho U, \tag{3.87}$$

E_τ is the process map associated with errors during the moment τ, and p_τ is the probability of any such errors occurring. In the interaction picture, the action of the circuit has been shifted into our basis states, and $U(\tau) = 1$. So, we may write our final state in the presence of error as

$$\rho^{(\mathrm{err})} = \frac{1}{\mathcal{N}} \left\{ \rho + \sum_{\tau_0 < \tau} p'_{\tau_0} E_{\tau_0}[\rho] \right. $$
$$\left. + \sum_{\tau_1 < \tau_0 < \tau} p'_{\tau_0} p'_{\tau_1} E_{\tau_0}\left[E_{\tau_1}[\rho]\right] + \dots \right\}, \tag{3.88}$$

where $\mathcal{N}$ is a normalization coefficient

$$\mathcal{N} = \prod_\tau \frac{1}{1 - p_\tau}, \tag{3.89}$$

and p'_τ are the rescaled probabilities

$$p'_\tau = \frac{p_\tau}{1 - p_\tau}. \tag{3.90}$$

If desired, one can recognise this also as a discrete form of the well-known time-ordered integrals in quantum mechanics — a time-ordered sum

$$\rho^{(\mathrm{err})} = \frac{1}{\mathcal{N}} \mathcal{T} \exp \left(\sum_\tau p'_\tau E_\tau \right) [\rho], \tag{3.91}$$

where $\mathcal{T}$ is the time-ordering operator. Our projection onto the verified subspace is linear, so we may consider it on each of the individual terms

in the sum. Assuming $p'(\tau)$ is small for all τ, the first-order corrections to $\rho^{(v)}$ occur from errors E_τ during a single timestep. These corrections take the form

$$P_{\mathrm{v}} p'_\tau E_\tau[\rho] P_{\mathrm{v}} = p'_\tau \begin{pmatrix} p_{0,\tau} & \frac{1}{2} g_\tau^{(\mathrm{err})}(t) \\ \frac{1}{2} g_\tau^{(\mathrm{err})\dagger}(t) & p_{1,\tau} \end{pmatrix}, \qquad (3.92)$$

where

$$p_{0,\tau} = \langle 0_{\mathrm{v}} | E_\tau[\rho] | 0_{\mathrm{v}} \rangle \qquad (3.93)$$

$$p_{1,\tau} = \langle 1_{\mathrm{v}} | E_\tau[\rho] | 1_{\mathrm{v}} \rangle \qquad (3.94)$$

$$g_\tau^{(\mathrm{err})}(t) = \langle 1_{\mathrm{v}} | E_\tau[\rho] | 0_{\mathrm{v}} \rangle \qquad (3.95)$$

The off-diagonal element in this matrix gives the contribution to the phase function $g(t)$

$$g(t) \to \frac{1}{\mathcal{N}} g(t) + \frac{1}{\mathcal{N}} \sum_\tau p'_\tau g_\tau^{(\mathrm{err})}(t). \qquad (3.96)$$

One may generalize this to higher-order terms. For example, the second-order contribution to the error takes the form

$$\frac{1}{\mathcal{N}} \sum_{\tau_0 < \tau_1} p'_{\tau_0} p'_{\tau_1} \langle 1_{\mathrm{v}} | E_{\tau_1}[E_{\tau_0}[\rho]] | 0_{\mathrm{v}} \rangle \qquad (3.97)$$

The mitigation power from verification requires two conditions: that the dependence of the normalization $\mathcal{N}$ on the simulated time t is simple, and that the off-diagonal error contributions (Eq. 3.95) are small. We expect both conditions to often be the case. The positivity of $E_\tau[\rho]$ implies that

$$g_\tau^{(\mathrm{err})} + p_{0,\tau} + \sum_{n=1}^{2^N - 2} \langle \rho_n^{(\mathrm{f})} | E_\tau[\rho] | 0_{\mathrm{v}} \rangle \le 1, \qquad (3.98)$$

where $|\rho^{(\mathrm{f})}\rangle$ is an appropriately chosen basis for $\mathcal{H}^{(\mathrm{f})}$. On average all terms are equally-weighted, so $g_\tau^{(\mathrm{err})} \sim 2^{-N}$. As such, negligible $g_\tau^{(\mathrm{err})}$ should be the norm rather than the exception; we need reason to expect that the error channel E_τ will not scatter us out of the verified subspace. If $g_\tau^{(\mathrm{err})} = 0$, the effect of E_τ on $g(t)$ occurs via the damping by $\mathcal{N}$, which itself may depend on t. However, $\mathcal{N}$ depends only on the rate at which errors occur, and is oblivious to their specific form. If a Hamiltonian is fast-forwardable, e^{iHt} may be implemented in time constant in t. Then assuming a constant error rate per moment of the circuit, our phase function is dampened by a

constant amount,

$$g^{\mathrm{err}}(t) = \frac{1}{\mathcal{N}} g(t), \qquad (3.99)$$

which may be corrected for by renormalization (Eq.30 of the main text). If a Hamiltonian is not fast-fowardable, e^{iHt} must take real time $\tau_{\max}(t) = \mathcal{O}(t)$ to simulate to constant error. Assuming this is the case, and that we have a constant error rate per moment of the circuit, the damping from each possible error E_τ is multiplicative, and our estimation takes the form

$$g^{\mathrm{err}}(t) = e^{-t/T_1} g(t). \qquad (3.100)$$

Here, T_1 is defined as the (simulated) time t over which enough errors E have accumulated that

$$\frac{1}{\mathcal{N}(\tau)}(E) = e^{-1}. \qquad (3.101)$$

This constant damping may be considered an imaginary shift to the eigenvalues E_j; $E_j \to E_j + \frac{1}{T_1}$. It may be removed by classical postprocessing techniques [124, 130, 137]. However, the shrinking of the signal increases the sampling requirements to estimate $g(t)$ exponentially in t.

Although random error channels are exponentially suppressed by verification (following Eq. 3.98), realistic error models are biased, and may apply undesired phases to $g_\tau^{\mathrm{err}}(t)$ instead of setting it to 0. The density matrix in Eq. 3.92 is not normalized, but it must be positive, which implies

$$|g_\tau^{(\mathrm{err})}|^2 < \frac{p_{0,\tau} p_{1,\tau}}{p_{0,\tau} + p_{1,\tau}}. \qquad (3.102)$$

This means that errors must either fail to scatter *both* $|0_{\mathrm{v}}\rangle$ and $|1_{\mathrm{v}}\rangle$, or rotate between these states and the failed state $|\rho^{(f)}\rangle$. When control-free methods are used, $|0_{\mathrm{v}}\rangle$ is separated from $|1_{\mathrm{v}}\rangle$ and $|\rho^{(f)}\rangle$ by highly non-local excitations, which are non-physical error channels. However, when single-control methods are used, $|0_{\mathrm{v}}\rangle$ is coupled to $|1_{\mathrm{v}}\rangle$ and $|\rho^{(f)}\rangle$ by control qubit errors. These control qubit errors deform the Bloch sphere defined by $|0_{\mathrm{v}}\rangle = |0\rangle|0\rangle$ and

$$g(t)|1_{\mathrm{v}}\rangle + \sqrt{1 - |g(t)|^2}|\rho^{(f)}\rangle = |0\rangle|\vec{1}\rangle. \qquad (3.103)$$

When this deformation is asymmetric around the z-axis, or a rotation, $g(t)$ may be quickly corrupted[4]. However, symmetric noise (such as a

---

[4]We have observed this for instance due to T_1 decay on the control qubit between final rotation and readout. However, we can correct for this easily by measuring in

depolarizing channel, or T_1 or T_2 channels during the bulk of the circuit) can be seen to simply dampen $g(t)$ in an identical manner to $\mathcal{N}$. That is, the dampening will depend only on the rate at which these errors occur. Such dampening will be cancelled by renormalization, as observed in Fig. 3.9.

Errors that do not rotate between $|0_v\rangle$ and $|1_v\rangle$, but still contribute non-trivially to $g_\tau^{(\mathrm{err})}(t)$ to first order must have both $|0_v\rangle$ and $|1_v\rangle$ as approximate eigenstates of the error channel. This suggests a reason why control-free VPE is more noise-robust to noise than single-control VPE: the starting and reference states are very different when looked at locally, which makes it less likely that a single local error will have both states as near-eigenstates. It also suggests a reason why we might expect the suppression of errors to only second-order: if the same error occurs in subsequent moments (in a local frame), and the basis states $|0_v\rangle = |0_v(\tau)\rangle$ have not evolved significantly between these moments, the second error will almost (but not completely) cancel out the first, driving the system back into the verified subspace in an uncorrectable manner. This implies that a circuit which more quickly scrambles the basis states $|0_v\rangle$ and $|1_v\rangle$ between moments should be less susceptible to error than one where the states evolve slowly. Understanding the dynamics of these noisy circuits in more detail is a clear target for future work.

3.B Effect of parallelizing QPE

In this appendix we investigate the phase function obtained during the parallel estimation of multiple commuting Hamiltonians, and demonstrate that the resulting expectation values from this estimation are not affected by the parallelization and verification process combined. Let us consider the case where we have two commuting Hamiltonians H_0, H_1. In this case, we may write a simultaneous eigenbasis $|E_j\rangle$ for both Hamiltonians — $H_b|E_j\rangle = E_j^{(b)}|E_j\rangle$. Let $|\psi_s\rangle = \sum_j a_j|E_j\rangle$, and we may calculate the controlled-time-evolved global state $|\Psi(t)\rangle$ to be

$$\sum_j a_j(|0\rangle + e^{iE_j^{(0)}t}|1\rangle)(|0\rangle + e^{iE_j^{(1)}t}|1\rangle)|E_j\rangle. \qquad (3.104)$$

---

the opposite basis 50% of the time

Tracing out control qubit 1 obtains the following reduced density matrix for the system + control qubit 0

$$\sum_{j,j'} a_j a_{j'}^* \left[1 + e^{i\left(E_j^{(1)} - E_{j'}^{(1)}\right)} \right]$$

$$\times \left[\left(|0\rangle + e^{iE_j^{(0)}t}|1\rangle \right) |E_j\rangle\langle E_{j'}| \left(\langle 0| + e^{-iE_{j'}^{(0)}t}\langle 1| \right) \right]. \tag{3.105}$$

The issue here then comes from this additional factor $\left[1 + e^{i\left(E_j^{(1)} - E_{j'}^{(1)}\right)} \right]$ at the front. Note that (as we should expect) this goes away upon tracing out the system register, as the trace over $|E_j\rangle\langle E_j'|$ yields (dropping all additional terms in the above expression).

$$\sum_l \langle E_l|E_j\rangle\langle E_{j'}|E_l\rangle = \delta_{j,j'} \tag{3.106}$$

However, post-selection implies that we take the expectation value with regards to $|\psi_{\mathrm{s}}\rangle$, obtaining

$$\langle \psi_{\mathrm{s}}|E_j\rangle\langle E_{j'}|\psi_{\mathrm{s}}\rangle = a_j^* a_{j'}. \tag{3.107}$$

The off-diagonal element of the control qubit density matrix can then be found to be

$$\frac{1}{2} \sum_{j,j'} |a_j|^2 |a_{j'}|^2 \left(1 + e^{i(E_j^{(1)} - E_{j'}^{(1)})t} \right) e^{iE_j^{(0)}t} \tag{3.108}$$

$$= \frac{1}{2} \sum_j |a_j|^2 e^{iE_j^{(0)}t}$$

$$+ \frac{1}{2} \sum_j |a_j|^2 e^{i(E_j^{(1)} + E_j^{(0)})t} \sum_{j'} |a_{j'}|^2 e^{-iE_{j'}^{(1)}t}. \tag{3.109}$$

One can see that this is a linear combination of products of the phase functions of H_0, H_1, and $H_0 + H_1$. In theory the eigenvalues $E_j^{(0)}$ and amplitudes squared $|a_j|^2$ are still present in this function, and could be extracted via classical postprocessing. However, the $\frac{1}{2}$ coefficient implies we need 4 times as many single-shot experiments for the estimation of $|a_j|^2$ to the same error (compared to a standard QPE experiment for H_0). Extending this to $L > 2$ summands, the off-diagonal for the sth control

qubit can be written:

$$\frac{1}{2^L} \sum_{j,j'} |a_j|^2 |a_{j'}|^2 e^{iE_j^{(s)}t} \prod_{s' \neq s} \left(1 + e^{i(E_j^{(s')} - E_{j'}^{(s')})t}\right), \tag{3.110}$$

and we see that the signal corresponding to 'just' $g(t)$ is exponentially small. However, all is not lost. Inspecting the form of Eq. 3.110, we see that we may expand this as a sum of $2^L J^2$ separate (possibly degenerate) spurious energies $F_{v,j,j'}^{(s)}$, indexed by a L-bit binary integer v and the original j and j' indices

$$F_{v,j,j'}^{(s)} = E_j^{(s)} + \sum_{s' \neq s} v_{s'}(E_j^{(s')} - E_{j'}^{(s')}), \tag{3.111}$$

with corresponding (v-independent) spurious amplitudes

$$B_{j,j'} = \frac{1}{2^L} |a_j|^2 |a_{j'}|^2. \tag{3.112}$$

(Note that as stated these energies are automatically at least doubly-degenerate as v_s does not appear in the equation for $F_{v,j,j'}^{(s)}$.) If we then calculate the weighted average of the $F_{v,j,j'}^{(s)}$ (which is what we would do if we processed the signal as if the parallelization had not occurred), we find

$$\begin{aligned} \sum_{v,j,j'} B_{j,j'} F_{v,j,j'}^{(s)} = &\frac{1}{2^L} \sum_{v,j,j'} |a_j|^2 |a_{j'}|^2 E_j^{(s)} \\ &+ \frac{1}{2^L} \sum_{v,j,j'} B_{j,j'} \sum_{s' \neq s} v_{s'} E_j^{(s')} \\ &- \frac{1}{2^L} \sum_{v,j,j'} B_{j,j'} \sum_{s' \neq s} v_{s'} E_{j'}^{(s')}. \end{aligned} \tag{3.113}$$

As j and j' are just dummy indices, and as $B_{j,j'} = B_{j',j}$, the last two terms cancel, and as $\sum_v = 2^L$ and $\sum_{j'} |a_{j'}|^2 = 1$, we have

$$\sum_{v,j,j'} B_{j,j'} F_{v,j,j'}^{(s)} = \sum_j |a_j|^2 E_j^{(s)} = \langle H_s \rangle. \tag{3.114}$$

This implies that expectation values may be extracted via parallel verified phase estimation, even though the signal itself may be significantly more complex. For the case of Pauli H_s operators, the spectrum $F_{v,j,j'}^{(s)}$ is highly

degenerate — it is the set of odd integers $\{-2L+1, -2L+3, \ldots, 2L-3, 2L-1\}$. (This parallels the spectrum of a spin-$\frac{2L-1}{2}$ operator, which one might not expect following Hund's rules for the combination of spin-$\frac{1}{2}$ systems, which is curious.) This must be taken into account when signal processing by amplitude-fitting, as one would otherwise miss components of the energy. However, the overhead for this is only linear in the number of simultaneously-estimated terms.

3.C Compensation for spurious eigenvalues due to sampling noise

When quantum phase estimation is used to estimate eigenvalues as well as amplitudes to sum together to give an expectation value (Eq. 53 of the main text), finite sampling noise introduces a small bias to this estimation that may be cancelled. This bias does not come from the QPE itself. The sampling noise has a white spectrum which is invariant under a Fourier transform, so classical post-processing of a noisy spectrum yields a set of spurious eigenvalue/amplitude pairs evenly distributed around the circle. However, in order to evaluate Eq. 53 of the main text, we have to make a branch cut in this circle. The resulting terms then average to bias the signal by a term $\Delta_{\mathrm{bias}} = \langle H \rangle - \overline{\langle H \rangle}$ towards the center of the resulting region. (Here, $\langle H \rangle$ is the true expectation value, and $\overline{\langle H \rangle}$ that measured naively.) For example, if we assume all eigenvalues $E_j \in [-\pi, \pi]$, this biases the signal towards zero. This bias is dependent on both the number of steps K, and the number of samples M used in the QPE process. Numerically, we find (Fig. 3.8):

$$\Delta_{\mathrm{bias}} = -\langle H \rangle \times (K-2)^{\frac{1}{2}} M^{-\frac{1}{2}}. \qquad (3.115)$$

Inverting this obtains

$$\langle H \rangle = \overline{\langle H \rangle} \left[1 + (K-2)^{\frac{1}{2}} M^{-\frac{1}{2}} \right]^{-1}, \qquad (3.116)$$

which is used in the estimation in Sec. V.D.

3
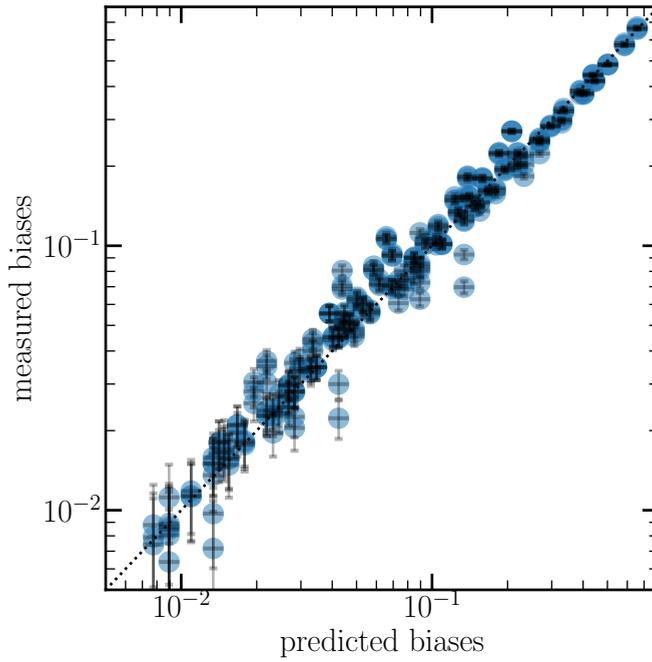
Figure 3.8: Predicted (Eq. 3.115) vs found bias from estimating expectation values using Prony's method.

3.D Demonstration of immunity to control noise in single-control VPE

One might expect that the discrepancy between the scaling of the error mitigation power of the control-free and single-control circuits seen throughout this work comes from accumulation of errors on the control qubit alone. In this appendix, we show that this is not the case. In Fig. 3.9, we see that removing all errors on the control qubit does little to reduce the total error rate (black crosses), whilst a model with noise only on the control qubit achieves an error limited by our use of Prony's method for post-processing. In App. 3.A we argue that the increased error suppression from control-free VPE comes from the large separation between reference and starting states. Errors will be removed by verification unless they maintain coherence between these states, which these error models fail to do.

3.E Use of a variational outer loop to mitigate constant unitary noise

One of the main uses of expectation values $\langle H \rangle$ in quantum experiments is to use them as a cost function in a variational outer loop. Optimizing the parameters $\vec{\theta}$ in a preparation unitary $U_\mathrm{p}(\vec{\theta})$ to minimize the expectation value of the prepared state $|\Psi(\vec{\theta})\rangle = U_\mathrm{p}(\vec{\theta})|0\rangle$ then gives an approximation for the true ground state of H. The variational optimization process is itself known to be robust against certain types of error [82, 144], in particular control errors. These occur when a signal meant to implement a gate $G(\theta)$ either drifts or is distorted and instead implements $G(\theta')$. As this error is often repeated throughout an experiment, i.e. every instance of G is miscalibrated by a similar amount, it will be repeated throughout the experiment. Verification can only correct single errors, and as such is not targeted for this type of noise. By contrast, the dominant source of errors in a VQE are often the incoherent errors that verification is designed to target. As such, verification and variational optimization provide cumulative mitigation by targeting sources of error the other lets through.

   To demonstrate the combined mitigation effects, we use verified control-free phase estimation of a Givens rotation circuit in the inner loop of a variational quantum eigensolver. In order to prevent oversimplifying the problem, we add a next-nearest-neighbour coupling and on-site potential
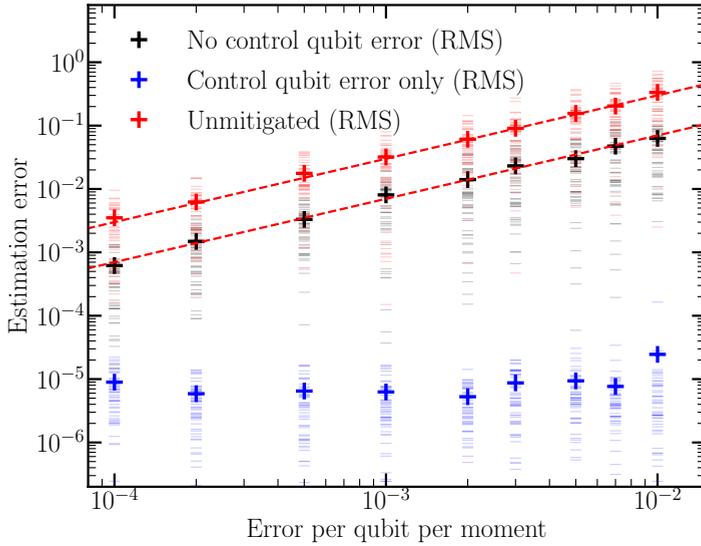
Figure 3.9: Mitigation of the same 4-qubit VHA circuit as in Fig. 4 of the main text, but with either depolarizing noise on only the system register (black) or only on the control qubit (blue). This is compared with the error in estimation using partial state tomography instead of VPE (red). For each dataset, the RMS error (crosses) is plotted over 50 different estimations for each error rate (with randomly-chosen ansatz parameters), and individual data points are plotted as dashes behind. For reference, dashed lines showing linear (red) dependence on the gate error rate are plotted.

to the Hamiltonian in Eq. 66 of the main text, yielding

$$H = H_1 + H_2 \tag{3.117}$$

$$H_1 = -t_1 \sum_{j=1}^{N} c_j^\dagger c_{j+1} + \text{h.c.} \tag{3.118}$$

$$H_2 = -t_2 \left(\sum_{j=1}^{N} c_j^\dagger c_{j+1} + \text{h.c.} + \sum_{j=1}^{N} c_j^\dagger c_j \right), \tag{3.119}$$

and estimate expectation values for H_1 and H_2 separately. Here, we again
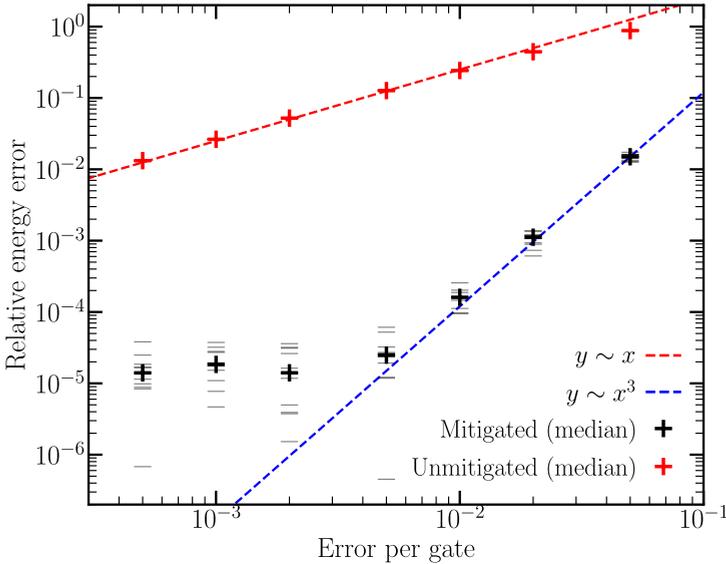
Figure 3.10: Error in estimating the ground state energy of a free-fermion system (Eq. 66 of the main text) of 4 fermions (on four qubits), using control-free verified phase estimation and a VQE. Noise model is a mixture of amplitude and phase damping and constant two-qubit control error (details in text). Median absolute errors for both verified estimation (black crosses) and standard partial state tomography (red crosses) are calculated over 10 different optimization attempts. Individual simulations are plotted behind (faint dashes) Each optimization started from a different parameter set and had different control rates set. Linear (red dashed) and cubic (blue dashed) lines are shown as guides.

take periodic boundary conditions for a $N = 4$-site system (i.e. all sums in indices are taken modulo 4), and fix $t_1 = 1$, $t_2 = 0.5$. This ensures that the ground state of the system is neither a ground state of H_1 or H_2 (in which case the compiled variational ansatz and basis rotation would cancel to become an identity circuit). For a simple model combining control error and incoherent noise, we fix p, draw a random offset $x_i \in [-\frac{p}{\pi}, \frac{p}{\pi}]$ for each two-qubit ISWAP gate, and decompose the variational circuit into ISWAP$^{1/2}$ gates. (Though not terribly well-known, the ISWAP$^{1/2}$ gate is a good native hardware gate for superconducting qubits, and decompositions of other gates into ISWAP$^{1/2}$ gates are known [3].) Then, throughout

the circuit, we implement $ISWAP^{1/2+x_i}$ gates in place of $ISWAP^{1/2}$ gate. We additionally add amplitude and phase damping noise at a rate $\frac{p}{2}$. In Fig. 3.10, we plot the result following optimization via the COBYLA algorithm implemented in scipy [152], in the absence of sampling noise. We see that the verification circuit is insensitive to the incoherent noise as expected, and behaves similarly to the effect of amplitude and phase damping alone (Fig. 3(right) of the main text).

3.F Term-wise comparison of VPE performance

To attempt to further understand the ability of VPE to mitigate errors, in this appendix we consider the effect of estimating different types of terms on the same preparation circuit. We consider the fermionic swap network used in Sec.V.C of the main text to prepare states for a H_2 Hamiltonian. When this was split into number-conserving Pauli operator sums (Fig.6(2a-2b) of the main text), different circuits had to be used to estimate individual terms. In Fig. 3.11, we show the result of estimating the expectation values of two of the individual terms used in the control-free Pauli operator decomposition under an amplitude-damping noise model (Fig.6(2b) of the main text). (Recall that this figure demonstrated first-order sensitivity to this error model, whilst the low-rank factorization demonstrated a third-order sensitivity to the same model.) We see that the $H_s = Z_0 Z_1$ term (left plot) shows the cubic dependence on error rate observed in previous amplitude-damping experiments, whilst the two-body scattering term (right plot)

$$H_s = X_0 Y_1 Y_2 X_3 + Y_0 X_1 X_2 Y_3 - X_0 X_1 Y_2 Y_3 - Y_0 Y_1 X_2 X_3, \qquad (3.120)$$

does not. This two-body scattering term is the only term contributing to the first-order decay of the VPE estimation observed in Fig.6(b2) of the main text — all other terms in the decomposition display similar decay to Fig. 3.11(left). This indicates that the errors to which we are first-order sensitive occur during the circuit implementation of $e^{iH_s t}$, and not the state preparation. The circuit implementing $e^{iH_s t}$ for the two-body scattering term is the only such circuit that does not conserve number throughout. (Instead, this evolution is achieved in two steps: a basis transform of $XY, YX \rightarrow IZ, ZI$ on pairs of qubits, ZZ rotations between the pairs and uncomputing, and then a basis transform of $XX, YY \rightarrow IZ, ZI$ on pairs of qubits, ZZ rotations between the pairs, and uncomputing again.)

Finding decompositions of these circuits more amenable to VPE is a clear target for future work.
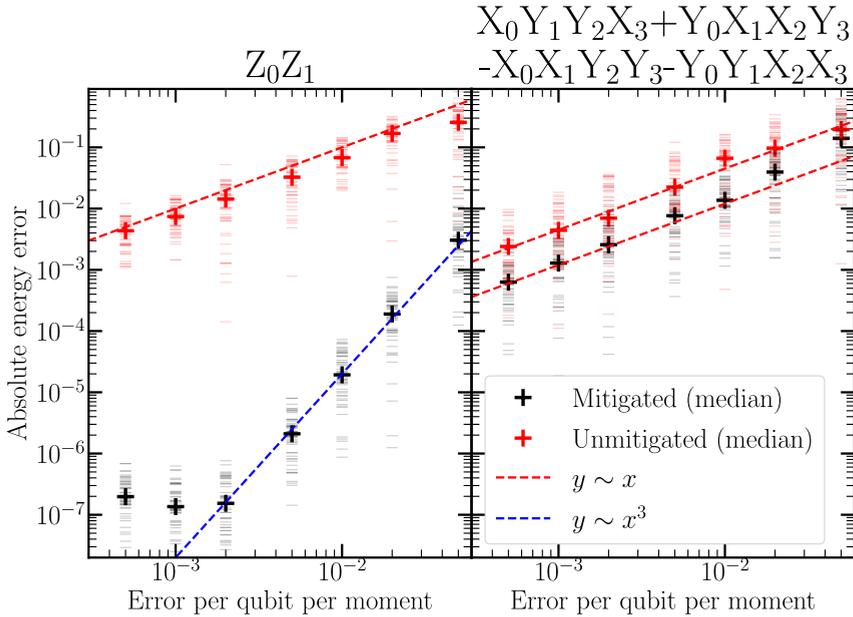
Figure 3.11: Expectation value estimation of two individual H_s terms from the control-free number-conserving Pauli operator decomposition of the H_2 Hamiltonian studied in Fig.6 of the main text on states prepared by a fermionic swap network. The two terms here comprise part of the sum (Eq.6 of the main text) for the expectation value of Fig.6(2b) of the main text — but are studied here without pre-factors (i.e. $\|H_s\| = 1$). Each figure is labeled with the studied term, and guide-lines (dashed red and blue) are given to show observed scaling laws. Data presented is the median (crosses) over 50 individual data points (faint dashes) of the absolute error in estimation using VPE (black) and standard partial state tomography (red).

3.G Comparison to symmetry verification

In this section we present a comparison of verified phase estimation and symmetry verification on a depolarizing noise model, using the experiment

in Fig.6(b1) of the main text. In order to improve performance, we choose to verify on the number operator $\sum_i Z_i$, instead of the parity $\prod_i Z_i$. To perform symmetry verification we take the quantum state prepared by the circuit, and directly project this into the number-conserving space. (In a real experiment simultaneous readout of the number operator and all terms is possible [103], but requires a slight addition of circuitry, which would increase the final error slightly.) In Fig. 3.12, we observe that while symmetry verification reduces error by around an order of magnitude, it does not provide the same asymptotic improvement as VPE. We also note that VPE improves over symmetry verification at all error rates, despite having circuit over 3 times as deep. This is to be expected; as phase (Z_i) errors commute with the number operator, these cannot be detected by symmetry verification and so contribute at first order to the final error rate.
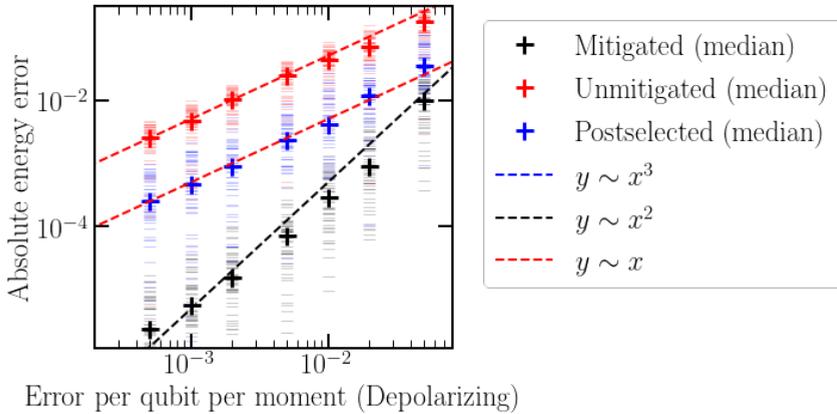
Figure 3.12: Comparison between verified phase estimation and symmetry verification. Both techniques are compared on the estimation of the expectation value of the electronic structure Hamiltonian for H_2 under a depolarizing noise model. Verified and unverified results are from the same simulated experiment as Fig.6(b1) of the main text. Although symmetry verification improves the energy error by around a factor 10, it still exhibits first-order scaling, as it cannot correct for phase (Z) errors during the experiment.

Optimizing the information extracted by a single qubit
measurement

4.1 Introduction

The largest bottleneck in quantum algorithm design is the encoding and
decoding of a quantum state. Although each full characterization of a
quantum state requires an exponentially large amount of information,
direct measurements of an N-qubit quantum state ρ extract only N bits
of information, and collapse ρ to a state described by those N bits alone
— erasing any other information. Performing this repeatedly allows the
estimation of an expectation value $\langle O \rangle := \text{Tr}[O\rho]$ of any operator O that is
diagonal in the measurement basis. The rate at which such a measurement
converges is known as the standard quantum or shot noise limit [153] -
after M repeated preparations, $\langle O \rangle$ can be estimated with variance

$$\text{Var}[O] = M^{-1} \left(\langle O^2 \rangle - \langle O \rangle^2 \right). \tag{4.1}$$

Though this rate can be improved upon [66, 129, 154, 155], doing so
requires implementing long coherent circuits or performing large correlated
measurements, which are not feasible in the current NISQ era [18].

Instead of using all N qubits to extract data from a quantum state,

one may perform a partial measurement that extracts less than N bits, and use the remaining qubits to detect and mitigate errors [29, 66, 122]. Error mitigation is key in obtaining precise results from NISQ circuits, such as variational algorithms [19, 82], where the output of the quantum algorithm is a set of estimates of expectation values. Echo verification (EV - see Section 4.2.2) [69, 156–158] allows one to strongly mitigate errors in a wide class of algorithms, by recasting measurements as Hadamard tests. In each EV circuit, a single bit of information is extracted from the system register as a measurement, freeing up the remainder of the register for error detection/mitigation. One may combine results of multiple EV circuits (through classical post-processing) into an error-mitigated estimator of any target quantity. However, the stringent requirement that only one bit of information be extracted from the device further tightens the bottleneck of quantum-classical I/O.

In this paper we study how we can optimize information extraction from a quantum system to estimate the expectation value of an observable O, under the restriction that only a single bit of information is measured per state preparation. This matches the requirements of EV, the rest of the information being reserved for error mitigation. We do not focus in this work on the effectiveness of EV as an error mitigation strategy, and consider only the case of error-free quantum simulation. We define measurements with a single-bit outcome in terms of the Hadamard test, use these to construct an expectation value estimator for a more complicated operator via a linear decomposition, and calculate the variance of this resulting estimator. We prove necessary conditions for such a linear decomposition to be optimal; i.e. to minimize the cost of expectation value estimation. We construct a provably optimal (in some sense) decomposition for a fast-forwardable operator, and give a general (albeit expensive) method to implement this decomposition through quantum signal processing [46–48]. We analyse our methods numerically, comparing the variance of estimators based on our optimal method with other known approaches such as Pauli decompositions and the Dirichelet kernel measurements introduced in [159]. We find an asymptotic improvement between our optimal decomposition and a simple Pauli decomposition of a factor $N^{0.7}$, which at 13 qubits gives already an order of magnitude improvement.

4.2 Single-qubit measurements

The most general measurement that extracts one bit of information from a N-qubit state $|\psi\rangle$ is a binary Positive-Operator Valued Measurement

(binary POVM); this is defined by two positive operators $\Pi_+, \Pi_- > 0$ such that $\Pi_+ + \Pi_- = \mathbb{1}$. The outputs of such measurement, which we label $+1$ and -1, have probabilities $p_\pm = \langle\psi| \Pi_\pm |\psi\rangle$. Schematically,
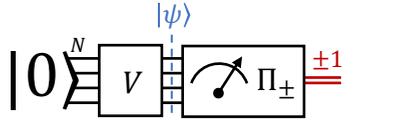
where we defined the unitary preparing the state $V|0\rangle := |\psi\rangle$. In section 4.2.1, we review the Hadamard test and we show that there exists a one-to-one equivalence between outcomes of Hadamard tests and binary POVMs.

Extracting only a single bit allows further processing of the quantum information remaining in the state register. For instance, inverting the unitary that prepared $|\psi\rangle$ and measuring in the computational basis yields a powerful error mitigation technique, echo verification [156–158], which we review in Section 4.2.2. In another example, the Hadamard test may be used to estimate the gradient of a cost function with respect to a variational term $\exp(iA\theta)$ in a circuit, as $\frac{d}{d\theta}\exp(iA\theta) = iA\exp(iA\theta)$ [160, 161]. Both these methods require operating on the system register after the binary measurement is performed, preventing further information extraction. (For the specific case of EV, we show in App. 4.B that extracting more than one bit of information is counterproductive.) Furthermore, this restricted output model of quantum computation can be relevant in quantum-enhanced metrology settings [154, 162], where a single-qubit probe is used [163]. A similar restricted access model has been studied in the context of Hamiltonian learning [164, 165]. Note that this single qubit access model is different to the one clean qubit model of computation (DQC-1) [166]; here we consider using a single qubit to extract information from a non-trivial quantum state.

4.2.1 The Hadamard test

A Hadamard test (HT) is a binary measurement performed on a state $|\psi\rangle_s$ in the N-qubit system register s. It is implemented through a control qubit c initialized in the state $|+\rangle_c$, a controlled unitary CU and a projective Pauli measurement X_c on the control qubit. As a quantum circuit this

can be written
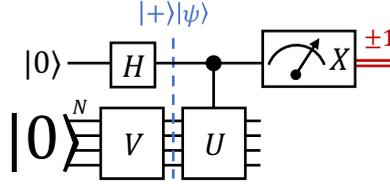
and the resulting state before measurement can be easily calculated to be

$$|\Phi\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle |\psi\rangle + |1\rangle U |\psi\rangle \Big). \tag{4.2}$$

Tracing out the system register then yields the following reduced density matrix on the control qubit,

$$\rho_{\mathrm{c}} = \frac{1}{2} \begin{pmatrix} 1 & \langle U \rangle \\ \langle U \rangle^* & 1 \end{pmatrix}. \tag{4.3}$$

One may estimate the expectation value of $\mathrm{Re}(U) := \frac{1}{2}(U + U^\dagger)$ by measuring the control qubit in the X basis, which returns $\mathrm{Tr}[X\rho_{\mathrm{c}}] = \langle \frac{1}{2}(U + U^\dagger) \rangle$.

To prove the equivalence between HT and binary POVM, we explicitly construct one from another. To construct the binary POVM corresponding to the HT, we define the measurement operators that represent the back-action of the measurement on the system register

$$M_\pm = \langle \pm |_{\mathrm{c}} CU |+\rangle_{\mathrm{c}} = \frac{\mathbb{1} \pm U}{2}, \tag{4.4}$$

and the relative positive operators $\Pi_\pm = M_\pm^\dagger M_\pm$ used to compute probabilities $p_\pm = \langle \psi | \Pi_\pm |\psi\rangle$ of measuring ± 1 on the ancilla. Vice versa, given a binary POVM $\{\Pi_+, \Pi_-\}$, we can construct a corresponding Hadamard test by choosing a unitary U that satisfies $\mathrm{Re}(U) = \Pi_+ - \Pi_-$,

$$U = \exp[i\arccos(\Pi_+ - \Pi_-)]. \tag{4.5}$$

This is always possible because $\Pi_+ - \Pi_-$ is Hermitian and $\|\Pi_+ - \Pi_-\| < \|\Pi_+ + \Pi_-\| = 1$. It is easy to check that the Hadamard test constructed from this unitary return the correct positive operators $\Pi_\pm$.

4.2.2 Echo verification

The name echo verification (EV) refers to a class of powerful error mitigation techniques [156–158], applicable in most algorithms that make use a Hadamard test to perform measurements on a system register. This technique was originally introduced by the name of *verified phase estimation* [156] as it considered estimating expectation values of multiple unitaries $U_l = e^{iHt_l}$, with an archetypal application in the context of single-ancilla phase estimation. However, in this work, we consider the more general expectation-value estimation subroutine yielding $\langle \mathrm{Re}(U) \rangle$. We prefer the name echo verification (used also in [69, 79]) due to the similarities to a Loschmidt echo.

Echo Verification relies on a key idea: exploiting the information left in the system register after the application of the controlled-unitary operator prescribed by the Hadamard test. This information is used to detect errors and mitigate their effect on estimated quantities. This is done by "echoing" the preparation unitary V, i.e. applying $V^\dagger$ after the controlled evolution, and verifying whether the register s returns to the initial state $|0\rangle$. The corresponding circuit is
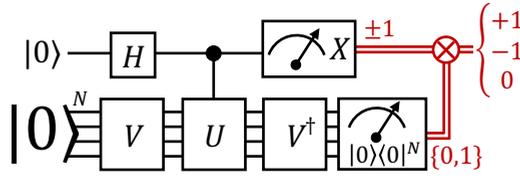
where the multiplication of the classical information channels (red double-lines) sets the circuit output to zero upon failed verification (i.e. if the final system state is orthogonal to $|0\rangle_s$), and to the output of the Hadamard test otherwise.

Let us denote the combined state after the controlled unitary as $|\Phi\rangle$, and let $\Pi_\psi = |\psi\rangle\langle\psi| = V^\dagger |0\rangle\langle 0| V$ be the projector on the state $|\psi\rangle_s$. The estimate of $\langle \mathrm{Re}(U) \rangle$ can be obtained by measuring the operators $X^{\mathrm{EV}} := X \otimes \Pi_\psi$ on $|\Phi\rangle$ (EV circuit), as opposed to $X_c := X \otimes \mathbb{1}$ (HT circuit). One can confirm that, in the absence of error, these operators have identical expectation values on the state at the end of the circuit [156]

$$\langle\Phi|X^{\mathrm{EV}}|\Phi\rangle = \langle\Phi|X_c|\Phi\rangle = \langle\psi|\mathrm{Re}(U)|\psi\rangle. \tag{4.6}$$

For an intuitive explanation, note that if the controlled unitary changes the state of the system register, the ancilla qubit must have been in the

$|1\rangle$ state, and $\langle 1|X|1\rangle = 0$. This implies that the expectation value of $X \otimes (\mathbb{1}_N - \Pi_\psi)$ is 0.

In the presence of a circuit error, verification is likely to fail. This decreases the expectation value measured by the error probability, which can be measured separately. Rescaling the result by the error probability yields a noise-mitigated estimate of the expectation value $\langle \psi| \operatorname{Re}(U)|\psi\rangle$. The error mitigation power of this method is explored in [79, 156–158, 167] and experimentally tested in [69]. In this work, we only consider noiseless circuits.

The EV circuit implements a ternary measurement, with outputs $+1, -1, 0$. Compared to a standard HT defined by the same unitary, the probabilities p_+ and p_- are reduced by the same amount ($\frac{p_0}{2}$), yielding a result with the same expected value. As a consequence, the variance of an EV measurement is always smaller than that of the corresponding HT (this is formalized in Appendix 4.A).

An extension of Echo Verification allows extracting more than one qubit of information per circuit run by using multiple auxiliary qubits. However, as the measurement is quadratic in $|\psi\rangle\langle\psi|$ (resulting by the use of two copies of V_ψ in the circuit), reconstructing the desired expectation values requires nonlinear processing of the measurement results. Furthermore, as each measurement interferes with the verification of the others, all the variances of estimated expectations increase. In appendix 4.B we explore this, and we prove that measuring more than one bit of information per EV experiment is always counterproductive in terms of final variance, for a fixed total number of shots.

4.2.3 Ancilla-free echo verification

The direct (control-based) measurement via the HT may often be replaced by an indirect measurement using an altered circuit [156, 168, 169], allowing control-free implementations of these single-bit measurements. We review briefly the control-free echo verification scheme.

In the Hadamard test, the control qubit provides a clock-reference state $|0\rangle\,|\psi\rangle$, which is not changed by the application of CU. This clock-reference state is necessary to give physical meaning to the phase U induces on the system register states, thus making it measurable. If U has a known eigenstate $U\,|\psi_r\rangle = e^{i\phi_r}\,|\psi_r\rangle$ orthogonal to $|\psi\rangle$, this state can be used as a clock-reference removing the need for a control qubit. In quantum simulation, this state can often be found thanks to the symmetries of the system. For example, in second-quantized simulation of particle systems the vacuum state $|0\rangle$ is an eigenstate of any particle-number preserving

operator.

The control-free EV scheme prescribes preparing a cat-state $\frac{1}{\sqrt{2}}(|\psi\rangle + |\psi_r\rangle)$, applying U, and measuring $X^{\text{CFEV}} = (|\psi\rangle\langle\psi_r| + |\psi_r\rangle\langle\psi|)$. This can be done with the circuit
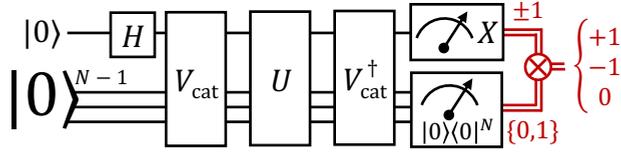
where $V_{\text{cat}}|0...00\rangle = |\psi_r\rangle$ and $V_{\text{cat}}|0...01\rangle = |\psi\rangle$. After the application of U, the state is $|\Phi\rangle = \frac{1}{\sqrt{2}}(U|\psi\rangle + e^{i\phi_r}|\psi_r\rangle)$, thus

$$\langle\Phi| X^{\text{CFEV}} |\Phi\rangle = \langle\psi| \text{Re}(Ue^{-i\phi_r}) |\psi\rangle. \tag{4.7}$$

If $\phi_r \neq 0$, the desired result $\langle\psi| \text{Re}(U) |\psi\rangle$ can be obtained by substituting $U \to Ue^{i\phi_r}$ or applying a phase gate $e^{-i\phi_r/2Z}$ to the first qubit before measurement.

4.2.4 Variance of a binary POVM

The Hadamard test differs from the projective measurement of $\text{Re}(U) := \frac{1}{2}(U + U^\dagger)$ (the Hermitian part of U). Each instance of the Hadamard test can only output $+1$ or -1, whereas the spectrum of $\text{Re}(U)$ can have up to 2^N distinct eigenvalues in the range $[-1, 1]$. This has a direct impact on the estimation uncertainty: performing the Hadamard test M times and measuring the control qubit in the X-basis yields an estimator of $\langle\text{Re}(U)\rangle = \text{Re}(\langle U\rangle)$ with a variance

$$\text{Var}^* \left[\langle\text{Re}(U)\rangle \right] = \frac{1 - \langle\text{Re}(U)\rangle^2}{M}, \tag{4.8}$$

which can be seen to be strictly larger than the variance one would obtain by performing a projective measurement of $\text{Re}(U)$ on M copies of $|\psi\rangle$ [Eq. (4.1)],

$$\text{Var} \left[\langle\text{Re}(U)\rangle \right] \leq \text{Var}^* \left[\langle\text{Re}(U)\rangle \right], \tag{4.9}$$

as $\langle\text{Re}(U)^2\rangle \leq 1$. Our goal is to optimize estimators of expectation values $\langle O\rangle$ of a given operator, which use data from multiple HTs with different unitaries U [each with the given variance Eq. (4.8)], and assuming one

test per state preparation. We want to minimize the total number of state preparations (distributed over different choices of U) needed to achieve an estimator of $\langle O \rangle$ with error smaller than a fixed ϵ.

4.3 Operator decompositions

It is common in quantum computing to estimate the expectation value of an operator O by writing O as a linear combination of simpler terms (a.k.a. a *decomposition*) which have their expectation values estimated independently [19, 54, 170]. In this work, we make use of this method, and consider estimating these simpler terms via Hadamard tests. Let us fix a decomposition[1] X,

$$O = \sum_{x \in X} c_x \operatorname{Re}(U_x) \leftrightarrow \langle O \rangle = \sum_{x \in X} c_x \langle \operatorname{Re}(U_x) \rangle, \qquad (4.10)$$

and consider estimating $\langle O \rangle$ by estimating each $\langle \operatorname{Re}(U_x) \rangle$ independently and summing the results. As $\operatorname{Re}(U_x)$ and O are Hermitian operators we may assume c_x to be real without loss of generality, and we may further assume $c_x \geq 0$ by absorbing a minus sign onto U_x. Note that the arrow in Eq. (4.10) points both ways as the set of expectation values on all states $|\psi\rangle$ uniquely defines an operator.

Once a suitable decomposition X of an operator O [Eq. (4.10)] has been chosen, to calculate the total cost of the algorithm we must allocate a number m_x of repeated single-shot HT experiments to estimate individual $\langle \operatorname{Re}(U_x) \rangle$. We assume a single-bit measurement per state preparation, i.e. each HT requires resetting the circuit and re-preparing $|\psi\rangle$, and the total number of re-preparations $M_X = \sum_{x \in X} m_x$ is the relevant cost of implementing our measurement scheme. If each $\langle \operatorname{Re}(U_x) \rangle$ is estimated independently, the variance on a final estimate of $\langle O \rangle$ can be calculated by standard propagation of variance

$$\operatorname{Var}_X^* \left[\langle O \rangle \right] = \sum_{x \in X} c_x^2 \operatorname{Var}^* \left[\langle \operatorname{Re}(U_x) \rangle \right] \qquad (4.11)$$

$$= \sum_{x \in X} \frac{c_x^2 (1 - \langle \operatorname{Re}(U) \rangle^2)}{m_x}. \qquad (4.12)$$

---

[1] In a slight abuse of notation, throughout this work we will use the same label (e.g. X) to represent the entire linear decomposition defined by the set $\{c_x, U_x\}$ in Eq. 4.10, and the set of labels x that we sum over.

Eq. (4.9) implies that under the same decomposition of O

$$\text{Var}_X\big[\langle O\rangle\big] := \sum_{x\in X} c_x^2\,\text{Var}[\langle\text{Re}(U_x)\rangle] \leq \text{Var}_X^*\big[\langle O\rangle\big], \qquad (4.13)$$

for all states ρ.

4.3.1 Adaptive shot allocation

Given a decomposition X and a total shot budget M_X, an optimal choice for the m_x may be found using Lagrange multiplier methods [121]

$$m_x = M_X\,\frac{c_x\sqrt{1-\langle\text{Re}(U_x)\rangle^2}}{\sum_{y\in X} c_y\sqrt{1-\langle\text{Re}(U_y)\rangle^2}}, \qquad (4.14)$$

recalling that $c_x \geq 0$. This yields a bound on the required M_X to estimate $\langle O\rangle$ with $\text{Var}^*\big[\langle O\rangle\big] = \epsilon^2$

$$M_X \geq \mathcal{M}_X := \epsilon^{-2}\bigg[\sum_{x\in X} c_x\sqrt{1-\langle\text{Re}(U_x)\rangle^2}\bigg]^2. \qquad (4.15)$$

We call $\mathcal{M}_X$ the *cost* of the decomposition X. This may be compared to well-known results for measurement bounds using standard tomography methods [54, 66, 68, 121] by substituting Var^* for Var in Eq. (4.11). Though exact values of $\langle U_x\rangle$ will not be known in advance, these can be estimated using a small initial fraction of measurements before a final distribution of measurements is allocated.

4.3.2 The decomposition hierarchy

We have shown above how to optimize measurement allocation given a linear decomposition X [Eq. (4.10)]. Let us now consider how to optimize X to minimize Eq. (4.15).

We first consider the effect of possible rescalings of $\text{Re}(U_x)$. If any term $c_x\,\text{Re}(U_x)$ has $\|\text{Re}(U_x)\| < 1$,[2] one can find some unitary $U_{x'}$ for which $\text{Re}(U_{x'}) = \text{Re}(U_x)/\|\text{Re}(U_x)\|$; substituting $U_x \to U_{x'}$ (and $c_x \to c_{x'}$ accordingly) will always improve the bound in Eq. (4.15). (For now we do not worry about how the unitaries may be implemented as quantum circuits; we will consider this issue later.)

---

[2] Unless stated otherwise, all norms in this work are the spectral norm.

One may next consider subdividing individual terms $\mathrm{Re}(U_x)$ of X, by writing

$$c_x \, \mathrm{Re}(U_x) = c_{x,0} \, \mathrm{Re}(U_{x,0}) + c_{x,1} \, \mathrm{Re}(U_{x,1}), \qquad (4.16)$$

where $U_{x,0}$ and $U_{x,1}$ are both unitary, and $c_x, c_{x,0}, c_{x,1} > 0$. As we can assume $\| \, \mathrm{Re}(U_x) \| = 1$, such a decomposition requires $c_{x,0} + c_{x,1} \geq c_x$, to preserve the spectral norm of $\mathrm{Re}(U_{x,0})$ and $\mathrm{Re}(U_{x,1})$. When this inequality is saturated, we call the sub-decomposition *norm-preserving*. It turns out that this condition is sufficient for the sub-decomposition to be non-increasing in the cost $\mathcal{M}$ of estimation [Eq. (4.15)], for all states $|\Psi\rangle$; formally:

Lemma 1
Given a linear decomposition X of a target operator O [Eq. (4.10)], a sub-decomposition X' [Eq. (4.16)] that is norm-preserving has non-increasing cost, $\mathcal{M}_{X'} \leq \mathcal{M}_X$ [Eq. (4.15)], for any state $|\Psi\rangle$.

We give a proof of this lemma in Appendix 4.C.1
We would like to extend the above lemma to a statement that norm-increasing subdecompositions of a linear decomposition X are always suboptimal in some sense. To achieve this, note that as a corollary to lemma 1, we can improve on all terms $c_x \, \mathrm{Re}(U_x)$ in a linear decomposition X by a norm-preserving identity shift

$$c_x \, \mathrm{Re}(U_x) = c_x (1 - \bar{\lambda}_x) \, \mathrm{Re}(U_{\tilde{x}}) + c_x \bar{\lambda}_x \mathbb{1}, \qquad (4.17)$$

where $\bar{\lambda}_x = \frac{1}{2}(\lambda_x^{\min} + \lambda_x^{\max})$, $\lambda_x^{\min}$ and $\lambda_x^{\max}$ are the lowest and highest eigenvalues of $\mathrm{Re}(U_x)$ respectively, and $\mathrm{Re}(U_{\tilde{x}})$ has the same eigenvectors of $\mathrm{Re}(U)$ (with its spectrum shifted and rescaled). We call the outcome decomposition $\tilde{X}$ of the procedure above the *center* of X. Though a norm-increasing subdecomposition of X may not be suboptimal relative to X, it is suboptimal relative to this center:

Lemma 2
Let X be a linear decomposition of O with all $\|\mathrm{Re}(U_x)\| = 1$; let $\tilde{X}$ be the center of X and let X' be a strictly norm-increasing sub-decomposition. There exists at least one state $|\Psi\rangle$ for which the cost $\mathcal{M}_{\tilde{X}} < \mathcal{M}_{X'}$.

We give a proof of this lemma in Appendix 4.C.2.
To recap, the above two lemmas show a) that norm-preserving sub-decompositions do not increase the cost of estimating expectation values via Hadamard tests on any given state, and b) norm-increasing sub-decompositions not only can increase expectation value estimation costs

on some states, but are guaranteed to do so on at least one. This result is in direct contrast to standard expectation value estimation, where independent estimation of $\langle A \rangle$ and $\langle B \rangle$ is sub-optimal to joint estimation of $\langle A + B \rangle$ whenever the latter is possible. This suggests a path towards optimizing HT expectation value estimation, by repeatedly dividing terms $\mathrm{Re}(U_x)$ in a norm-preserving manner, until no further sub-decomposition can reduce the cost any state. It turns out that not all choices of division lead to the same end-point, however all end points of this procedure have one common property (proven in Appendix 4.C.3):

Lemma 3
A decomposition X of an operator O has no non-trivial norm-preserving sub-decompositions if and only if all operators $\mathrm{Re}(U_x)$ in X are reflections: $\mathrm{Re}(U_x)^2 = 1$.

It should be no surprise that we find reflection operators $\mathrm{Re}(U_x)^2 = 1$ to be a crucial ingredient to optimize HT tomography, as these are the only operators that saturate the bound in Eq. (4.9) for all states $|\Psi\rangle$. We call a decomposition X that consists of reflection operators only a *reflection decomposition*. We give some simple examples of these in Appendix 4.C.4.

4.3.3 Optimizing reflection decompositions

Above we demonstrated that, for a decomposition X of an operator O to be optimal with regards to the cost $\mathcal{M}_X$ of estimating expectation values on a set of states (Eq. 4.15), all terms in X must be reflection operators. Otherwise, we demonstrated a means of sub-dividing single terms in the distribution to generate a new distribution with lower cost. However, this is not to say that all reflection decompositions X have the same cost $\mathcal{M}_X$. (These two statements are consistent as we cannot transform between reflection decompositions using subdivision.) The set of reflection decompositions of O form a convex set that is $2^{2^N - N}$-dimensional if all U_x are diagonal in the eigenbasis of O. This raises two questions: is there an optimal decomposition amongst the set of reflection decompositions, and does it achieve the von Neumann bound [Eq. (4.13)]?

Lemma 4
Let O be an operator and Π_j be projectors onto the eigenvalues of O;

$O\Pi_j = \Pi_j O = \lambda_j \Pi_j$. *The Ξ-decomposition of O, given by*

$$O = \frac{\lambda_0 + \lambda_J}{2} \, \mathbb{1} + \sum_{x=1}^{J-1} \frac{\delta\lambda_x}{2} \, \Xi_x \tag{4.18}$$

$$\Xi_x = \mathbb{1} - \sum_{j<x} 2\Pi_j \,, \quad \delta\lambda_x = \lambda_x - \lambda_{x-1}, \tag{4.19}$$

uniquely achieves the bound $\mathrm{Var}_\Xi^*[O] = \mathrm{Var}[O]$ *on all states* $|\Psi\rangle$ *with support on up to two eigenstates of O. No such decomposition achieves this bound on all states* $|\Psi\rangle$ *with support on three or more eigenstates of O.*

We prove this lemma in Appendix 4.C.5. Note that the Ξ-decomposition can be immediately restricted to any subspace of the full-2^N-dimensional Hilbert space containing $|\Psi\rangle$ (i.e. if we knew that due to a symmetry or by virtue of being a low-energy state, $|\Psi\rangle$ had support only on such a space), and the optimality result still holds. This implies in turn that no linear decomposition X can achieve the von Neumann variance bound even for as small as a 3-dimensional subspace. This makes sense, as our restriction to measure one bit of information per state preparation forms a bottleneck with respect to the 3 nonzero-probability outcomes of a Von Neumann measurement on this space.

4.3.4 Implementing the optimal decomposition

In order to realize the Ξ-decomposition estimator, we need to implement HT circuits that (approximately) estimate $\langle \Xi_x \rangle$. This may be achieved by realising that

$$\Xi_x = \mathrm{sgn}[O - \mu_x], \quad \mu_x = \frac{\lambda_{x-1} + \lambda_x}{2}, \tag{4.20}$$

where sgn is the sign function. An approximation of this unitary operator can then be realized using quantum signal processing (QSP) [46–48] of the sign function [86], requiring only one additional ancillary qubit. The QSP

circuit is given by
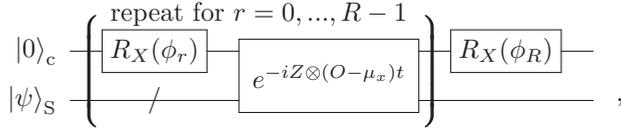
where $R_X(\phi_r) = e^{-i\frac{X}{2}\phi_r}$, implements a unitary block encoding $Q_{\boldsymbol{\phi}}$ of a degree-R trigonometric polynomial $S_{\boldsymbol{\phi}}$ of the operator $(O - \mu_x)t$:

$$\langle 1|_{\mathrm{c}}\, Q_{\boldsymbol{\phi}}\, |0\rangle_{\mathrm{c}} = \sum_{r=0}^{R} c_r(\boldsymbol{\phi}) e^{-ir(O-\mu_x)t} := S_{\boldsymbol{\phi}}[(O - \mu_x)t]. \qquad (4.21)$$

Here, $\boldsymbol{\phi}$ is a vector containing the individual angles ϕ_r implemented during the QSP circuit. We can then sample $\langle \mathrm{Re}\{S_{\boldsymbol{\phi}}[(O - \mu_x)t]\}\rangle$ through HT (or EV), using another qubit controlling all gates in the QSP circuit. To approximate Eq. (4.20) with our block-encoded operator $S_{\boldsymbol{\phi}}$, we must choose $t < \frac{\pi}{\|O-\mu_x\|}$ to avoid aliasing, and find the optimal $\boldsymbol{\phi}$

$$\boldsymbol{\phi} = \underset{(\phi_r = -\phi_{R-r})}{\arg\min} \int_{0+\delta}^{\pi-\delta} d\omega \Big[\mathrm{sgn}(\omega) - \mathrm{Im}[S_{\phi}(\omega)] \Big]. \qquad (4.22)$$

Here, the constraint $\phi_r = -\phi_{R-r}$ ensures $\mathrm{Im}[S_{\phi}(\omega)]$ is an odd function of ω. A resolution parameter $\delta \geq 0$ can be introduced to improve the approximation away from the nodes $\omega = \{0, \pm\pi\}$ of $S_{\phi}(\omega)$. In Appendix 4.D we give further details of this decomposition, and analyse the approximation error numerically. We find that this error converges exponentially in the number of circuit blocks R.

4.4 Numerical experiments

To investigate performance of various decompositions on states that have support on more than two eigenstates of O, and therefore are not covered by Lemma 4, we perform numerical simulations using random variationally-generated states and a simple toy operator $O = \sum_j Z_j$. (In appendix 4.F, we report this scaling for other systems.) We measure the variances on states generated by a hardware-efficient ansatz [59] with random input parameters using PennyLane [171]. For each datapoint 100 random states are generated. We consider estimating $\langle O \rangle$ in a realistic scenario where
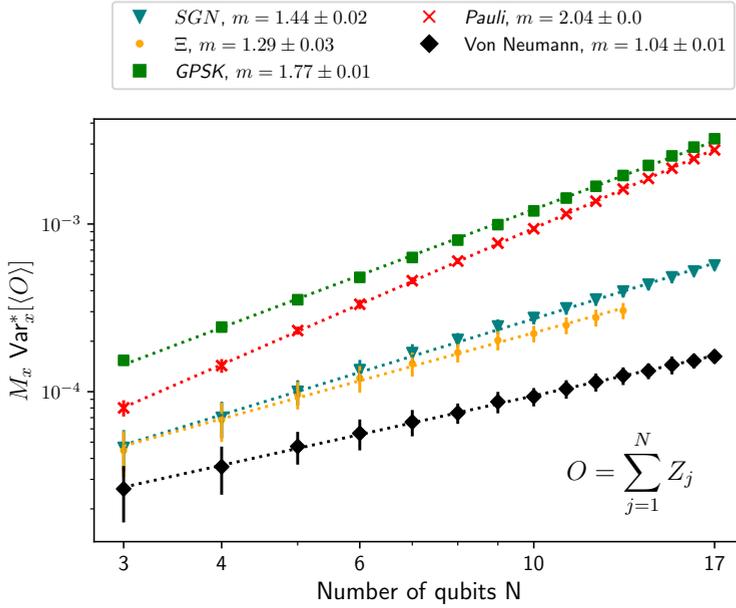
Figure 4.1: Comparison study of variances of different decompositions on random states generated by a hardware-efficient ansatz (see text for details). Different colours correspond to different decompositions [Eq.(4.10)] of the target operator O (see text for the description of all decompositions). Dashed lines are power-law fits to the data (obtained exponents are given in legend).

the $\langle \mathrm{Re}(U_x) \rangle$ values will not be known in advance to optimally choose m_x via Eq. (4.14). Instead, for each random state we generate a prior estimate of each $\langle \mathrm{Re}(U_x) \rangle$ from 10^5 measurements of the state, and use these to determine m_x (which are then only approximately optimal). This leaves the total shot count M_X as a free parameter; we resolve this in Fig. 4.1 by calculating $M_X \mathrm{Var}_X^*[\langle O \rangle]$. (This gives a quantity that is relevant regardless of the number of the shots actually used to estimate $\langle O \rangle$.)

An average of $M_X \mathrm{Var}_X^*[\langle O \rangle]$ over the 100 states is formed and plotted in Fig. 4.1 for each grouping method. This is compared to the Von Neumann measurement variance $\mathrm{Var}[O]$, which does not require any shot allocation, and sets a lower limit to the other estimators [see App. 4.C.5, Eq. (4.52)].

The Ξ-decomposition [orange, 'Ξ'] has the best asymptotic scaling of all decompositions, being suboptimal to Var[O] by a factor $\approx N^{1/3}$. The QSP approximation of Ξ, [teal, 'SGN'], has a slightly worse asymptotic scaling, which we associate to the error in approximating sgn($O - \mu_j$). At the largest considered $N = 13$, these two decompositions suffer approximately a factor 2 penalty in their total cost compared to Var[$\langle O \rangle$]. The generalized parameter-shift kernel decomposition [159] [green, 'GPSK', described in Appendix 4.E] has the worst overall performance out of the investigated estimators, due to the constant factor. It has however a better asymptotic scaling than a simple Pauli decomposition $U_x = Z_j$ [red, 'Pauli', Appendix 4.C.4]. In Appendix 4.F we investigate the scaling of different sets of observables. We observe that the order of the performance of the different decompositions remains consistent throughout, but the relative gains and losses in performance can be significantly different.

4.5 Conclusion

In this work we studied the optimization of expectation value estimation for a quantum state in the case where we are only allowed to measure a single qubit per state preparation (e.g. through Hadamard tests, with relevant application to echo verification). We calculated the cost of estimating the expectation value of an operator O by linearly decomposing O into a linear combination of sub-unitary terms, assuming an optimal shot allocation. We demonstrated that this cost is strictly non-increasing when terms are further subdivided, under the constraint that this subdivision preserves the induced 1-norm of the term coefficients. We showed that the end-points of this procedure of repeated division are linear decompositions of O where all terms are reflection operators; a so-called 'reflection decomposition'. We identified one such decomposition, the Ξ-decomposition, as unique in its ability to estimate $\langle O \rangle$ with a variance matching the Von Neumann measurement limit on any linear combination of up to 2 eigenstates of O. We demonstrated how the Ξ-decomposition may be approximately implemented through quantum signal processing. Numerical results demonstrate that on simple systems, the Ξ-decomposition and its approximate couterpart demonstrate clear constant and asymptotic improvements over other reflection decompositions (in the cost of estimating $\langle O \rangle$ on random states), with up to a factor $10\times$ improvement for estimation on 20 qubits.

Though these results are encouraging, the significant discrepancy between $\text{Var}^*_\Xi[O]$ and $\text{Var}[O]$ is worrying for NISQ algorithms that already incur a significant cost to tomograph complex Hamiltonians [54, 66, 122, 172–

174]; either one incurs a large overhead for measurement due to the need to invoke quantum signal processing or incur the clear asymptotic scaling cost that comes with measuring single Pauli terms per state preparation. Given that echo verification has a sampling cost scaling as $1/F^2$ (for a circuit fidelity F) [156], this result adds to the unlikelihood of beyond-classical NISQ variational algorithms in chemistry. Finding reflection decompositions with lower circuit depth is a clear avenue for future work.

4.A Echo verification estimators

The estimator used for echo verification is not identical to the one studied in the main text, and so its variance is not quite identical. In particular, we have $(X^{\mathrm{EV}})^2 = \Pi_\psi$ which implies that the variance on an estimate of $\langle\Phi|X_{\mathrm{EV}}|\Phi\rangle$ is

$$\mathrm{Var}^*_{\mathrm{EV}}[\langle\mathrm{Re}(U)\rangle] = \frac{\langle\Phi|\,\mathbb{1}\otimes\Pi_\psi\,|\Phi\rangle - \langle\psi|\,\mathrm{Re}(U)\,|\psi\rangle^2}{M}. \qquad (4.23)$$

Clearly $|\langle\Phi|\psi\rangle|^2 \leq 1$, which implies $\mathrm{Var}^*_{\mathrm{EV}}[\mathrm{Re}(U)] \leq \mathrm{Var}^*[\mathrm{Re}(U)]$ [by comparison with Eq (4.8)]. In other words, the varianc of the EV estimator is always smaller or equal to the variance of the relative HT estimator. It is easy to calculate from the circuit above that

$$\langle\Phi|\,\Pi_\psi\otimes\mathbb{1}\,|\Phi\rangle = \frac{1}{2}|1 + \langle\psi|\,U\,|\psi\rangle|^2, \qquad (4.24)$$

(noting that $\langle\mathrm{Re}(U)\rangle = \mathrm{Re}(\langle U\rangle)$, which can be subtituted back into our variance estimate to obtain

$$\mathrm{Var}^*_{\mathrm{EV}}[\mathrm{Re}(U)] \geq \frac{1 - \langle\psi|\mathrm{Re}(U)|\psi\rangle^2}{2M}, \qquad (4.25)$$

Thus, we have

$$\mathrm{Var}^*[\langle\mathrm{Re}(U)\rangle] \geq \mathrm{Var}^*_{\mathrm{EV}}[\langle\mathrm{Re}(U)\rangle] \geq \frac{\mathrm{Var}^*[\langle\mathrm{Re}(U)\rangle]}{2}. \qquad (4.26)$$

This justifies our focus in the main text on optimizing the estimator from a standard Hadamard test; this estimator is simpler to analyse, more general, and differs from the EV estimator (that motivated this work) by at most a factor 2.

4.B Parallelizing echo verification

In absence of echo verification, we can trivially parallelize Hadamard tests measuring K commuting operators $\{\mathrm{Re}(U_0), ..., \mathrm{Re}(U_{K-1})\}$ using K ancillary qubits, one controlling each U_k. If each U_k is controlled by a separate ancillary qubit (labeled k, where $C_k U_k$ represents the k-th unitary controlled by the k-th control qubit), the combined state of the system register s and ancillary qubits after all the unitaries are applied will be

$$\bigotimes_k C_k U_k \left|+\right\rangle_k \left|\psi\right\rangle . \tag{4.27}$$

The probabilities of obtaining ± 1 when measuring X on the j-th control qubit are

$$p_{j\pm} = \left\| \bigotimes_{k \neq j} C_k U_k \left|+\right\rangle_k \frac{1 \pm U_j}{2} \left|\psi\right\rangle \right\|^2 \tag{4.28}$$

$$= \frac{1}{4} \left\langle\psi\right| (1 \pm U_j^\dagger)(1 \pm U_j) \left|\psi\right\rangle \tag{4.29}$$

which coincides with the probabilities of a single Hadamard test with unitary U_j.

When performing echo verification, parallelization is more complicated. The result of verification (the measurement of $\Pi_\psi = |\psi\rangle\langle\psi|$ on the system register) is affected by all the controlled-U_k, and thus its result cannot be simply associated to one specific ancilla being in the state $|1\rangle$. To mitigate errors, all the cases in which the register is found in a state orthogonal to $|\psi\rangle$ should be considered as null towards all of the ancilla measurement results. The echo-verified probability of measuring the binary string $\vec{\sigma} = (\sigma_0, ..., \sigma_k)$, where each σ_k is ± 1 corresponding to the state $|\pm\rangle$ measured on the k-th ancilla, is then

$$p_{\vec{\sigma}}^{\mathrm{EV}} = \left| \left\langle\psi\right| \prod_k \left\langle\sigma_k\right|_k C_k U_k \left|+\right\rangle_k \left|\psi\right\rangle \right|^2$$

$$= \frac{1}{4^K} \left| \left\langle\psi\right| \prod_k (1 + \sigma_k U_k) \left|\psi\right\rangle \right|^2 . \tag{4.30}$$

The product in this equation can be then developed into a linear combination of 2^K expectation values (note that, as all U_k commute, the order

does not matter). Under the assumption that all these expectation values are real [granted if $U_k = \mathrm{Re}(U_k)$] Eq. (4.30) defines a quadratic system of 2^K equations with $2^K - 1$ unknowns[3]. Solving such system we find that the expectation value of a single $\mathrm{Re}(U_j)$ can be estimated by processing the sampled $p_{\vec{\sigma}}^{\mathrm{EV}}$ as

$$\langle \mathrm{Re}(U_j) \rangle = \underbrace{\left(\sum_{\vec{\sigma}:\sigma_j=+1} \sqrt{p_{\vec{\sigma}}^{\mathrm{EV}}} \right)^2}_{p_{j+}^{\mathrm{EV}}} - \underbrace{\left(\sum_{\vec{\sigma}:\sigma_j=-1} \sqrt{p_{\vec{\sigma}}^{\mathrm{EV}}} \right)^2}_{p_{j-}^{\mathrm{EV}}}, \qquad (4.31)$$

where we denoted $p_{j\pm}^{\mathrm{EV}}$ the terms that reproduce the probabilities that would be returned by a single, un-parallelized EV experiment

$$p_{j\pm}^{\mathrm{EV}} = \frac{1}{4} |\langle \psi | 1 \pm U_j | \psi \rangle|^2. \qquad (4.32)$$

We assume $p_{\vec{\sigma}}^{\mathrm{EV}}$ are sampled by averaging M shots of the parallel EV experiment. These are probabilities of mutually-exclusive measurements, thus the covariance matrix of the $p_{\vec{\sigma}}^{\mathrm{EV}}$ estimators is defined by

$$\mathrm{Var}[p_{\vec{\sigma}}^{\mathrm{EV}}] = \frac{1}{M} p_{\vec{\sigma}}^{\mathrm{EV}} (1 - p_{\vec{\sigma}}^{\mathrm{EV}}), \qquad (4.33)$$

$$\mathrm{Cov}[p_{\vec{\sigma}}^{\mathrm{EV}}, p_{\vec{\rho}}^{\mathrm{EV}}] = -\frac{1}{M} p_{\vec{\sigma}}^{\mathrm{EV}} p_{\vec{\rho}}^{\mathrm{EV}} \quad \text{if} \quad \vec{\sigma} \neq \vec{\rho}. \qquad (4.34)$$

We can then propagate the error through Eq. (4.31) to obtain the variance on the parallel-EV (PEV) estimator of $\langle \mathrm{Re}(U_j) \rangle$

$$M \, \mathrm{Var}_{\mathrm{PEV}}^*[\langle \mathrm{Re}(U_j) \rangle] =$$

$$\sum_{\vec{\sigma}} \frac{p_{j\sigma_j}^{\mathrm{EV}}}{p_{\vec{\sigma}}^{\mathrm{EV}}} p_{\vec{\sigma}}^{\mathrm{EV}} (1 - p_{\vec{\sigma}}^{\mathrm{EV}}) - \sum_{\vec{\sigma} \neq \vec{\rho}} \sigma_j \rho_j \frac{\sqrt{p_{j\sigma_j}^{\mathrm{EV}}}}{\sqrt{p_{\vec{\sigma}}^{\mathrm{EV}}}} \frac{\sqrt{p_{j\rho_j}^{\mathrm{EV}}}}{\sqrt{p_{\vec{\rho}}^{\mathrm{EV}}}} p_{\vec{\sigma}}^{\mathrm{EV}} p_{\vec{\rho}}^{\mathrm{EV}}$$

$$= \sum_{\vec{\sigma}} p_{j\sigma_j}^{\mathrm{EV}} - \langle \mathrm{Re}(U_j) \rangle$$

$$= 2^{K-1} (p_{j\sigma_+}^{\mathrm{EV}} + p_{j-}^{\mathrm{EV}}) - \langle \mathrm{Re}(U_j) \rangle. \qquad (4.35)$$

---

[3] In the case of a more general $U = \mathrm{Re}(U) + i \, \mathrm{Im}(U)$, a similar system can be constructed by measuring each U_k and iU_k with $2K$ ancillas. Showing this is besides the scope of our work, and for the sake of simplicity we restrict ourselves to the case of Hermitian $U_k = \mathrm{Re}(U_k)$.

which explodes exponentially with the size of the parallelization K.

More generally, we can compute the covariance matrix for all the $p_{j,\sigma_j}^{\mathrm{EV}}$ through error propagation

$$\mathrm{Var}[p_{j\sigma_j}^{\mathrm{EV}}] = p_{j\sigma_j}^{\mathrm{EV}}(2^{K-1} - p_{j\sigma_j}^{\mathrm{EV}}) \tag{4.36}$$

$$\mathrm{Cov}[p_{j\sigma_j}^{\mathrm{EV}}, p_{k\rho_k}^{\mathrm{EV}}] = \delta_{j,k}\sqrt{p_{j\sigma_j}^{\mathrm{EV}} p_{k\rho_k}^{\mathrm{EV}}} - p_{j\sigma_j}^{\mathrm{EV}} p_{k\rho_k}^{\mathrm{EV}} \tag{4.37}$$

[where the covariance assumes $(j, \sigma_j) \neq (k, \rho_k)$]. This shows that, increasing K, we effectively add to the covariance matrix a positive semi-definite term with a norm that scales exponentially in K. As all the decompositions Eq. (4.10) are ultimately to be estimated as linear combinations of the sampled probabilities $p_{j\sigma_j}^{\mathrm{EV}}$, parallelizing error verification is counterproductive.

4.C Proof of decomposition optimality hierarchy

In this section we build up to the proof that the Ξ-decomposition is optimal in terms of cost (4.15), by proving the lemmas introduced in the main text. We first prove that a norm-preserving sub-decomposition has non-increasing cost with respect to its parent decomposition, for all states $|\psi\rangle$. We then prove that a sub-decomposition that does not have the norm-preserving property is always sub-optimal (i.e. it has strictly greater cost than an alternative norm-preserving sub-decomposition). The iteration of the norm-preserving sub-decomposition procedure leads to one of many alternative improving sequences of decompositions. The endpoint of each sequence is a norm-preserving linear decomposition of O for which all unitaries are reflection operators. Finally, we prove that one of such decompositions (the Ξ-decomposition) achieves the Von-Neumann measurement variance bound on a certain set of states, and that no unbiased estimator based on single-qubit measurements can achieve this bound on a larger set of states.

4.C.1 Proof of Lemma 1, and corollaries

Given a linear decomposition X of an operator O [Eq. (4.10)], consider a norm-preserving sub-decomposition X' where a single term $x \in X$ is split according to Eq (4.16). The bound on the total number of shots Eq. (4.15)

will then change:

$$\mathcal{M}_X \to \mathcal{M}_{X'} = \epsilon^{-2} \left[\sum_{y \neq x} c_y \sqrt{1 - \langle \mathrm{Re}(U_y) \rangle^2} \right. \tag{4.38}$$

$$\left. + c_{x,0} \sqrt{1 - \langle \mathrm{Re}(U_{x,0}) \rangle^2} + c_{x,1} \sqrt{1 - \langle \mathrm{Re}(U_{x,1}) \rangle^2} \right]^2.$$

[with the change with respect to Eq.(4.15) being the second row]. This results in a reduction of the cost, as can be seen by calculating

$$c_x^2 \left[1 - \langle \mathrm{Re}(U_x) \rangle^2 \right]$$
$$= (c_{x,0} + c_{x,1})^2 - \left(c_{x,0} \langle \mathrm{Re}(U_{x,0}) \rangle + c_{x,1} \langle \mathrm{Re}(U_{x,1}) \rangle \right)^2$$
$$= c_{x,0}^2 \left[1 - \langle \mathrm{Re}(U_{x,0}) \rangle^2 \right] + c_{x,1}^2 \left[1 - \langle \mathrm{Re}(U_{x,1}) \rangle^2 \right]$$
$$\quad + 2 c_{x,0} c_{x,1} \left(1 - \langle \mathrm{Re}(U_{x,0}) \rangle^2 \langle \mathrm{Re}(U_{x,1}) \rangle^2 \right)$$
$$\geq c_{x,0}^2 \left[1 - \langle \mathrm{Re}(U_{x,0}) \rangle^2 \right] + c_{x,1}^2 \left[1 - \langle \mathrm{Re}(U_{x,1}) \rangle^2 \right]$$
$$\quad + 2 c_{x,0} c_{x,1} \sqrt{ \left[1 - \langle \mathrm{Re}(U_{x,0}) \rangle^2 \right] \left[1 - \langle \mathrm{Re}(U_{x,1}) \rangle^2 \right] }$$
$$= \left[c_{x,0} \sqrt{1 - \langle \mathrm{Re}(U_{x,0}) \rangle^2} + c_{x,1} \sqrt{1 - \langle \mathrm{Re}(U_{x,1}) \rangle^2} \right]^2, \tag{4.39}$$

where, in the center inequality we have used the fact that for $0 \leq a, b \leq 1$,

$$1 - ab \geq \sqrt{(1 - a^2)(1 - b^2)}. \tag{4.40}$$

As a corollary and example, we look at identity shifts of a term $x \in X$. For $\mathrm{Re}(U_x)$ with unit norm, we can assume without loss of generality the largest eigenvalue is $\lambda_{\max} = 1$, and the smallest is $\lambda_{\min}$. We can then perform the simple norm-preserving decomposition

$$c_x \mathrm{Re}(U_x) = c_x (1 - \bar{\lambda}) \mathrm{Re}(U_{x'}) + c_x \bar{\lambda} \mathbb{1} \tag{4.41}$$

with $\bar{\lambda} = \frac{1}{2}(\lambda_{\min} + \lambda_{\max})$. The resulting $\mathrm{Re}(U_{x'})$ has maximum eigenvalue $+1$ and minimum eigenvalue -1, thus it does not admit non-trivial identity shift.

A norm-preserving sub-decomposition Eq. (4.16) of a term with $|\mathrm{Re}(U_x)| = 1$ will only admit terms with $|\mathrm{Re}(U_{x,i})| = 1$. (This can be checked by taking the expectation value of both sides of Eq. (4.16) on the eigenstate on which $|\langle \mathrm{Re}(U_x) \rangle| = 1$.) By the same reasoning, terms with $\mathrm{Re}(U_x)$ having maximum eigenvalue $+1$ and minimum eigenvalues -1 [like those

obtained by the identity shifts Eq. (4.41)] only admit sub-decompositions whose terms have the same property.

4.C.2 Proof of Lemma 2

In this appendix we compare the costs of two decompositions derived by an original decomposition X: the center $\tilde{X}$ where all terms are transformed according to Eq. (4.17), and the norm-increasing subdecomposition X' where a term $x \in X$ is changed according to Eq. (4.16) assuming $c_{x,0} + c_{x,1} > c_x$. Remembering that all coefficients are positive $c_y > 0$, the cost of each decomposition Eq. (4.15) is the square of a sum of positive values; the terms in this sum for $y \neq x$ do not change for $X \to X'$, and have a non-increasing value for $X \to \tilde{X}$. We thus focus only on the term $x \in X$ and the derived ones, highlighted here

$$\mathcal{M}_{X'} = \epsilon^{-2} \left[\overbrace{\sum_{j \in \{0,1\}} c_{x,j} \sqrt{1 - \langle \mathrm{Re}(U_{x,j}) \rangle}}^{m'} + ... \right]^2, \qquad (4.42)$$

$$\mathcal{M}_{\tilde{X}} = \epsilon^{-2} \left[\underbrace{c_x(1 - \bar{\lambda}_x) \sqrt{1 - \langle \mathrm{Re}(U_{\tilde{x}}) \rangle^2}}_{\tilde{m}} + ... \right]^2. \qquad (4.43)$$

We now prove there exists a state $|\Psi\rangle$ for which $\tilde{m} < m'$, which implies $\mathcal{M}_{\tilde{X}} < \mathcal{M}_{X'}$.

Let $|\psi_+\rangle$ and $|\psi_-\rangle$ be eigenvectors of $\mathrm{Re}(U_{\tilde{x}})$ with eigenvalue $+1$ and -1 respectively. We consider three cases:

1. $|\langle \psi_\sigma | \mathrm{Re}(U_{x,j}) | \psi_\sigma \rangle| < 1$ for at least one combination of $\sigma \in \{+, -\}$ and $j \in \{0, 1\}$. In this case, on the state $|\Psi\rangle = |\psi_\sigma\rangle$ we get $\tilde{m} = 0 < m' \neq 0$.

2. $\langle \psi_\sigma | \mathrm{Re}(U_{x,j}) | \psi_\sigma \rangle = \sigma$ for all combinations of $\sigma \in \{+, -\}$ and $j \in \{0, 1\}$. By combining Eq. (4.16) and Eq. (4.17) and taking the expectation value on $|\psi_\sigma\rangle$ we obtain $\sigma[c_{x,0} + c_{x,1} - c_x(1 - \bar{\lambda}_x)] = c_x \bar{\lambda}_x$, which implies $c_{x,0} + c_{x,1} = c_x$, violating one of the hypotheses of the lemma.

3. $\langle \psi_\sigma | \mathrm{Re}(U_{x,j}) | \psi_\sigma \rangle = (-1)^j \sigma$ for all combinations of $\sigma \in \{+, -\}$ and $j \in \{0, 1\}$. We define the state $|\Psi\rangle = \frac{|\psi_+\rangle + |\psi_-\rangle}{\sqrt{2}}$, on which $\langle \mathrm{Re}(\tilde{U}_x) \rangle = \langle \mathrm{Re}(\tilde{U}_{x,0}) \rangle = \langle \mathrm{Re}(\tilde{U}_{x,1}) \rangle = 0$. On this state, the costs

are $\mathcal{M}_{\tilde{X}} = \epsilon^2 c_x^2 (1 - \bar{\lambda}_x)^2$ and $\mathcal{M}_{X'} = \epsilon^2 (c_{x,0} + c_{x,1})^2$. As $\bar{\lambda}_x \geq 0$ and $c_{x,0} + c_{x,1} > c_x$, $\mathcal{M}_{X'} < \mathcal{M}_{\tilde{X}}$.

4.C.3 Proof of Lemma 3

In this appendix, we prove that the end-point of norm-preserving decomposition sequences are reflection operators. In other terms, if $\mathrm{Re}(U_x)$ is a reflection operator, it only admits a norm-preserving sub-decomposition [Eq. (4.16)] if $\mathrm{Re}(U_{x,0}) = \mathrm{Re}(U_{x,1}) = \mathrm{Re}(U_x)$.

To prove this, consider a state $|\psi\rangle$ in the $+1$ eigenspace of $\mathrm{Re}(U_x)$. For a norm-preserving decomposition, we must have

$$
\begin{aligned}
c_{x,0} + c_{x,1} = c_x &= c_x \langle \psi | \mathrm{Re}(U_x) | \psi \rangle \\
&= c_{x,0} \langle \psi | \mathrm{Re}(U_{x,0}) | \psi \rangle + c_{x,1} \langle \psi | \mathrm{Re}(U_{x,1}) | \psi \rangle.
\end{aligned}
\tag{4.44}
$$

As $\| \mathrm{Re}(U_{x,0}) \|, \| \mathrm{Re}(U_{x,1}) \| \leq 1$, this equality can only be satisfied if $|\psi\rangle$ is also a $+1$ eigenstate of both $U_{x,0}$ and $U_{x,1}$. A similar argument holds for all -1 eigenstates of U_x, and so $U_{x,0}, U_{x,1}$ and U_x share the same eigenstates and eigenvalues and must be equal. Taking such a sub-decomposition has no effect on the estimator of $\langle O \rangle$, as the same HT are performed and the total number of shots doesn't change, i.e. $\mathcal{M}_{X'} = \mathcal{M}$ in Eq. (4.38).

4.C.4 Examples of reflection decompositions

The simplest example of a reflection-based decomposition is a decomposition in terms of Pauli operators

$$
O = \sum_j^J c_j Z_j,
\tag{4.45}
$$

with $c_j \geq 0$. We could be tempted to measure $\langle O \rangle$ with a single HT circuit (assuming access to a block-encoding of $\frac{O}{\|O\|}$, which is optimal). In this case, as $O = \|O\| \mathrm{Re}(U)$, the bound Eq. (4.15) is

$$
M \geq \epsilon^{-2} \|O\|^2 \left[1 - \frac{\langle O \rangle^2}{\|O\|^2} \right].
\tag{4.46}
$$

To improve on this, we can estimate each $\langle Z_j \rangle$ separately, each with a Hadamard test with $U_j = \langle Z_j \rangle$ (a binary operator). As the spectral norm of O is equal to the induced 1-norm $\|O\|_1 = \sum_j^J c_j$, Eq. (4.45) is a

norm-preserving decomposition. The bound Eq. (4.15) then becomes

$$M \geq \epsilon^{-2} \left[\sum_j c_j \sqrt{1 - \langle Z_j \rangle^2} \right]^2 , \qquad (4.47)$$

which is always smaller or equal than Eq. (4.46) [easily proven through Eq. (4.40)]. This inequality is only saturated when the considered state ρ has support only on the $\|O\|^2$-eigenvalue subspace of O^2; the operator O projected on this subspace is effectively a binary operator.

Norm-preserving decompositions do not need to involve only mutually commuting Pauli operators. As a practical example, we consider the two-qubit operator $O = XX + YY$, which appears commonly in quantum Hamiltonians. As $O = 2 \operatorname{Im}[\text{iSWAP}]$, this operator can be measured with a single Hadamard test circuit. Furthermore, in the context of electronic structure Hamiltonians, O preserves particle number, so in general a control-free scheme using the vacuum as reference state can be employed for the measurement. This operator has three eigenvalues $\{0, \pm 1\}$, which means we can improve its measurement by decomposing it in binary operators. We propose three decompositions $O = \operatorname{Re} U_0 + \operatorname{Re} U_1$ The obvious Pauli decomposition $U_0 = XX, U_1 = YY$ has the downside of not conserving particle number. To fix this, we can take

$$U_j = \frac{1}{2}[(XX + YY) + (-1)^j (Z\mathbb{1} + \mathbb{1}Z)]. \qquad (4.48)$$

These are particle-number preserving, reflection operators and can be easily implemented by combining iSWAP with single-qubit $e^{\pm iZ\pi/4}$ rotations on both qubits. The last decomposition,

$$U_j = \frac{1}{2}[(XX + YY) + (-1)^j (ZZ + \mathbb{1}\mathbb{1})], \qquad (4.49)$$

uses particle-preserving reflection operators with different eigenvalue multiplicities: unlike Pauli operators, the ± 1-eigenvalue subspaces of U_j have unequal dimension 1 and 3. For any state in the 0-eigenvalue subspace, spanned by $\{|00\rangle, |11\rangle\}$, the estimate variance $\operatorname{Var}^*[\langle \operatorname{Re}(U_j) \rangle] = 0$ for decomposition Eq. (4.49). This is not true for the other two decompositions, which indicates that not all decompositions in binary operators are born equal. We will deal with this in the next section. Another example of a few-qubit reflection operator that is a sum of non-commuting Pauli

operators is the three-spin all-to-all Heisenberg coupling

$$O = \frac{1}{3} \sum_{l=1}^{2} \sum_{m=0}^{l-1} X_m X_l + Y_m Y_l + Z_m Z_l, \qquad (4.50)$$

which appears e.g. in the Kagome-Heisenberg Hamiltonian.

4.C.5 Proof of Lemma 4

In this appendix we prove Lemma 4, which formally states the optimality and uniqueness of the Ξ-decomposition. To do this, we first define a variance bound for a class of estimators of $\langle O \rangle$ on a state $|\psi\rangle$. We prove that the bound is achieved on all eigenstates of O if all the sampled operators $\mathrm{Re}(U_x)$ are diagonal in the eigenbasis of O. We then construct the Ξ-decomposition, and prove that the related estimator saturates the bound on the set S_2 of all states with support on at most two eigenstates of O. Finally, we prove no other decomposition satisfies this requirement (i.e. the Ξ-decomposition is unique), and no decomposition satisfies the bound on a superset $S \supset S_2$.

A decomposition X [Eq. (4.10)] of an operator O is optimal on a state $|\psi\rangle$ if no other decomposition produces an estimator with lower cost [Eq. (4.15)] for that state. Optimality can be defined for a set S of states: X is optimal on S if, for each $|\psi\rangle \in S$, no decomposition X' has lower cost $M_{X'} < M_X$. (Note that this can be readily generalized to mixed state, without changing any of our next results.) Lemmas 1-3 imply a necessary condition for optimality on the whole Hilbert space: X can only be optimal on all states if it has the form

$$O = \bar{\lambda}_O \mathbb{1} + \sum_{x \in X} c_x \mathrm{Re}(U_x), \qquad (4.51)$$

$$c_x > 0, \quad |\bar{\lambda}_O| + \sum_{x \in X} c_x = \|O\|, \quad \mathrm{Re}(U_x)^2 = \mathbb{1}$$

where $\bar{\lambda}_O$ is the average of the largest and smalles eigenvalues of O. In other words, X is a norm-preserving decomposition of the center of O where all sampled terms are reflection operators. This condition is not sufficient: as many non-equivalent instances of such decompositions exist, as exemplified in Appendix 4.C.4.

We now construct a bound on the variance of the estimator of $\langle O \rangle$ based on the decomposition X: saturating this bound on all $|\psi\rangle \in S$ implies

optimality of X on S. [The cost of the decomposition Eq. (4.15) is defined as the minimum value of M required to achieve target variance ϵ^2, so minimum variance at fixed M implies minimum cost at fixed ϵ.]

$$\text{Var}_X^*[\langle O \rangle] = \frac{1}{M} \left[\sum_x c_x \sqrt{1 - \langle \text{Re}(U_x) \rangle^2} \right]^2 \geq \frac{\text{Var}[O]}{M}. \tag{4.52}$$

This bound is implied by Eq. (4.9) and Eq. (4.12), with the choice of optimal shot allocation Eq. (4.14). It physical interpretation is rooted in the following observation: a Von Neumann measurement of O is the lowest-variance unbiased estimator of $\langle O \rangle$ when given access to a single state preparation. Thus, given M independent experiments each with a single state preparation, the mean of Von Neumann measurements is the lowest-variance unbiased estimator.

We first consider the set S_1 of all eigenstates of O. For any $|\phi\rangle \in S_1$, the value of the bound in Eq. (4.52) becomes $\text{Var}[O] = 0$. The bound is thus saturated only if we choose all reflection operators $\text{Re}(U_x)$ diagonal in any eigenvector basis of O, i.e. $[U_x, O] = 0$ and $U_x |\phi\rangle = \pm |\phi\rangle$ for any $|\phi\rangle \in S_1$. For any decomposition of this form, we can write all U_x in terms of the eigenspace projectors of O:

$$U_x = \sum_{j=0}^{J-1} \xi_{x,j} \Pi_j, \quad \xi_{x,j} \in \{\pm 1\}, \tag{4.53}$$

where Π_j is the projector on the (eventually degenerate) λ_j-eigenspace of O, J is the number of distinct eigenvalues $\{\lambda_j\}$ of O, and without loss of generality we assume $\lambda_j > \lambda_{j-1}$. The coefficients will then have to satisfy the relation $\lambda_j = \sum_x c_x \xi_{x,j}$.

We define the Ξ-decomposition based on Eq. (4.53), by choosing $\xi_{x,j} = -1$ if $j < x$, and $+1$ otherwise. The resulting decomposition is presented in Lemma 4, Eq. (4.18). The operators Ξ_x are reflections by definition, and it is easy to check that the decomposition satisfied the necessary condition Eq. (4.51). Note that $c_0 = (\lambda_0 + \lambda_j)/2$ defines the optimal identity shift (producing the center of O) and the $c_x = (\lambda_x - \lambda_{x-1})/2$ complete the decomposition.

We now prove that the Ξ-decomposition is optimal on the set S_2 of states with support on two eigenstates of O,

$$S_2 = \left\{ \frac{\alpha |\lambda_m\rangle + \beta |\lambda_n\rangle}{\sqrt{\alpha^2 + \beta^2}} : |\lambda_m\rangle, |\lambda_n\rangle \in S_1 \right\}. \tag{4.54}$$

4

On a general state $|\psi\rangle$ with eigenspace occupations $a_j = \langle\psi|\,\Pi_j\,|\psi\rangle$, the estimator based on the Ξ-decomposition has variance

$$\mathrm{Var}^*_\Xi[\langle O\rangle] = \frac{1}{M}\left[\sum_j \frac{\delta\lambda_j}{2}\sqrt{4(\sum_{i<j} a_i)(\sum_{i\geq j} a_i)}\right]^2. \tag{4.55}$$

For a state $|\phi\rangle \in S_2$, only two occupations are nonzero $a_m, a_n \neq 0$ (we assume w.l.g. $m < n$), thus the term under square root is reduced to $4a_m a_n$ if $m < j \leq n$ and 0 otherwise. The resulting variance

$$\mathrm{Var}^*_\Xi[\langle\phi|\,O\,|\phi\rangle] = \frac{1}{M}\left[\frac{\lambda_n - \lambda_m}{2}\sqrt{4a_m a_n}\right]^2 \tag{4.56}$$

$$= \frac{1}{M}a_n a_m(\lambda_n - \lambda_m) = \mathrm{Var}[\langle O\rangle]$$

thus saturating the bound Eq. (4.12).

We now prove that the only optimal decomposition on S_2 is the Ξ-decomposition (or equivalent up to relabeling and trivial subdecompositions). First of all, $S_1 \subset S_2$, so the terms of the decomposition need to be of the form of Eq. (4.53). Consider a family of states $\sqrt{a_m}\,|\lambda_m\rangle + \sqrt{a_n}\,|\lambda_n\rangle$ for any $n > m$, with only two nonzero eigenstate occupations $a_m + a_n = 1$. On such a state,

$$\mathrm{Var}^*_X[O] = \frac{1}{M}\left[\sum_x c_x\sqrt{1 - [a_m\xi_{x,m} + a_n\xi_{x,n}]^2}\right]^2$$

$$= \frac{a_m a_n}{M}\left[\sum_x 2c_x\frac{1 - \xi_{x,m}\xi_{x,n}}{2}\right]^2. \tag{4.57}$$

The bound Eq. (4.52) is then saturated when

$$\left[\sum_x 2c_x\frac{1 - \xi_{x,m}\xi_{x,n}}{2}\right]^2 = \lambda_n - \lambda_m, \tag{4.58}$$

where we simplified out the free parameter $\frac{a_m a_n}{M}$. This can be rewritten as

$$\sum_x c_x\xi_{x,n}(\xi_{x,n} - \xi_{x,m}) = \sum_x c_x(\xi_{x,n} - \xi_{x,m}) \tag{4.59}$$

using the condition on the decomposition coefficients $\lambda_j = \sum_x c_x\xi_{x,j}$. This

implies that, if $\xi_{x,n} = -1$ then $(\xi_{x,m} - \xi_{x,n}) = 0$ (recall that $c_x > 0$), i.e. $\xi_{x,m} = -1$. Thus the only U_x that can appear in this decomposition, are of the same form as the operators in the Ξ decomposition ($\xi_{j,m} = -1, \xi_{j,n} = +1$ for $m < j \leq n$), and thus X is either Ξ or a trivial sub-decomposition of it.

We now show that the Ξ-decomposition does not saturate the bound Eq. (4.52) for a state $|\psi\rangle$ with three non-zero occupations, $a_m, a_n, a_p \neq 0$ ($m < n < p$). On this state we can write

$$\mathrm{Var}^*_\Xi[\langle O \rangle] = \frac{1}{M} \left[(\lambda_n - \lambda_m) \sqrt{a_m(a_n + a_p)} \right. \tag{4.60}$$

$$\left. + (\lambda_p - \lambda_n) \sqrt{a_p(a_m + a_n)} \right]^2 . \tag{4.61}$$

Subtracting from this $\mathrm{Var}[\langle O \rangle]$, expanding and then collecting terms we get

$$\mathrm{Var}^*_\Xi[\langle O \rangle] - \mathrm{Var}[\langle O \rangle] =$$
$$= [(\lambda_n - \lambda_p)(\lambda_n - \lambda_m)] \cdot \tag{4.62}$$
$$\cdot \left[a_m a_p - \sqrt{a_m a_p (a_m + a_n)(a_p + a_n)} \right] > 0,$$

as both the terms in square brackets are strictly smaller than zero. This (along with the uniqueness of Ξ as the optimal estimator on S_2) implies that no HT-based estimator can saturate the bound Eq. (4.52) for arbitrary states.

In fact, the bound can only be saturated on states in S_2: on these states the Von Neumann measurement has only two possible outcomes (λ_m and λ_n) with nonzero probability. The adaptive shot allocation scheme then ensures (for a large enough M) that most of the measurements we take (Ξ_x with $m \leq x < n$) reproduce the statistics of the Von Neumann measurement, with the single bit we sample in every experiment always distinguishing between λ_m and λ_n. On any state $|\psi\rangle \in \overline{S_2}$, the Von Neumann measurement has three or more outcomes with non-zero probability, and we cannot repoduce its statistics by sampling a single qubit per experiment. This, along with the uniqueness of Ξ, implies that no decomposition can satisfy the sufficient condition for optimality on a superset $S \supset S_2$. The numerical results presented in this paper quantify the increase in variance with respect to the bound, along with confirming the Ξ-decomposition outperforms other decompositions on all states.

4

4.D Implementation of the Ξ decomposition via quantum signal processing

Verifiable samping of QSP polynomials — To measure the operators in the Ξ decomposition Eq. (4.20), we implement a Hadamard test (or EV) on trigonometric polynomials of $(H - \mu_x)t$ generated by the quantum signal processing. We tune the QSP coefficients such that the polynomials approximate the sign function in a suitable range. In this section we display and analyse this technique.

The full circuit we use to achieve this is:

The first control qubit (labeled HT) takes care of the Hadamard test. The second ancilla (labeled QSP) manages the quantum signal processing subroutine, extended through the sign-controlled evolution $e^{-iZ\otimes(O-\mu_x)t}$ to implements a quantum signal processing (QSP) on the operator $e^{(O-\mu_x)t}$. We now describe how the measurement scheme works, and how to select the ϕ parameters to approximate a measurement of $\mathrm{sgn}[(O - \mu_x)t]$ in the interval $[-\pi, \pi]$.

First, we analyze the QSP routine. Let us assume $|\psi\rangle$ to be an eigenstate of $(O - \mu_x)t$ with eigenvalue $\omega \in (-\pi, \pi)$, and only consider the effect of the controlled gates (removing the HT qubit). Then, we can reduce the cicuit to an effective single-qubit gate on the QSP qubit, with action

$$Q_\phi(\omega) = e^{-i\frac{Y}{2}\pi} e^{-i\frac{X}{2}\phi_R} \left[\prod_{r=1}^{R} e^{-i\frac{Z}{2}2\omega} e^{-i\frac{X}{2}\phi_{R-r}} \right]$$

$$= \begin{pmatrix} S_\phi(\omega) & \cdot \\ \cdot & \cdot \end{pmatrix} \tag{4.63}$$

which is a block encoding of $S_\phi(\omega)$, a degree-R trigonometric polynomial of ω. For the sake of simplicity we inserted the final gate $e^{-i\frac{Y}{2}\frac{\pi}{2}} = -iY$, shifting the polynomial of interest S from the block $\langle 1| Q |0\rangle$ to $\langle 0| Q |0\rangle$. We ensure $S_\phi(\omega)$ is real and odd by constraining

$$\phi_r = -\phi_{R-r} \implies S(\omega) = -S(-\omega) \in \mathbb{R}. \tag{4.64}$$

Re-introducing the system register, i.e. taking a general $|\psi\rangle_S$, can be done by linearity taking $Q(\omega) \mapsto Q[(O - \mu_x)t]$ and recovering the circuit above.

The result of the verified Hadamard test (or EV) is obtained by measuring on the output state of the circuit the expectation value of Z_{HT} (or $Z_{HT} \otimes |0\rangle\langle 0|_{QSP} \otimes |\psi\rangle\langle\psi|_S$). (In the absence of noise these two expectation values are equal. In the presence of noise, an additional measurement at $t = 0$ can be taken to mitigate errors. For more details on the technique we refer the reader to the original work on EV [156].)

Approximating the sign function — To approximate the operators Eq. (4.20) that make up the Ξ decomposition, we need to choose the QSP parameter ϕ such that $S_\phi(\omega)$ in Eq. (4.63) approximates sgn$[\omega]$. The polynomial $S_\phi(\omega)$ is odd, real, and 2π-periodic – thus having nodes $S_\phi(0) = S_{\phi(\pm\pi)} = 0$. To account for the approximation error in the neighborhood of these nodes, we introduce a resolution parameter $\delta \geq 0$, and request the approximation to be effective only in the $[\delta, \pi - \delta]$ interval. Choosing $\delta > 0$ implies accepting a larger error in approximating the sign function close to zero. For example, we know the eigenvalues of $(O - \mu_x)t$ closest to zero have absolute value $\frac{\delta\lambda_x}{2}t$, we can use this knowledge to choose δ.

We define a loss function to characterize the quality of the approximation: the average error

$$\mathcal{L}_\delta(\phi) = \frac{1}{\pi - 2\delta} \int_\delta^{\pi+\delta} d\omega \left[\text{sgn}(\omega) - \text{Im}[S_\phi(\omega)] \right]. \tag{4.65}$$

To choose the optimal parameters ϕ, we minimize this loss under the constraints (4.64). Although an analytical approach to this problem is possible building on the techniques described in [48], we take the numerical route to this approximation (which is efficient, scalable and easy to implement). The integral is thus substituted with a sum on a grid with a number of points much larger than the degree of the trigonometric polynomial. We plot in Fig 4.2 the minimized cost function, as a function of the approximation's order R and of the resolution parameter δ. We find that the loss always decays exponentially with an increasing order R, with a decay rate depending on δ.
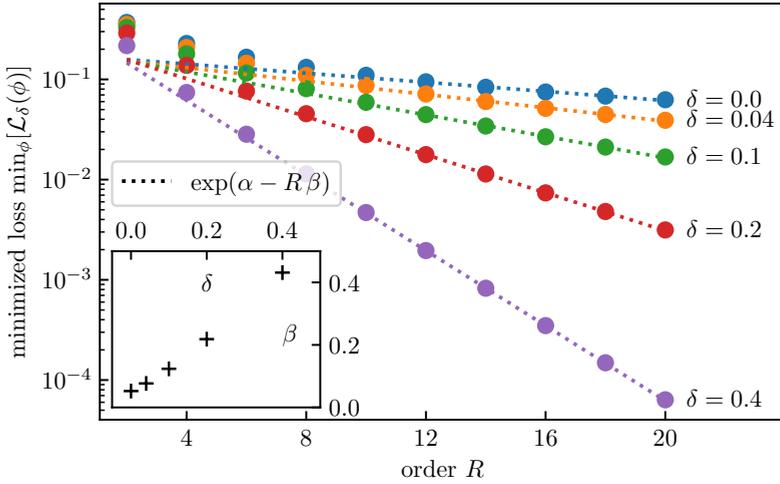
Figure 4.2: Loss Eq (4.65) for the optimal choice of QSP parameters ϕ, as a function of the order R (number of QSP layers) and resolution parameter δ. The dotted lines are log-lin fits for $R > 10$. The dependence of the fit parameter β on the resolution δ is shown in the inset.

4.E The generalized parameter-shift kernel decomposition of a diagonal operator with ladder spectrum

In [159] the authors propose techniques to estimate derivatives $\left\langle \frac{d}{dt} U(t) \right\rangle$ of a unitary $U(t) = e^{iOt}$ generated by O, by sampling $\langle \sin(Ot_l) \rangle = \langle \mathrm{Re}[U(t_l)] \rangle$ at a discrete set of points $\{t_l\}$. This technique can be used to estimate expectation values of O, as $\langle O \rangle = \left\langle [-i \frac{d}{dt} e^{iOt}]_{t=0} \right\rangle$, and it is clearly compatible with Hadamard test or EV measurements (as it only requires sampling $\langle \mathrm{Re}\, U(t_l) \rangle$).

For an operator O with equispaced eigenvalues $\Omega, 2\Omega, ..., R\Omega$ (commonly referred to as a "ladder spectrum"), the authors give a choice of $\{t_l\}$ and explicit coefficients $c_l(t)$ for the linear combination $\left\langle \frac{d}{dt} U(t) \right\rangle = \sum_l c_l \langle \mathrm{Re}[-iU(t_l)] \rangle$. Assuming $\Omega = 1$ (which can be considered a choice of units for the energy), the time points are chosen as $\{t_l = \frac{2l}{2R+1}\pi\}$. We

can then define a modified version of the Dirichelet kernel,

$$\tilde{D}_l(t) = \frac{1}{R}\cos(t_l)\left[\frac{1}{2}\sin(Rt) + \sum_{j=1}^{R-1}\sin(jt)\right], \qquad (4.66)$$

which satisfies $\tilde{D}_l(t_{l'}) = \delta_{ll'}$. This is a linear combination of the R basis functions $\{\sin(jt)\}_{j=1,\dots,R}$, like $\langle\sin(Ot)\rangle$. Thus, as the equality

$$\langle\sin(Ot)\rangle = \sum_{l=1}^{R}\langle\sin(Ot_l)\rangle\,\tilde{D}_l(t) \qquad (4.67)$$

holds for all $\{t_l\}_{l=1,\dots,R}$, it must to hold for all t. We can then differentiate the kernel rather than the expectation value itself. Evaluating $[\frac{\partial}{\partial t}\tilde{D}_l(t)]_{t=0}$ and combining the equations above we obtain

$$\langle O \rangle = \sum_{l=1}^{R}\frac{(-1)^{l-1}}{2R\sin^2(\frac{1}{2}t_l)}\,\langle\sin(Ot_l)\rangle \qquad (4.68)$$

$$= \sum_{l=1}^{R}c_l\,\langle\mathrm{Re}[-iU(t_l)]\rangle\,. \qquad (4.69)$$

This matches the form of decompositions Eq. (4.10). We call thi the generalized parameter shift kernel (GPSK) decomposition. Under the optimal shot allocation choice [Eq. (4.14)], the shot-variance of the estimator based on this decomposition is

$$M\,\mathrm{Var}^*_{\mathrm{GPSK}} = \left[\sum_{l=1}^{R}\frac{\sqrt{1-\langle\sin(Ot_l)\rangle^2}}{|2R\sin^2(\frac{1}{2}t_l)|}\right]^2 \qquad (4.70)$$

4.F Details on numerical simulations and further numerical results

We measure the variances on random states generated by hardware-efficient ansatzes using PennyLane [171]. For each value of N, 100 random set of parameters (and therefore 100 random states) are generated and measured for all decompositions. For each decomposition X, we first use 10^5 shots (allocated proportionally to the weight of each term) to obtain a rough
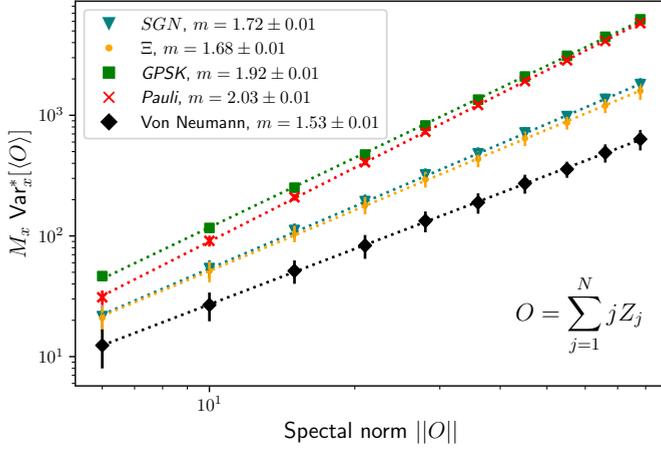
Figure 4.3: Comparison study of variances of different decompositions on random states generated by a hardware-efficient ansatz (see text for details). Different colours correspond to different decompositions [Eq.(4.10)] of the target operator O (see text for the description of all decompositions). Dashed lines are exponential fits ($a \exp(mN + b)$) to the data (the parameter m is given in legend).

estimate of the expectation value of each term $\langle \mathrm{Re}(U_x) \rangle$ for $x \in X$. These values are plugged in Eq. (4.14) to get an estimate of the optimal shot allocation ratios $r_x = \frac{m_x}{M_x}$. The variance of each term $\mathrm{Var}^*[\langle \mathrm{Re}(U_x) \rangle]$ is obtained by Eq. (4.8) (or by sampling in the case of the QSP-approximation decomposition 'SGN'). With these we compute the final shot-variance $M_X \mathrm{Var}^*_X[\langle O \rangle] = \sum_{x \in X} r_x^{-1} \mathrm{Var}^*[\langle \mathrm{Re}(U_x) \rangle]$. Finally, we average the values of $M_X \mathrm{Var}^*_X[\langle O \rangle]$ obtained for each random state. This average is the quantity reported in Fig. 4.1, Fig. 4.3 and Fig. 4.4.

The terms Ξ_x are constructed as per Eq. (4.18) using the known eigenvectors of O, and projectively measured on the prepared state (as these are reflection operators, Hadamard test samples match projective measurement samples). The terms in the Pauli decomposition are also directly measured on the prepared state. The GPSK-decomposition is constructed as described in 4.E and measured through a Hadamard test. The Von Neumann variance $\mathrm{Var}[O]$ is computed analytically.

The QSP approximation of Ξ (denoted SGN from the sign term approx-
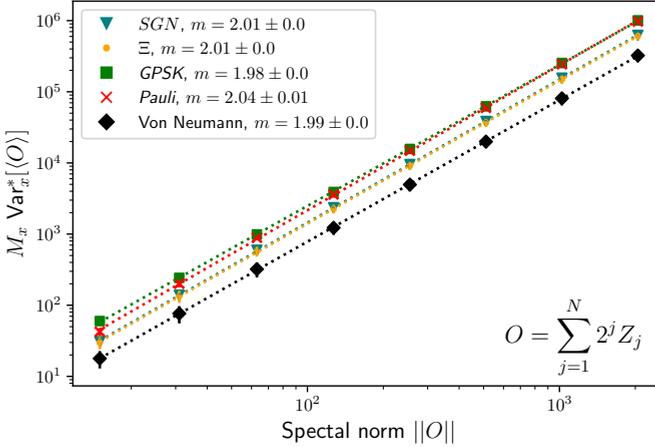
Figure 4.4: Comparison study of variances of different decompositions on random states generated by a hardware-efficient ansatz (see text for details). Different colours correspond to different decompositions [Eq.(4.10)] of the target operator O (see text for the description of all decompositions). Dashed lines are exponential-power-law fits $(\exp(aN^2 + bN + c))$ to the data (the dominant scaling parameter a is given in the legend).

imation) is implemented as described in Appendix 4.D for $R = 20$ and $\delta = 0$. For fair comparison with the other methods, echo verification is not used. The comparison between the Ξ and SGN decomposition shows how the approximation increased the final variance. (The approximation also introduces a bias, see Appendix 4.D.

All the simulations assume Hadamard-test-based measurement in an ideal circuit simulation: no circuit-level noise is considered and EV is not implemented.

We additionally report scaling results for the shot-variances of two other observables, $O = \sum_j j Z_j$ and $O = \sum_j 2^j Z_J$. The overall scaling of all decompositions matches the scaling of the operator norm $\|O\|$. Similarly to the case of Fig. 4.1, the Ξ decomposition performs best, the SGN approximation has a relatively small effect on the shot-variance, and the Pauli decomposition shows the worst scaling.

Virtual mitigation of coherent non-adiabatic transitions by echo verification

5.1 Introduction

The study of quantum many-body systems requires the precise estimation of observables. Quantum state preparation is naturally a prerequisite to this end, which is the rationale behind quantum computers or quantum simulators. The adiabatic algorithm has demonstrated large success in a variety of platforms [175, 176]. Still, the performance of current devices is hindered by noise, which cannot be error corrected, yet. Therefore, error mitigation techniques have been explored both theoretically and experimentally and can significantly improve the estimation of observables [69, 79, 177]. Surprisingly, there have been few synergies jointly considering error mitigation for the adiabatic algorithm.

Any quantum circuit can be efficiently simulated by the adiabatic algorithm [178]. In adiabatic quantum computation, the system is initialized in the ground state of a trivial Hamiltonian and one seeks to prepare the ground state of the final Hamiltonian by slowly interpolating between the two. The success of the algorithm is determined by the speed of the adiabatic passage and spectral properties of the Hamiltonians [49, 179]. More precisely, the total evolution time, or circuit depth, depends inverse
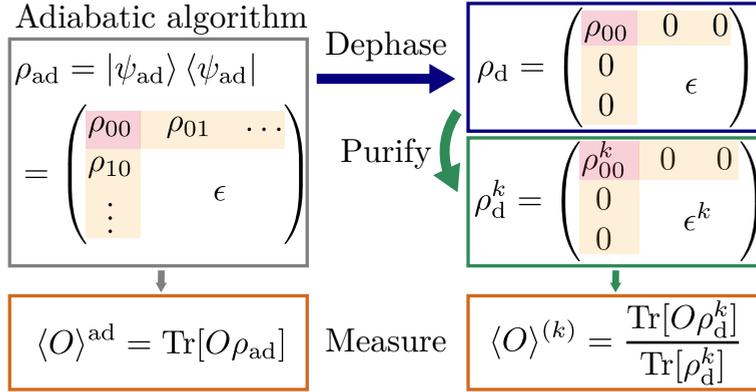
Figure 5.1: Schematic overview of the method. Density matrices are expressed in the energy eigenbasis of the target Hamiltonian. The pure state after the adiabatic evolution (ρ_{ad}) approximates the true ground state. Via a dephasing operation, the coherent error is promoted to an incoherent error in ρ_{d} such that error mitigation techniques can be applied. This allows measuring the kth degree purified observable $\langle O \rangle^{(k)}$ which yields a lower bias than evaluating the state directly after the adiabatic preparation $\left[\langle O \rangle^{\mathrm{ad}} \right]$.

polynomially on the minimum spectral gap between the ground state and the first excited state along the adiabatic path. These relations are quantified by the adiabatic theorem and versions thereof [51–53]. The adiabatic algorithm is especially suited for devices that implement dynamics natively without any Trotter overhead [180–183].

To address the restrictions in current hardware, various error mitigation techniques have been explored in recent years to improve the usefulness of a noisy quantum computation [79]. These methods include zero-noise extrapolation, exploiting symmetry or purity constraints, and several other approaches. Here, we focus on purity methods, which aim to suppress stochastic errors by projecting the noisy state ρ onto the closest pure state, given by the dominant eigenvector of ρ.

The purification can in general be achieved by collective measurements of several copies of ρ, known as virtual state distillation [148] or error suppression by derangement [184]. Echo verification (EV) achieves this using two copies of ρ multiplexed in time, rather than in space [156–158]. In EV, a desired state is prepared, an observable is measured
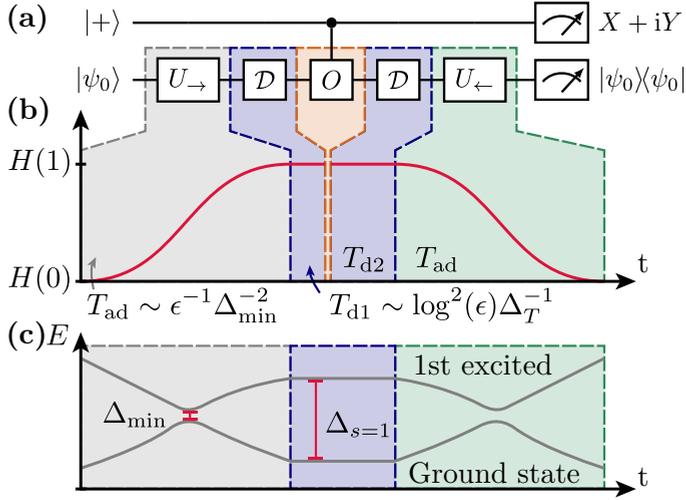
Figure 5.2: **(a)** Quantum circuit for adiabatic echo verification to estimate an observable $\langle O \rangle$. A quasi-adiabatic sweep $U_\rightarrow$ is followed by an approximate ground state dephasing operation $\mathcal{D}$. After the controlled application of a unitary observable O and dephasing again, the sweep is performed backward $U_\leftarrow (\neq U_\rightarrow^\dagger)$. Postprocessing the measurement result, including the success information of the ground state projection, allows to extract an improved expectation value. **(b)** Schematic of the Hamiltonian dynamics. Approximate dephasing is implemented by evolving with the target Hamiltonian at $s = 1$ for a random time. Typically, this time is much smaller than the time required for the adiabatic algorithm as depicted in **(c)**, where we sketch a corresponding low-energy spectrum.

controlled by an auxiliary qubit, and the state is then uncomputed. This allows to access expectation values of the so-called 2nd degree purified state of ρ: $\langle O \rangle_{\mathrm{EV}} = \mathrm{Tr}\big[O\rho^2\big] / \mathrm{Tr}\big[\rho^2\big]$. Recently, purification-based error mitigation has been tested experimentally in the context of the variational quantum eigensolver [69]. Error mitigation methods tailored specifically to the adiabatic algorithm have been explored considerably less in the literature. Few exceptions consider error suppression and correction [185] or symmetry-protection for Trotter dynamics [186].

In this work, we present a mitigation technique for estimating observables on quasi-adiabatically-prepared states, in the spirit of echo verification (Fig. 5.1). Along with stochastic device noise, our method seeks to sup-

press the coherent error due to non-adiabatic transitions. Our method which we denote *Adiabatic Echo Verification* (AEV) relies on dephasing operations to promote the coherent errors to random errors, which can then be mitigated. Similar to the original echo verification technique, the leading order error in the ground state expectation value of an observable is suppressed quadratically. In particular, we consider an imperfect implementation of the dephasing operation using random-time evolution. Related random-time dynamics have been successfully used in the context of Zeno-type protocols [187]. The overhead from this dephasing operation is only poly-logarithmic in the accuracy of the dephasing operation for estimating observables of states within gapped phases. We discuss how the protocol compares favorably against doubling the total evolution time in the standard adiabatic algorithm. A key feature of our technique is that hardware noise is also mitigated naturally through the EV method. Our protocol only requires implementing positive-time evolution and applying the operator of interest in a controlled way. Hence, the protocol is not only suitable for purely gate-based quantum devices but also for hybrid quantum simulators, e.g. using neutral Rydberg atoms [182].

5.2 The adiabatic algorithm and purification-based error mitigation

In order to be able to measure observables on the ground state $|E_0\rangle$ of a target Hamiltonian H_T, a state approximating $|E_0\rangle$ with sufficient precision needs to be prepared. The quantum adiabatic algorithm (QAA) is a suitable algorithm for this task. At the heart of the QAA is the adiabatic theorem, which states that a system remains in an instantaneous eigenstate if the Hamiltonian is changed sufficiently slowly and the eigenstate is separated from other eigenstates by a minimum spectral gap $\Delta_{\min}$ throughout the transition [50]. Hence, the desired ground state $|\psi_T\rangle$ of a Hamiltonian of interest H_T can be prepared by interpolating from a suitable Hamiltonian H_0 with a trivial ground state $|\psi_0\rangle$ as

$$H(s) = (1 - s)H_0 + sH_T. \tag{5.1}$$

where $s = t/T$ is the parametrized time. The folk version of the adiabatic theorem states that a total time $T = \mathcal{O}\left(\Delta_{\min}^{-2}\epsilon^{-1/2}\right)$ suffices to prepare the ground state up to fidelity $1 - \epsilon$. Rigorous versions of the adiabatic theorem give a bound $T = \mathcal{O}\left(\Delta_{\min}^{-3}\epsilon^{-1/2}\right)$ if $H(s)$ is twice differentiable [51, 52]. Given a finite coherence time, the QAA prepares an approximation

to the target state $|\psi_{\mathrm{ad}}\rangle = \sqrt{1-\epsilon}\,|E_0\rangle + \sqrt{\epsilon}\,\big|E_0^\perp\big\rangle$ where $\langle E_0|E_0^\perp\rangle = 0$. Measuring an observable O, we obtain an approximation to the true value $\mathrm{Tr}\,[O\,|\psi\rangle\langle\psi|_{\mathrm{ad}}] = (1-\epsilon)\langle O\rangle_{|E_0\rangle} + \mathcal{O}(\sqrt{\epsilon})$.

Purification methods such as echo verification (EV) or virtual state distillation improve the quality of an expectation value measurement on a noisy (incoherent) approximation ρ of a pure state $|\psi\rangle\langle\psi|$. This is achieved by effectively measuring the expectation value $\mathrm{Tr}\big[O\rho^k\big]$ of O on the k-th power of the density matrix. Raising ρ to the k-th power suppresses the eigenvectors with smaller eigenvalues, increasing the relative weight of the dominant eigenvector which, for small enough noise, should be $|\psi\rangle$. As ρ^k is non-normalized, purification methods prescribe to independently measure $\mathrm{Tr}\big[\rho^k\big]$ to calculate the desired estimator

$$\langle O\rangle^{(k)} := \mathrm{Tr}\big[O\rho^k\big]\big/\,\mathrm{Tr}\big[\rho^k\big]. \tag{5.2}$$

If ρ has an eigenstate $|E_0\rangle$ with large weight $c_0 = 1-\epsilon$ (small positive ϵ), we can write the density matrix as $\rho = c_0\,|E_0\rangle\langle E_0| + \epsilon\rho_\perp$ with $\rho_\perp$ a density matrix orthogonal to $|E_0\rangle$ (i.e., $\rho_\perp\,|E_0\rangle = 0$). The kth degree purified estimator is then

$$\langle O\rangle^{(k)} = \frac{c_0^k\,\langle E_0|\,O\,|E_0\rangle + \epsilon^k\,\mathrm{Tr}\big[\rho_\perp^k O\big]}{c_0^k + \epsilon^k\,\mathrm{Tr}\big[\rho_\perp^k\big]} \tag{5.3}$$

$$= \langle E_0|\,O\,|E_0\rangle + \mathcal{O}(\epsilon^k\,\mathrm{Tr}\big[\rho_\perp^k\big]\|O\|), \tag{5.4}$$

where $\|\cdot\|$ is the operator norm. Echo verification implements purification for $k = 2$ using a single register by multiplexing two state-(un)preparation oracles in time. The method suppresses the error contributions $\mathcal{O}(\epsilon)$ such that the leading order becomes $\mathcal{O}(\epsilon^2)$.

5.3 Mitigating coherent errors in adiabatic state preparation

Our main contribution is to propose a method where the echo verification technique is applied to coherent errors. We focus on an application where the coherent error arises in the adiabatic algorithm due to finite algorithm runtimes. However, as the state prepared by a noiseless implementation of the adiabatic algorithm is pure, naive purification will not have any effect.

To recover the error mitigation power on $\rho_{\mathrm{ad}} = |\psi_{\mathrm{ad}}\rangle\langle\psi_{\mathrm{ad}}|$, we introduce

an ideal dephasing channel that turns coherent errors into incoherent noise,

$$\mathrm{deph}_{H}[\rho] := \sum_{j} |E_j\rangle\langle E_j| \, \rho \, |E_j\rangle\langle E_j| = \mathrm{diag}[\rho], \tag{5.5}$$

where we sum over an eigenbasis $\{|E_j\rangle\}_j$ of the target Hamiltonian H_T. Here, we assume a nondegenerate spectrum and give an extension for degenerate spectra in the Supplement. The dephasing channel projects a density matrix onto its diagonal in the energy eigenbasis, removing the off-diagonal coherences. Applying the channel to the state prepared by the adiabatic algorithm yields

$$\rho_{\mathrm{d}} := \mathrm{deph}_{H}[\rho_{\mathrm{ad}}] = c_0 \, |E_0\rangle\langle E_0| + \epsilon\rho_\perp = \begin{pmatrix} c_0 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \epsilon\rho_\perp & \\ 0 & & & \end{pmatrix} \tag{5.6}$$

with $\rho_\perp = \sum_{j\neq 0} \rho_{jj} \epsilon^{-1} \, |E_j\rangle\langle E_j|$. Then, using the echo verification technique on the dephased state, which is a mixed state, we obtain the following result for the observable O:

$$\frac{\mathrm{Tr}[O\rho_{\mathrm{d}}^k]}{\mathrm{Tr}[\rho_{\mathrm{d}}^k]} = (1-\gamma)\,\langle E_0| \, O \, |E_0\rangle + \gamma \frac{\mathrm{Tr}[O\rho_\perp^k]}{\mathrm{Tr}[\rho_\perp^k]}, \tag{5.7}$$

with $\gamma = \left[1 + c_0^k / \left(\epsilon^k \, \mathrm{Tr}[\rho_\perp^k]\right)\right]^{-1} \sim \mathcal{O}\left(\epsilon^k \, \mathrm{Tr}[\rho_\perp^k]\right)$.

To implement echo verification, typically, an inverse pair of unitaries $(U_\rightarrow, U_\rightarrow^\dagger)$ would be required [156]. The unpreparation $U_\rightarrow^\dagger$ then uses negative-time dynamics, which is generally not available in analog simulators. For our purposes, however, we can consider the two states $\rho_{\mathrm{ad}} = U_\rightarrow \, |\psi_0\rangle\langle\psi_0| \, U_\rightarrow^\dagger$ and $\sigma_{\mathrm{ad}} = U_\leftarrow^\dagger \, |\psi_0\rangle\langle\psi_0| \, U_\leftarrow$, where $U_\leftarrow$ is a positive-time adiabatic evolution with an inverted schedule from $s = 1$ to $s = 0$ [cf. Fig. 5.2(b)]. Both states have the same guaranteed fidelity with the target state $|E_0\rangle$ from the adiabatic theorem and ground state coherences are surpressed after the dephasing operation. This allows to use positive-time dynamics for the unpreparation step in AEV.

Next, we consider the implementation of the dephasing channel. Importantly, we observe that a channel that dephases only the ground state would also be sufficient to achieve our goal, producing a state of the form Eq. 5.6 with a more general, non-diagonal $\rho_\perp$, provided that c_0 still dominates. In the following part, we analyze such an approximate dephasing operation using positive-time dynamics.

5.4 Implementation and cost of the dephasing

We can implement an approximation of the dephasing channel [Eq. 5.5] by a random-time evolution $\exp(-iH_T\tau)$, with τ sampled from a probability distribution $P(\tau)$, as follows: We limit the support of P to the interval $\tau \in [0, T_{\mathrm{d}}]$. This ensures the dephasing can be realized naturally in quantum simulators and limits the time overhead of the dephasing operation to $2\,T_{\mathrm{d}}$ for the AEV circuit. We define the approximate dephasing channel

$$\operatorname*{deph}_{H,P}[\rho] := \int_0^{T_{\mathrm{d}}} \mathrm{d}P(\tau) e^{-iH_T\tau} \rho\, e^{iH_T\tau} \tag{5.8}$$

$$= \sum_{j,k} \mathcal{F}_{jk} \left|E_j\middle\rangle\middle\langle E_j\right| \rho \left|E_k\middle\rangle\middle\langle E_k\right|, \tag{5.9}$$

where $\mathcal{F}_{jk} := \mathcal{F}[P](E_j - E_k)$ is the Fourier transform of the random-time distribution at the transition energies. We will make use of the shorthand $\mathcal{D}[\rho] := \operatorname{deph}_{H,P}[\rho]$. As we only need to dephase the ground state, we require $\max_{j>0} |\mathcal{F}_{0j}| < \delta$. Evaluating the adiabatic echo verification circuit (Fig. 5.2) with the *approximate* dephasing channel $\mathcal{D}[\rho]$ yields an estimator with expectation

$$\langle O \rangle^{\mathrm{AEV}} = \frac{\operatorname{Tr}[O\tilde{\rho}\tilde{\sigma}]}{\operatorname{Tr}[\tilde{\rho}\tilde{\sigma}]} \tag{5.10}$$

where $\tilde{\rho}_{jk} = \mathcal{F}_{jk}[\rho_{\mathrm{ad}}]_{jk}$ and $\tilde{\sigma}_{kl} = \mathcal{F}_{kl}^*[\sigma_{\mathrm{ad}}]_{kl}$, expressed as matrix elements in the eigenbasis of the target Hamiltonian (cf. Supplement). We can bound the deviation of the AEV estimator from the ground state expectation value as

$$\left| \langle O \rangle^{\mathrm{AEV}} - \langle E_0 | O | E_0 \rangle \right| \lesssim \|O\|(\epsilon^{1/2}\delta + \epsilon^2) \tag{5.11}$$

with a small prefactor. To ensure this error is bounded by $\mathcal{O}(\epsilon^2)$, it is then sufficient to take $\delta \sim \epsilon^{3/2}$.

An upper bound on the $|\mathcal{F}_{0j}|$ can be obtained as a functional of the distribution $P(\tau)$. We can thus redefine

$$\delta := \max_{\Delta > \Delta_T} \left| \mathcal{F}[P](\Delta) \right| \tag{5.12}$$

where $\Delta_T < E_1 - E_0$ is a lower bound on the target Hamiltonian ground state gap. In principle, different distributions can be chosen. We might, for example, simply choose a uniform distribution $P(\tau) = 1/T_{\mathrm{d}}$ for $\tau \in [0, T_{\mathrm{d}}]$. As its Fourier transform is the cardinal sine function $\sin(x)/x$, we obtain

$\delta \sim (\Delta T_{\mathrm{d}})^{-1}$. However, discontinuities in P or its derivatives limit the asymptotic decay of $\mathcal{F}[P]$ to a polynomial. We can improve upon this without increasing the maximal evolution time by choosing a mollifier, i.e. a smooth distribution supported on $[0, T_{\mathrm{d}}]$. A suitable example for our purposes is the rescaled bump function

$$P_{T_{\mathrm{d}}}(\tau) = \begin{cases} \frac{2}{\mathcal{N} T_{\mathrm{d}}} \exp\left[\frac{T_{\mathrm{d}}^2}{4\tau(\tau - T_{\mathrm{d}})}\right] & \text{if} \quad \tau \in [0, T_{\mathrm{d}}], \\ 0 & \text{otherwise}, \end{cases} \tag{5.13}$$

where $\mathcal{N} \approx 2.25$ is a normalization factor. The Fourier transform of this function decays super-polynomially. Adapting the results from Ref. 188, we recover

$$\delta < \sqrt{\frac{8\pi}{\sqrt{e}}} \, (T_{\mathrm{d}} \Delta_T)^{-3/4} \exp\left[-\sqrt{T_{\mathrm{d}} \Delta_T / 2}\right], \tag{5.14}$$

with the full derivation included in the Supplement. A dephasing time $T_{\mathrm{d}} \sim \Delta_T^{-1} \log^2[\epsilon]$ is thus sufficient to achieve an overall error $\mathcal{O}(\epsilon^2)$. Often, one is interested in observables of states in gapped phases [189], such that only the poly-logarithmic term contributes to a non-constant overhead.

5.5 Comparison with standard adiabatic algorithm.

We seek to compare the method proposed here with the trivial alternative for improving the performance of the adiabatic algorithm, which is simply doubling the evolution time in the QAA. In the standard adiabatic theorem, there is a polynomial relationship between the accuracy and the evolution time [51]. In principle, the adiabatic theorem can be improved towards an exponential error dependence by assuming a sufficiently smooth schedule with vanishing derivatives at the beginning and end of the schedule [53]. However, this is at the cost of passing the minimal spectral gap at a faster rate, which, in general, leads to more transitions. In practice, we can observe an exponential scaling (*Landau-Zener* regime) transitioning into an inverse-quadratic scaling for longer times [190].

Regarding our method, we therefore conclude that if the error dependence was indeed exponential, as in a Landau-Zener problem, the AEV would yield a performance comparable to the QAA with double the evolution time. Compared to the standard theorems with a polynomial dependence, our method improves up to quadratically. For the sake of
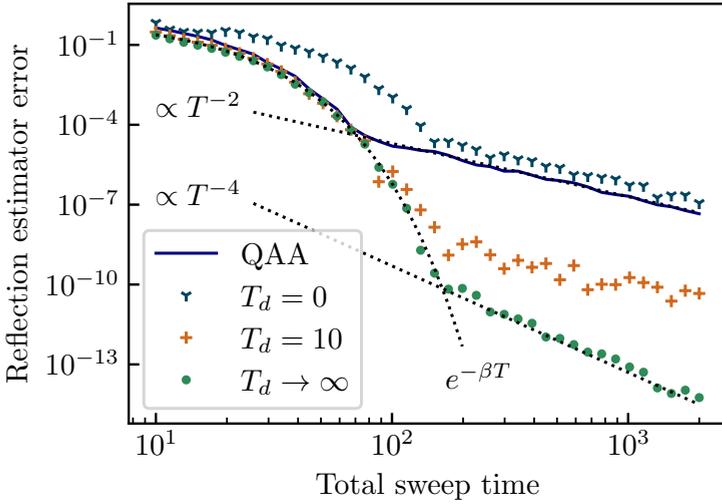
Figure 5.3: Comparing the QAA and AEV for different dephasing times T_d, as a function of the total sweep time T (QAA: $T = T_\mathrm{ad}$, AEV: $T = 2\,T_\mathrm{ad}$). AEV improves over simply doubling the QAA sweep time in the regime of polynomial error dependence. The respective estimator bias from $\langle E_0 | O | E_0 \rangle$ is shown, where $O = 1 - 2 |E_0\rangle\langle E_0|$ is a reflection on the target state. The target Hamiltonian is an Ising model $H_T = 0.2 \sum_{j=1}^{5} \sigma_j^z - \sum_{j=1}^{4} \sigma_j^z \sigma_{j+1}^z$, and $H_0 = \sum_{j=1}^{5} \sigma_j^x$, using a linear schedule. We perform density-matrix simulations; time-dependent evolution is implemented by Euler integration, ensuring a sufficiently small error when discretizing the sweep. Approximate dephasing is implemented with $P(\tau)$ as in Eq. 5.13.

concreteness, we include numerical benchmarks in Fig. 5.3 showing an advantage for preparing the ground state of a transverse field Ising model.

5.6 Discussion and practical considerations.

In this letter, we introduced *Adiabatic Echo Verification* (AEV), a scheme to mitigate the coherent errors that characterize adiabatic state preparation. Our method is tailored to current quantum devices, which lack the possibility to correct errors. AEV requires doubling the circuit time compared to standard adiabatic state preparation, but improves up to

quadratically in the estimator bias. The additional features of the protocol are the following. First, in order to implement the verification part of the circuit, the path of the quasi-adiabatic evolution is simply reversed. Moreover, we show how the dephasing operation can be approximately implemented with positive-time dynamics. Hence, only positive-time evolution is required in AEV. This makes our method suitable for quantum computers that operate in a hybrid mode of digital gates and analog simulation. Rydberg atom arrays have recently demonstrated such capabilities [182, 191]. Additionaly, AEV naturally mitigates non-coherent hardware noise through echo verification. While our focus in the paper was not on the technical analysis of hardware noise mitigation, the results in echo verification literature fully apply to our case [69, 156].

We note that our method is compatible with arbitrary sweep profiles in the QAA. This is especially helpful as it is well known that slowing down the adiabatic sweep at the position of the minimum spectral gap mitigates transitions out of the ground state [192, 193]. More generally, our technique can be applied to other coherent approximate state preparation approaches, such as variational quantum algorithms (VQAs) [194]. This applies to VQAs that prepare a pure state heuristically by a parametrized operation $U(\theta)$ aiming at approximating the desired ground state. By dephasing the prepared state and using the echo verification technique, unpreparing the state with $U(\theta)^\dagger$, we expect that the performance of VQAs can be improved.

We note that the control-free versions of echo verification [69, 156], which employ a reference state instead of a control qubit, are not naively available for AEV. This is due the dephasing channel annihilating coherences between the reference state and the state of interest. Recently, a method for rescaling survival probabilities was considered that has similarities with control-free echo verification [195]. While their unnormalized estimator $\text{Tr}[\rho O \rho O^\dagger]$ differs from the echo verification counterpart $\text{Tr}[O \rho^2]$, it would still allow for mitigating errors for certain interesting observables such as out-of-time-order correlators (OTOCs) [146]. Not requiring the implementation of a controlled operation can significantly simplify experiments. This is why an extension of AEV without a control qubit is an interesting direction for future work. Another promising research direction is the combination of AEV with other purification-based error mitigation methods such as virtual state distillation [148, 184]. Using multiple copies of the quasi-adiabatically prepared state, further improvements for surpressing errors seem possible.

5.A Dephasing operation on a degenerate spectrum

In Eq. 5.5 we define the perfect dephasing channel for an operator H with non-degenerate spectrum. If the spectrum of H contains degeneracies, the dephasing channel will project ρ to a block-diagonal operator, where the blocks are defined by the (degenerate) eigenspaces of H:

$$\underset{H}{\mathrm{deph}}[\rho] = \sum_j \Pi_j \rho \Pi_j \tag{5.15}$$

where $\Pi_j = \delta(H - E_j)$ is the projector on the eigenspace of H with eigenvalue E_j.

For AEV, we are only interested in dephasing the ground state with respect to the rest of the spectrum; we only require the ground state of H to be non-degenerate for Eq. 5.6 and the subsequent analysis to be valid. This is anyway a typical requirement in adiabatic state preparation.

5.B Evaluation of AEV estimator with approximate dephasing

In this section, we evaluate the error on the Adiabatic Echo Verification (AEV) estimator with respect to the target value $\langle E_0 | O | E_0 \rangle$. We bound it as a function of the adiabatic state (un)preparation error ϵ and the dephasing approximation error δ. The AEV circuits we consider only require positive-time evolution with respect to the adiabatic Hamiltonian Eq. 5.1, and the ability to perform a controlled-O operation (or a decomposition thereof) to implement the Hadamard test.

We recall the echo verification (EV) expectation value estimator [69, 196] is defined as

$$\langle O \rangle^{\mathrm{EV}} := \frac{\mathbb{E}[\text{verified Hadamard test circuit (VHT)}]}{\mathbb{E}[\text{echo circuit (Echo)}]}. \tag{5.16}$$

In our case, the verified Hadamard test circuit is

$$
\text{VHT} := \quad
\begin{array}{c}
|+\rangle \\
|\psi_0\rangle
\end{array}
\quad
\boxed{U_\rightarrow}\ \boxed{\mathcal{D}}\ \boxed{O}\ \boxed{\mathcal{D}}\ \boxed{U_\leftarrow}
\quad
\begin{array}{c}
X + iY \\
|\psi_0\rangle\langle\psi_0|
\end{array}
\quad ,
$$

(5.17)

where $U_\rightarrow$ is the adiabatic state preparation, $U_\leftarrow$ is the state unpreparation, and deph $:= \text{deph}_{H,P}$ is the approximate dephasing. At the end of the circuit we need to measure $X \otimes |\psi_0\rangle\langle\psi_0|$ and $Y \otimes |\psi_0\rangle\langle\psi_0|$ to recover the real $\text{Re}[\langle O\rangle^{\text{EV}}]$ and imaginary $i\,\text{Im}[\langle O\rangle^{\text{EV}}]$ parts of the expectation value, respectively. The output of a single sample of the circuit will be the result of the Pauli X (± 1) or Y ($\pm i$) on the control qubit if the system register returns to the state $|\psi_0\rangle$, and 0 otherwise. Our notation supposes that O is a unitary operator, and its application is controlled by the state of the control qubit. If O is not unitary, we can rewrite it as a decomposition $O = \sum_x a_x \text{Re}[U_x] + b_x \text{Im}[U_x]$ and measure the terms of the decomposition separately [196]. The echo circuit, which is used to compute the normalization of $\langle O\rangle^{\text{EV}}$, is given as

$$
\text{Echo} := \quad |\psi_0\rangle \quad \boxed{U_\rightarrow}\ \boxed{\mathcal{D}}\ \boxed{\mathcal{D}}\ \boxed{U_\leftarrow} \quad |\psi_0\rangle\langle\psi_0| \quad\quad (5.18)
$$

and obtained by substituting the operator O with the identity in the previous circuit.

The adiabatic preparation and the adiabatic unpreparation are defined as

$$
U_\rightarrow = \mathcal{T} \exp\left\{ -i \int_0^{T_{\text{ad}}} H[s(t)]\,\mathrm{d}t \right\} ; \tag{5.19}
$$

$$
U_\leftarrow = \mathcal{T} \exp\left\{ -i \int_0^{T_{\text{ad}}} H[s(T_{\text{ad}} - t)]\,\mathrm{d}t \right\}, \tag{5.20}
$$

where $\mathcal{T} \exp$ notates the time-ordered exponential, $H(s)$ is the adiabatic Hamiltonian (5.1), $s(t)$ the adiabatic schedule with $s(0) = 0$ and $s(T_{\text{ad}}) = 1$, and T_{ad} is the total evolution time of the adiabatic algorithm. Note that $\mathrm{d}t$ is always positive, thus negative-time evolution is not required to implement $U_\rightarrow$ and $U_\leftarrow$. Typically, in an EV circuit, if the preparation unitary is U, then the unpreparation is performed with its conjugate transpose $U^\dagger$ such

that $UU^\dagger = \mathbb{1}$. Here, however, this is not the case for the two operations $U_\leftarrow$ and $U_\rightarrow$. We will show that the added dephasing indeed removes this requirement for our purposes.

In our calculations, we only assume that $U_\rightarrow$ ($U_\leftarrow$) implement an approximate state (un)preparation of $|E_0\rangle$ with a fidelity of at least $1 - \epsilon$ with small $\epsilon > 0$. Concretely, we define

$$|\langle E_0|\, U_\rightarrow\, |\psi_0\rangle\,|^2 = 1 - \epsilon_\leftarrow\,, \quad |\langle\psi_0|\, U_\leftarrow\, |E_0\rangle\,|^2 = 1 - \epsilon_\rightarrow, \qquad (5.21)$$

such that $\epsilon = \max\{\epsilon_\leftarrow, \epsilon_\rightarrow\}$. It is reasonable to assume the two adiabatic processes will have a similar error, as any adiabatic theorem bounds both in the same way.

The approximate dephasing channel, of the form Eq. 5.8, is defined via a matrix of Fourier coefficients

$$\mathcal{F}_{jk} := \mathcal{F}[P](E_j - E_k) \in \mathbb{C}\,, \qquad \mathcal{D}[\rho]_{jk} = \mathcal{F}_{jk}\rho_{jk} \qquad (5.22)$$

We denote $A_{jk} = \langle E_j|\, A\, |E_j\rangle$ the matrix elements of an operator in the eigenbasis of H_T.

The only requirement on the dephasing channel is that the Fourier coefficients are bounded $\max_{j>0} |\mathcal{F}_{0j}| < \delta$, which imposes that the coherences between the ground state and any other eigenstate are suppressed by a factor smaller than δ. In the main text, we relate this factor to the dephasing time and to ground state gap of the target Hamiltonian.

The expectation value of circuit Eq. 5.17 is

$$\mathbb{E}[\mathrm{VHT}] = \mathrm{Tr}\left\{U_\leftarrow\ \mathcal{D}\Big[\mathrm{Ctrl}O\ \mathcal{D}[U_\rightarrow(|+\rangle\langle+|\otimes|\psi_0\rangle\langle\psi_0|)U_\rightarrow^\dagger]\mathrm{Ctrl}O^\dagger\Big]\right.$$
$$\left.U_\leftarrow^\dagger\ (|\psi_0\rangle\langle\psi_0|\otimes 2\,|0\rangle\langle1|)\right\} =$$

$$= \mathrm{Tr}\left\{\underbrace{U_\leftarrow^\dagger\,|\psi_0\rangle\langle\psi_0|\,U_\leftarrow}_{\sigma}\ \mathcal{D}\Big[O\ \mathcal{D}[\ \underbrace{U_\rightarrow\,|\psi_0\rangle\langle\psi_0|\,U_\rightarrow^\dagger}_{\rho}\ ]\Big]\right\} =$$

$$= \sum_{jk} \mathcal{F}_{jk}\ \mathrm{Tr}\left\{U_\leftarrow^\dagger\,|\psi_0\rangle\langle\psi_0|\,U_\leftarrow\cdot\right.$$
$$\left.\cdot\mathcal{D}\Big[O\,|E_j\rangle\underbrace{\langle E_j|\,U_\rightarrow\,|\psi_0\rangle\langle\psi_0|\,U_\rightarrow^\dagger\,|E_k\rangle}_{\rho_{jk}}\,\langle E_k|\Big]\right\} =$$

$$= \sum_{jkl} \mathcal{F}_{jk}\mathcal{F}_{lk}\sigma_{kl}\,O_{lj}\,\rho_{jk} \qquad (5.23)$$

where we expand the dephasing channels, and we define the density matrices ρ and σ, corresponding respectively to the pure states

$$U_\rightarrow \left| \psi_0 \right\rangle = \sqrt{1-\epsilon} \left| E_0 \right\rangle + \sqrt{\epsilon} \sum_{j>0} \alpha_j \left| E_j \right\rangle , \quad \sum_{j>0} |\alpha_j|^2 = 1; \qquad (5.24)$$

$$\left\langle \psi_0 \right| U_\leftarrow = \sqrt{1-\epsilon} \left\langle E_0 \right| + \sqrt{\epsilon} \sum_{j>0} \beta_j^* \left\langle E_j \right| , \quad \sum_{j>0} |\beta_j|^2 = 1. \qquad (5.25)$$

We can then absorb the dephasing coefficients into $\tilde{\rho}_{jk} = \mathcal{F}_{jk}\rho_{jk}$ and $\tilde{\sigma}_{kl} = \mathcal{F}_{lk}\sigma_{kl} = \mathcal{F}_{lk}^*\sigma_{kl}$, simplifying

$$\mathbb{E}[\text{VHT}] = \text{Tr}[\tilde{\rho}\tilde{\sigma}O], \quad \mathbb{E}[\text{Echo}] = \text{Tr}[\tilde{\rho}\tilde{\sigma}], \quad \langle O \rangle^{\text{EV}} = \frac{\text{Tr}[\tilde{\rho}\tilde{\sigma}O]}{\text{Tr}[\tilde{\rho}\tilde{\sigma}]}. \qquad (5.26)$$

Comparing this result to the standard purification estimator Eq. 5.2, we see that the ρ^2 is substituted by $\tilde{\rho}\tilde{\sigma}$. The explicit expression for this operator in the H_{T} eigenbasis is

$$\begin{aligned}
\tilde{\rho}\tilde{\sigma} =& \left[(1-\epsilon)^2 + \epsilon(1-\epsilon) \sum_{j>0} \alpha_j^* \beta_j \mathcal{F}_{j0}^* \mathcal{F}_{0j} \right] \left| E_0 \right\rangle \left\langle E_0 \right| \\
&+ \sqrt{\epsilon}\sqrt{1-\epsilon} \sum_{j>0} \left[(1-\epsilon)\alpha_j \mathcal{F}_{j0} + \epsilon \sum_{l>0} \alpha_j \alpha_l^* \beta_l \mathcal{F}_{jl} \mathcal{F}_{l0}^* \right] \left| E_j \right\rangle \left\langle E_0 \right| \\
&+ \sqrt{\epsilon}\sqrt{1-\epsilon} \sum_{j>0} \left[(1-\epsilon)\beta_j^* \mathcal{F}_{j0} + \epsilon \sum_{l>0} \beta_j^* \alpha_l^* \beta_l \mathcal{F}_{jl} \mathcal{F}_{l0}^* \right] \left| E_0 \right\rangle \left\langle E_j \right| \\
&+ \epsilon \sum_{j,k>0} \left[(1-\epsilon)\alpha_j \beta_k^* \mathcal{F}_{j0} \mathcal{F}_{0k}^* + \epsilon \sum_{l>0} \alpha_j \alpha_l^* \beta_l \beta_k^* \mathcal{F}_{jl} \mathcal{F}_{lk}^* \right] \left| E_j \right\rangle \left\langle E_k \right| \\
=& [\tilde{\rho}\tilde{\sigma}]_{00} \left| E_0 \right\rangle \left\langle E_0 \right| + \sum_{j>0} [\tilde{\rho}\tilde{\sigma}]_{j0} \left| E_j \right\rangle \left\langle E_0 \right| + \\
&+ \sum_{j>0} [\tilde{\rho}\tilde{\sigma}]_{0j} \left| E_0 \right\rangle \left\langle E_j \right| + \sum_{j,k>0} [\tilde{\rho}\tilde{\sigma}]_{jk} \left| E_j \right\rangle \left\langle E_k \right| . \qquad (5.27)
\end{aligned}$$

We proceed by bounding the error of $\langle O \rangle^{\text{EV}}$ with respect to the target $\left\langle E_0 \right| O \left| E_0 \right\rangle$,

$$\begin{aligned}
\text{error} :=& \left| \frac{\text{Tr}[O\tilde{\rho}\tilde{\sigma}]}{\text{Tr}[\tilde{\rho}\tilde{\sigma}]} - \left\langle E_0 \right| O \left| E_0 \right\rangle \right| \\
=& \left| \text{Tr}[O\tilde{\rho}\tilde{\sigma}] - O_{00} \text{Tr}[\tilde{\rho}\tilde{\sigma}] \right| \cdot \left| \text{Tr}[\tilde{\rho}\tilde{\sigma}] \right|^{-1} . \qquad (5.28)
\end{aligned}$$

In the H_T eigenbasis, the relevant terms read

$$\text{Tr}[O\tilde{\rho}\tilde{\sigma}] = [\tilde{\rho}\tilde{\sigma}]_{00}O_{00} + \sum_{j>0}[\tilde{\rho}\tilde{\sigma}]_{0j}O_{j0} + \sum_{k>0}[\tilde{\rho}\tilde{\sigma}]_{k0}O_{0k} + \sum_{j,k>0}[\tilde{\rho}\tilde{\sigma}]_{jk}O_{kj},$$
(5.29)

$$\text{Tr}[\tilde{\rho}\tilde{\sigma}] = [\tilde{\rho}\tilde{\sigma}]_{00} + \sum_{j>0}[\tilde{\rho}\tilde{\sigma}]_{jj}.$$
(5.30)

We focus first on bounding the first factor on the right-hand side of Eq. 5.28,

$$\left|\text{Tr}[O\tilde{\rho}\tilde{\sigma}] - O_{00}\text{Tr}[\tilde{\rho}\tilde{\sigma}]\right| = \left|\text{Tr}[(O - O_{00})\tilde{\rho}\tilde{\sigma}]\right|.$$
(5.31)

We separate this expression through the triangle inequality,

$$\left|\text{Tr}[O\tilde{\rho}\tilde{\sigma}] - O_{00}\text{Tr}[\tilde{\rho}\tilde{\sigma}]\right| =$$
$$= \left|\sum_{j>0}[\tilde{\rho}\tilde{\sigma}]_{0j}O_{j0} + \sum_{k>0}[\tilde{\rho}\tilde{\sigma}]_{k0}O_{0k} + \sum_{j,k>0}[\tilde{\rho}\tilde{\sigma}]_{jk}O_{kj} - \sum_{j>0}[\tilde{\rho}\tilde{\sigma}]_{jj}O_{00}\right|$$
$$\leq \|O\|\left(\left|\sum_{j>0}[\tilde{\rho}\tilde{\sigma}]_{0j}\right| + \left|\sum_{j>0}[\tilde{\rho}\tilde{\sigma}]_{j0}\right|\right) + \left|\sum_{j,k>0}[\tilde{\rho}\tilde{\sigma}]_{jk}O_{kj} - \sum_{j>0}[\tilde{\rho}\tilde{\sigma}]_{jj}O_{00}\right|,$$
(5.32)

where $\|\cdot\|$ is the operator norm. To bound the first term, we apply again the triangle inequality,

$$\left|\sum_{j>0}[\tilde{\rho}\tilde{\sigma}]_{0j}\right| \leq \left|\sum_{j>0}\tilde{\rho}_{00}\tilde{\sigma}_{0j}\right| + \left|\sum_{j,k>0}\tilde{\rho}_{0k}\tilde{\sigma}_{kj}\right|$$
(5.33)

and bound both resulting terms through Cauchy-Schwartz inequalities,

$$\left|\sum_{j>0}\tilde{\rho}_{00}\tilde{\sigma}_{0j}\right| \leq (1-\epsilon)^{3/2}\,\epsilon^{1/2}\left|\sum_j \mathcal{F}_{j0}\beta_j^*\right|$$
$$\leq (1-\epsilon)^{3/2}\,\epsilon^{1/2}\,|\vec{\beta}|\max_{k>0}|\mathcal{F}_{0k}| = (1-\epsilon)^{3/2}\,\epsilon^{1/2}\delta; \quad (5.34)$$

5

$$\left| \sum_{j,k>0} \tilde{\rho}_{0k}\, \tilde{\sigma}_{kj} \right| = (1-\epsilon)^{1/2}\, \epsilon^{3/2} \left| \sum_{j,k>0} \mathcal{F}_{0k}\alpha_k^*\beta_k \mathcal{F}_{jk}\beta_j^* \right|$$

$$\leq (1-\epsilon)^{1/2}\, \epsilon^{3/2}\, |\vec{\beta}| \sum_{j>0} \left| \sum_{k>0} \mathcal{F}_{jk}\alpha_k^*\beta_k \mathcal{F}_{0k} \right|$$

$$\leq (1-\epsilon)^{1/2}\, \epsilon^{3/2}\, |\vec{\beta}|^2\, |\vec{\alpha}| \max_{j,k>0} |\mathcal{F}_{jk}| \max_{k>0} |\mathcal{F}_{0k}|$$

$$\leq (1-\epsilon)^{1/2}\, \epsilon^{3/2}\delta \tag{5.35}$$

where we note that $\vec{\alpha}$ and $\vec{\beta}$ are normalized by defintion, $\mathcal{F}_{jk} \leq 1$ and $\max_{k>0} |\mathcal{F}_{0k}| = \delta$. The same bound applies to the second term in the parentheses in Eq. 5.32, $\left| \sum_{j>0} [\tilde{\rho}\tilde{\sigma}]_{j0} \right|$. The last term of Eq. 5.32 can be rewritten as

$$\sum_{j,k>0} [\tilde{\rho}\tilde{\sigma}]_{jk} O_{kj} - \sum_{j>0} [\tilde{\rho}\tilde{\sigma}]_{jj} O_{00} = \mathrm{Tr}[\Pi_> \tilde{\rho}\tilde{\sigma}\, \Pi_> (O - O_{00}\mathbb{1})] \tag{5.36}$$

where $\Pi_> = \mathbb{1} - |E_0\rangle\langle E_0|$ is the projector on the subspace orthogonal to $|E_0\rangle$. We can then use the Von Neumann inequality to bound

$$\left| \mathrm{Tr}[\Pi_> \tilde{\rho}\tilde{\sigma}\, \Pi_> (O - O_{00}\mathbb{1})] \right| \leq \|O - O_{00}\mathbb{1}\| \cdot \|\Pi_> \tilde{\rho}\tilde{\sigma}\, \Pi_>\|_1 \tag{5.37}$$

where $\|A\|_1 = \mathrm{Tr}\sqrt{A^\dagger A}$ is the trace norm.

Now, by virtue of the triangle inequality, and the fact that $\||u\rangle\langle v|\|_1 = \|u\|\|v\|$, for any vectors $|u\rangle, |v\rangle$, we have

$$\|\Pi_> \tilde{\rho}\tilde{\sigma}\Pi_>\|_1 \leq \epsilon(1-\epsilon) \left\| \sum_{j>0} \alpha_j \mathcal{F}_{j0} |E_j\rangle \right\| \left\| \sum_{k>0} \beta_k \mathcal{F}_{0k} |E_k\rangle \right\| +$$

$$+ \epsilon^2 \sum_{l>0} |\alpha_l||\beta_l| \left\| \sum_{j>0} \alpha_j \mathcal{F}_{jl} |E_j\rangle \right\| \left\| \sum_{k>0} \beta_k \mathcal{F}_{lk} |E_k\rangle \right\|$$

$$\leq \epsilon(1-\epsilon)\delta^2 + \epsilon^2 \left(\max_{j,k>0} |\mathcal{F}_{jk}| \right)^2 \sum_{l>0} |\alpha_l||\beta_l|$$

$$\leq \epsilon(1-\epsilon)\delta^2 + \epsilon^2 \tag{5.38}$$

where we used Cauchy-Schwartz in the last line.

Combining the bounds from Eqs. (5.34), (5.35) and (5.38), and using

$\|O - O_{00}\mathbb{1}\| \leq 2\|O\|$, we get

$$\left|\mathrm{Tr}[O\tilde{\rho}\tilde{\sigma}] - O_{00}\,\mathrm{Tr}[\tilde{\rho}\tilde{\sigma}]\right| \leq$$
$$\leq 2\|O\|\left[(1-\epsilon)^{3/2}\,\epsilon^{1/2}\delta + 3(1-\epsilon)^{1/2}\,\epsilon^{3/2}\delta + \epsilon(1-\epsilon)\delta^2 + \epsilon^2\right]. \quad (5.39)$$

Next, to bound the factor $|\mathrm{Tr}[\tilde{\rho}\tilde{\sigma}]|^{-1}$ in Eq. 5.28, we apply the reverse triangle inequality to $|\mathrm{Tr}[\tilde{\rho}\tilde{\sigma}]|$:

$$\left|\mathrm{Tr}[\tilde{\rho}\tilde{\sigma}]\right| = \left|(1-\epsilon)^2 + 2\epsilon(1-\epsilon)\,\mathrm{Re}\left(\sum_{j>0}\alpha_j\beta_j^*\mathcal{F}_{j0}^2\right) + \epsilon^2\sum_{j,l>0}\alpha_j\alpha_l^*\beta_l\beta_j^*\mathcal{F}_{jl}^2\right|$$
$$\geq \left|(1-\epsilon)^2 - 2\epsilon(1-\epsilon) - \epsilon^2\right| \quad (5.40)$$

where we used that $\delta \leq 1$. For $\epsilon < \sqrt{3/2} - 1$, the argument in Eq. 5.40 is strictly positive, so we can remove the absolute value signs. The dominant terms in the error Eq. 5.28 are then

$$\mathrm{error} \sim \|O\|(\epsilon^{1/2}\delta + \epsilon^2). \quad (5.41)$$

We can verify that for $\delta \to 0$ we recover the error scaling with ϵ^2, as expected from perfect dephasing. To achieve the same scaling, it is in fact sufficient to choose $\delta = \epsilon^{3/2}$.

5.C Dephasing time for a smooth probability distribution

In this section we motivate the choice of the rescaled bump function in Eq. 5.13 for the distribution $P(\tau)$ used to implement dephasing by random-time evolution (Eq. 5.8). We recall that we require $P : [0, T_\mathrm{d}] \to \mathbb{R}_+$ to have support on $[0, T_\mathrm{d}]$. This ensures we only need to evolve for positive times and the maximal dephasing time is T_d. The performance for the dephasing operation on the ground state is measured by $\delta = \max_{\Delta > \Delta_T} |\mathcal{F}[P](\Delta)|$, which is essentially a bound on the decay of the Fourier transformation of P. As τ and Δ are conjugate dimensionful variables, we can equivalently study

$$\delta = \max_{\omega > \Delta_T T_\mathrm{d}} |\mathcal{F}[\tilde{P}](\omega)| \quad \text{for} \quad \tilde{P} : [0,1] \to \mathbb{R}_+, \quad (5.42)$$

where $\tilde{P}(\tau\Delta_T) = P(\tau)$.

To obtain the best possible asymptotic decay of the Fourier transform of a function $\mathcal{F}[f]$, we should choose f to be smooth. In fact, requiring the

Fourier transform of f to decay as $\mathcal{F}[f](k) \lesssim |k|^{-(r+1+\epsilon)}$ (for any choice of $\epsilon > 0$) implies that

$$\exists L_r > 0 : \forall x \quad \left| \frac{\mathrm{d}^r f(x)}{\mathrm{d}x^r} \right| = \left| \int \mathrm{d}k\, e^{ikt} \underbrace{k^r \mathcal{F}[f](k)}_{\lesssim |k|^{-(1+\epsilon)}} \right| < L_r, \qquad (5.43)$$

because $|k|^{-(1+\epsilon)}$ is absolutely integrable away from 0 and $\mathcal{F}[f](k)$ is bounded. This implies that f and all its derivatives up to order $r-1$ are Lipschitz continuous. Thus, to achieve a Fourier transform decaying faster than any polynomial $\mathcal{F}[f](k) = o(1/\mathrm{poly}(k))$, we have to choose $f(x) \in C_\infty$ a smooth function.

One smooth function with compact support is the bump function

$$f(x) = \begin{cases} e^{-(1-x^2)^{-1}} & \text{if} \quad -1 < x < 1, \\ 0 & \text{otherwise,} \end{cases} \qquad (5.44)$$

we define its norm $\mathcal{N} := \int_{-1}^{1} f(x)\, \mathrm{d}x \approx 2.25$. Based on this function, we define the probability distribution

$$\begin{aligned} P_{T_\mathrm{d}}(\tau) &= \frac{2}{T_\mathrm{d}\mathcal{N}} f\left(2\frac{\tau}{T_\mathrm{d}} - 1 \right) \\ &= \begin{cases} \frac{2}{T_\mathrm{d}\mathcal{N}} \exp\left(\left[4(\frac{\tau}{T_\mathrm{d}} - 1)\frac{\tau}{T_\mathrm{d}} \right]^{-1} \right) & \text{if} \quad 0 < \tau < T_\mathrm{d}, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \qquad (5.45)$$

which is normalized, smooth, and has support on $[0, T_\mathrm{d}]$. The Fourier transform of this function can be estimated through the saddle point approximation; we build on the results of Ref. 188 which provide a bound the Fourier transform $\mathcal{F}[\mathcal{N}f]$ of the normalized $f(x)$:

$$\mathcal{F}[\mathcal{N}f](k) \approx 2\, \mathrm{Re}\left[\sqrt{\frac{-i\pi}{\sqrt{2i}}} e^{ik - \frac{1}{4} - i\sqrt{k}} \right] k^{-\frac{3}{4}} e^{-\sqrt{k}}. \qquad (5.46)$$

We construct a monotonic envelope for this oscillating function by substituting the real part for an absolute value, and we perform a change of variables obtaining the bound

$$\delta = \max_{\Delta > \Delta_T} |\mathcal{F}[P_{T_\mathrm{d}}](\Delta)| < \sqrt{\frac{8\pi}{\sqrt{e}}} (T_\mathrm{d}\Delta_T)^{-3/4} e^{-\sqrt{T_\mathrm{d}\Delta_T/2}}. \qquad (5.47)$$

The validity of this bound is also verified numerically. This translates to a statement on the dephasing time T_{d} required to achieve a target dephasing performance δ_{target} for a given gap Δ_T between the ground state and the first excited state of H_T. Note that the inverse is defined in terms of the principal branch W_0 of the Lambert W function. We obtain

$$T_{\mathrm{d}} = \frac{9}{2\,\Delta_T} W_0 \left[\frac{2\,\sqrt{2}\,\pi^{1/3}}{3\,e^{1/6}\,\delta_{\mathrm{target}}^{2/3}} \right] \lesssim O\left(\log^2(\delta_{\mathrm{target}}^{-1}) \Delta_T^{-1} \right). \qquad (5.48)$$

Thus, T_{d} grows linearly with Δ_T and poly-logarithmically with $\delta_{\mathrm{target}}^{-1}$.

5

A hybrid quantum algorithm to detect conical intersections

6.1 Introduction

Conical intersections (CI) are degeneracy points in the Born-Oppenheimer molecular structure Hamiltonians, where two potential energy surfaces cross. Similar to Dirac cones in graphene [197], these intersections are protected by symmetries of the Hamiltonian which guarantee that any loop in parameter space around a conical intersection has a quantized Berry phase [198]. CIs play an important role in photochemistry [199, 200], as they mediate reactions such as photoisomerization and non-radiative relaxation, which are key steps in processes such as vision [201] and photosynthesis [202]. Therefore, detecting the presence and resolving the properties of CIs is important for computing reaction and branching rates in photochemical reactions [203, 204]. Nevertheless, the study of such processes requires electronic structure methods capable of accurately modelling both the shape and the relative energies of the two intersecting potential energy surfaces, a requirement that poses challenges for the current available methods [205]. Given the need to develop novel methods for identifying and characterizing CIs, quantum computers present themselves as a highly promising option for this task.

Quantum computing has long been driven by the desire to simulate interacting physical systems, such as molecules, as a novel means of investigating their properties [2, 70]. This is typically achieved by preparing eigenstates of molecular Hamiltonians in quantum devices which can natively store and process quantum states. This task would otherwise require an exponentially-scaling classical memory. Recently, with the first noisy and intermediate-scale quantum devices (NISQ) [18] being built, it became increasingly important to research tailored and robust algorithms that minimize the quantum device requirements [18]. Variational quantum algorithms (VQA), such as the variational quantum eigensolver [19, 82] (VQE) and its variations, caught the spotlight in this context, as they allow to prepare and measure quantum states with circuits of relatively low depth. The key feature of VQAs is the repeated execution of short parameterized quantum circuits on the quantum device, from which measurement results are sampled. These results are used to estimate a cost function, which is then minimized by varying the parameters defining the gates of the quantum circuit. Due to the noise introduced by sampling, a relatively large number of circuit runs and measurements are typically needed to estimate the cost function accurately. In chemistry, where VQEs are often proposed as a method to resolve ground state energies to high accuracy, the number of required samples to achieve such accuracy can become prohibitively large [54]. Furthermore, the convergence of the cost function to an optimum is typically only suggested heuristically, and it is proven to be problematic in some cases that lack such heuristic structure [206]. Therefore, it is compelling to suggest VQAs that can access quantities that are less reliant on the precision of both the optimization process and the measurement procedure.

A promising target for VQAs is the computation of the Berry phase $\Pi_{\mathcal{C}}$, which can be used to resolve the existence of CIs. More specifically, Π_C is defined as the geometric phase acquired by an eigenstate of a parameterized Hamiltonian over a closed adiabatic path $\mathcal{C}$ in parameter space [198]. Most importantly, it is known that in the presence of certain symmetries, the Berry phase will be quantized to values 0 or π. This quantization is exactly what makes the Berry phase an attractive target for a VQAs, as it implies the final result of the computation need only be accurate to error $< \frac{\pi}{2}$. Quantum algorithms to compute Berry phases have been already proposed, both variational [207, 208] and Hamiltonian-evolution based [209]. Moreover, the long-known effects of Berry phase on nuclear dynamics around a conical intersection [210–212] have been explored recently in analog quantum simulation experiments [213–215]. Nevertheless, previous proposals did not attempt to detect CIs in realistic

176

quantum chemistry problems with an efficient algorithm that can be run on NISQ devices.

In this chapter, we propose a hybrid quantum algorithm to compute the quantized Berry phase for ground states of a family of parameterized real Hamiltonians. We focus on the specific application to molecular Hamiltonians, where we can identify a conical intersection by measuring the Berry phase along a loop in atomic coordinates space. We first review the definition of CIs and Berry phases in Sec. 6.2. Then, we present all the ingredients of the proposed algorithm in Sec. 6.3, which is similar to a VQA in spirit, but it does not require full optimization; rather, the variational parameters are updated by a single Newton-Raphson step for each molecular geometry along a discretization of the loop. In Sec. 6.4, we prove a convergence guarantee for the algorithm under certain assumptions on the ansatz, by providing sufficient condition bounds on the total number of steps and the acceptable sampling noise. Finally, we adapt our algorithm to a specific ansatz in Sec. 6.5 and we benchmark it on a model of the formaldimine molecule $H_2C=NH$ in Sec. 6.6. Section 6.7 presents our conclusions, a discussion of potential application cases for our algorithm and an outlook on possible enhancements.

The core code developed for the numerical benchmarks, which provides a flexible implementation of an orbital-optimized variational quantum ansatz, is made available in a GitHub repository [216].

6.2 Background

6.2.1 Conical intersections

Let us consider a molecular electronic structure Hamiltonian $H(\mathbf{R})$ parameterized by the nuclear geometry $\mathbf{R}$ in some configuration space $\mathcal{R}$. A conical intersection is a point $\mathbf{R}^\times \in \mathcal{R}$ where two potential energy surfaces become degenerate, leading to non-perturbatively large non-adiabatic couplings, and thus a breakdown of the Born-Oppenheimer approximation [217, 218]. Conical intersections extend to a manifold of dimension $\dim[\mathcal{R}] - 2$, and lead to the two potential energy surfaces taking the form of cones in the remaining two directions $\hat{x}$, $\hat{z}$. These two potential energy surfaces can be described as eigenstates of the effective Hamiltonian

$$H^{\mathrm{eff}}(\mathbf{R}) = h_x[\mathbf{R} - \mathbf{R}^\times]_x \sigma_x + h_z[\mathbf{R} - \mathbf{R}^\times]_z \sigma_z. \qquad (6.1)$$

The Pauli terms σ_x and σ_z form a complete basis for two-dimensional real symmetric matrices. Thus, a single conical intersection cannot be

lifted by any real-valued and continuous perturbation of $H^{\mathrm{eff}}(\mathbf{R})$; such a perturbation would only shift the value of $\mathbf{R}^{\times}$. Moreover, the presence of such an effective Hamiltonian implies that in any direction other than R_x or R_z the energies must be degenerate.

Simulations involving conical intersections are challenging due to the degeneracy of the two states involved, as the character of both states needs to be considered. Active space methods are often used for organic molecules, which involve selecting the chemical bonds that are formed or broken in the reaction pathway as well as the most significant spectator or correlating orbitals [219]. The situation is more complicated when transition metals are involved because the close energetic spacing of d-orbitals typically requires including all five d-orbitals of such a metal into the active space. An additional complication arises if the two crossing states correspond to different atomic configurations: a situation that is not uncommon for the early or late d- (or f-) metals, for which configurations with a different d- (f-) population are energetically close. In such cases, one may need to either work with non-orthogonal orbitals [220] or add an additional d-shell to the active space [221] to qualitatively describe the nature of both states. For cases of practical interest in which one wants to characterize and simulate the internal conversion processes in a complex photo-excited system, the presence of transition metals may easily lead to large active space requirements. These can not be met by classical algorithms and would be highly challenging for quantum algorithms as well.

One way to reduce the complexity of the problem is to first focus on the presence or absence of conical intersections that connect the ground and excited states. The measurement of the Berry phase in chemical systems allows this: without explicitly computing the excited state surface and non-adiabatic couplings, it should be possible to detect whether a loop in the nuclear coordinate space encloses a conical intersection or not. In this manner, one may alleviate the requirements for the active space selection and orbital optimization and quickly establish the region in the potential energy surface that contains an intersection with another surface and needs to be scrutinized further [222]. Information about the location of conical intersections is of interest also for ground-state dynamics; the CIs and the Berry phase they induce influence the propagation of nuclear wave packets on the adiabatic ground state surface and thereby affect the branching rates and efficiency of reactions or isomerizations [223]. For both types of applications, precise study of dynamics on ground state surfaces as well as characterizing the efficiency of radiationless decay, it is of interest to explore the possibilities offered by quantum algorithms.

6.2.2 Berry phases in real Hamiltonians

The Berry phase $\Pi_{\mathcal{C}}$ is the geometric phase acquired by some eigenstate $|\Phi(\mathbf{R})\rangle$ of a system with parameterized Hamiltonian $H(\mathbf{R})$ as it is adiabatically transported around a closed loop in parameter space $\mathcal{C} \subset \mathcal{R}$. $\Pi_{\mathcal{C}}$ can be defined as the closed line integral of the Berry connection along the loop $\mathcal{C}$

$$\Pi_{\mathcal{C}} = -i \oint_{\mathcal{C}} d\mathbf{R} \cdot \langle \Phi(\mathbf{R})| \nabla_{\mathbf{R}} |\Phi(\mathbf{R})\rangle . \tag{6.2}$$

The integrand must be imaginary, as

$$\nabla_{\mathbf{R}} \langle \Phi(\mathbf{R})|\Phi(\mathbf{R})\rangle = 2\,\mathrm{Re}[\langle \Phi(\mathbf{R})| \nabla_{\mathbf{R}} |\Phi(\mathbf{R})\rangle] = 0, \tag{6.3}$$

thus $\Pi_{\mathcal{C}}$ is real. In this work, we assume $|\Phi(\mathbf{R})\rangle$ is the ground state of $H(\mathbf{R})$.

We need to parameterize the loop $\mathcal{C} = \{\mathbf{R}(t), t \in [0,1]\}$ in order to evaluate the integral. Moreover, it is possible to multiply the ground state $|\Phi(\mathbf{R})\rangle$ by a t-dependent phase, resulting in a $U(1)$-gauge a transformation which leaves all physical quantities invariant. We take

$$|\Psi(t)\rangle = e^{i\Theta(t)}|\Phi(\mathbf{R}(t))\rangle, \tag{6.4}$$

which allows us to rewrite the Berry phase as

$$\Pi_{\mathcal{C}} = -i \int_0^1 dt \langle \Psi(t)|\partial_t|\Psi(t)\rangle + \int_0^1 dt\, \partial_t \Theta(t). \tag{6.5}$$

If there is a representation for which each Hamiltonian $H(\mathbf{R})$ is real, it is possible to choose eigenstates that have all real components. In this case, we can choose $\Theta(t)$ such that $|\Psi(t)\rangle$ has real expansion, which implies the first integrand is real; as $\partial_t \langle \Psi(t)|\Psi(t)\rangle = 0$ this must also to be imaginary, therefore $\langle \Psi(t)|\partial_t|\Psi(t)\rangle = 0$. Under this choice, we can evaluate the Berry phase as a boundary term

$$\Pi_{\mathcal{C}} = \int_0^1 dt\, \partial_t \Theta(t) = \Theta(1) - \Theta(0) = \arg\left[\langle \Psi(0)|\Psi(1)\rangle\right], \tag{6.6}$$

where the last equality is obtained using the definition in Eq. (6.4). Furthermore, $|\Psi(1)\rangle$ and $|\Psi(0)\rangle$ are real by construction, which implies $\Pi_{\mathcal{C}}$ can only take two values (modulo 2π): 0 or π.

The quantization of $\Pi_{\mathcal{C}}$ implies that it is invariant for topological deformations of $\mathcal{C}$. If $\mathcal{C}$ can be contracted to a point, then $\Pi_{\mathcal{C}} = 0$. A non-trivial

6

$\Pi_{\mathcal{C}} = \pi$ can only occur when $\mathcal{C}$ encircles a degeneracy. One can check with the effective Hamiltonian Eq. (6.1) that any loop encircling $\mathbf{R}^{\times}$ has a Berry phase of π. This extends by continuity to any region of $\mathcal{R}$ around the CI, as long as $\mathcal{C}$ does not enclose a second degeneracy point. Thus, we have a one-to-one correspondence between CIs and the nontrivial Berry phase.

6.2.3 Measuring Berry phase with a variational wavefunction

There have been various proposals in the literature for computing Berry phases using a gate-based quantum device [207, 209]. In this work, we propose to use a variational algorithm to track $|\Psi(t)\rangle$ when parallel-transported around the loop $\mathcal{C}$, in the spirit of the variational adiabatic method described in Ref. [54, 224]. We approximate

$$|\Psi(t)\rangle \approx |\psi(\boldsymbol{\theta}_t^*)\rangle := U(\boldsymbol{\theta}_t^*) |\psi_0\rangle, \tag{6.7}$$

where $|\psi(\boldsymbol{\theta})\rangle$ is a variational ansatz state, and $\boldsymbol{\theta}_t^*$ continuously tracks a local minimum $[\nabla_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta}_t^*) = 0]$ of the variational energy

$$E(t, \boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta})| H(\mathbf{R}(t)) |\psi(\boldsymbol{\theta})\rangle. \tag{6.8}$$

The angle $\boldsymbol{\theta}_t^*$ is well-defined as long as the Hessian $\nabla_{\boldsymbol{\theta}}^2 E(t, \boldsymbol{\theta})$ remains positive definite in a neighbourhood of $\boldsymbol{\theta}_t^*$ for all t, ensuring the $\boldsymbol{\theta}_t^*$ is continuous in t and non-degenerate. Although our treatment naturally extends to any variational ansatz that continuously parametrizes normalized states $|\psi(\boldsymbol{\theta})\rangle$ (including classical ansätze like e.g. matrix-product states), we assume the operator $U(\boldsymbol{\theta})$ is implemented by a parameterized quantum circuit (PQC) acting on an initial state $|\psi_0\rangle$; this implies that information about the state needs to be extracted from a quantum device through sampling.

6.3 Methods

In this section, we detail all the ingredients needed to implement our hybrid algorithm to resolve quantized Berry phases with a variational quantum ansatz. Initially, in Sec. 6.3.1, we discuss how selecting an ansatz that preserves the Hamiltonian's symmetries establishes a natural gauge, leading to the reduction of the Berry phase integral to the boundary term Eq. (6.6). In Sec. 6.3.2, we introduce our parameter update approach,

which employs single Newton-Raphson steps to trace the variational state along a discretization of the loop $\mathcal{C}$. In Sec. 6.3.3 we explain how to employ a basic regularisation technique to handle the potential non-convexity of the cost function and in 6.3.4 we explain how to measure the final overlap using an ancilla-free Hadamard test. Finally, in Sec. 6.3.5, we provide a full overview of the algorithm.

6.3.1 Fixing the gauge with a real ansatz

As discussed in Sec. 6.2, the quantization of the Berry phase is granted by the symmetries of the Hamiltonian family $H(\mathbf{R})$, which ensure the existence of a basis for which each $H(\mathbf{R})$ has a real representation. When it comes to electronic structure Hamiltonians, it is always possible to find a real representation for time-reversal symmetric Hamiltonians with integer total spin [225]. Moreover, real noninteger-spin Hamiltonians are also found throughout nonrelativistic quantum chemistry.

As our variational ansatz state [in Eq. (6.7)] is defined by a family of unitary operators, it inherits a natural gauge from $U(\boldsymbol{\theta})$. In particular, if $U(\boldsymbol{\theta})$ is written as a product of real rotations in the basis in which $H(\mathbf{R})$ is real, then we force $|\psi(\boldsymbol{\theta})\rangle$ to have real components as well, which fixes a global $U(1)$ phase. This can be obtained by constructing the PQC with a sequence of parameterized unitaries such as

$$U_j(\theta_j) = e^{A_j \theta_j} \tag{6.9}$$

generated by antisymmetric operators A_j that are real in the chosen representation. (We choose dimensional units such that $\|A_j\| = 1$ without loss of generality, see Appendix 6.A). Examples from electronic structure include real fermionic (de-)excitations, such as unitary singles ($A_{pq} = \hat{a}_p^\dagger \hat{a}_q - \hat{a}_q^\dagger \hat{a}_p$) and doubles ($A_{pqrs} = \hat{a}_p^\dagger \hat{a}_q \hat{a}_r^\dagger \hat{a}_s - \hat{a}_q^\dagger \hat{a}_p \hat{a}_s^\dagger \hat{a}_r$). Many PQC ansätze commonly proposed for quantum chemistry, such as unitary coupled cluster (UCC) [19, 226, 227] and quantum-number preserving gate fabrics (NPF) [57], are composed from these elementary rotations. Formally, our ansatz state can then be defined as

$$|\psi(\boldsymbol{\theta})\rangle = \prod_{j=1}^{n_p} U_{n_p-j}(\theta_{n_p-j}) |\psi_0\rangle, \tag{6.10}$$

where U_k are the aforementioned parameterized rotations applied in circuit-composition order.

In this case, the Berry phase can be estimated by

$$\Pi_{\mathcal{C}} = \arg\left[\,\langle\psi(\boldsymbol{\theta}_{t=0}^*)|\psi(\boldsymbol{\theta}_{t=1}^*)\rangle\,\right], \tag{6.11}$$

which corresponds to the boundary term in Eq. (6.6). This implies that, in the case of a non-trivial Berry phase, the path traced by $\boldsymbol{\theta}_t^*$ will not close up on itself (i.e. $\boldsymbol{\theta}_1^* \neq \boldsymbol{\theta}_0^*$), highlighting an important difference between the ansatz parameters $\boldsymbol{\theta}$, which fix the gauge of $|\psi(\boldsymbol{\theta})\rangle$, and the Hamiltonian parameters $\mathbf{R}$, which define a ground state $|\Phi(\mathbf{R})\rangle$ up to a $U(1)$ gauge freedom. However, the change in optimal parameters is insufficient to prove the existence of a nontrivial Berry phase; we must both successfully track the minimum $\boldsymbol{\theta}_t^*$ as t traces from 0 to 1, and estimate the final overlap $\langle\psi(\boldsymbol{\theta}_0^*)|\psi(\boldsymbol{\theta}_1^*)\rangle$. While the argument (sign) of the overlap will yield the Berry phase, its absolute value is a proxy of success as it certifies the initial and final states are physically equivalent.

6.3.2 Avoiding full optimization via Newton-Raphson steps

As mentioned above, the Berry phase $\Pi_{\mathcal{C}}$ is a discrete quantity, and therefore we only need to estimate it to accuracy $< \frac{\pi}{2}$. In Appendix 6.A, we show that this implies that we can accept an error on the estimate $\tilde{\boldsymbol{\theta}}_1$ of the final optimum $\boldsymbol{\theta}_1^*$ bounded in 1-norm by $\|\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^*\|_1 < 1$. Thus, we are not required to exactly track $|\psi(\boldsymbol{\theta}_t^*)\rangle$ and as a result, the variational energy Eq. (6.8) does not need to be fully re-optimized at every time-step t. Instead, it suffices to keep the estimate $\tilde{\boldsymbol{\theta}}_t$ of the optimal parameters within the basis of convergence of the true minimum $\boldsymbol{\theta}_t^*$.

To achieve this, we still need an initial optimum as an input, which is obtained by running one full optimization. If possible, the initial point is selected such that optimization is simplest. Then, we propose to use a single step of the Newton-Raphson algorithm at points $t \in \{\Delta t, 2\,\Delta t, \ldots, 1 - \Delta t, 1\}$.

The Newton-Raphson (NR) algorithm determines the update of the estimate of the minimum $\tilde{\boldsymbol{\theta}}$ through the gradient $\mathcal{G}(t, \boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta})$ and Hessian $\mathcal{H}(t, \boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}}^2 E(t, \boldsymbol{\theta})$ of the variational energy Eq. (6.8). The derivatives can be computed using either finite-difference methods or parameter-shift rules [228]; either method requires sampling the variational energy $E(t, \boldsymbol{\theta})$ at a number of different parameter points $\boldsymbol{\theta}$. Given the estimate $\tilde{\boldsymbol{\theta}}_t$ of the optimum at point t as an initial guess, the NR step with

6

cost $E(t + \Delta t, \boldsymbol{\theta})$ prescribes the update $\tilde{\boldsymbol{\theta}}_{t+\Delta t} = \tilde{\boldsymbol{\theta}}_t + d\boldsymbol{\theta}_{t,\Delta t}^{\text{NR}}$, with

$$d\boldsymbol{\theta}_{t,\Delta t}^{\text{NR}} = -\mathcal{H}^{-1}(t + \Delta t, \tilde{\boldsymbol{\theta}}_t)\boldsymbol{\mathcal{G}}(t + \Delta t, \tilde{\boldsymbol{\theta}}_t). \tag{6.12}$$

The Newton-Raphson method is well-known to have a finite-sized basin of quadratic convergence, as long as the cost function is strongly convex at the optimum,

$$m(\boldsymbol{\theta}^*) := \min_{\mathbf{v}} \frac{\mathbf{v}^T \mathcal{H}(\boldsymbol{\theta}^*)\mathbf{v}}{\mathbf{v}^T \mathbf{v}} > 0. \tag{6.13}$$

In other words, the lowest eigenvalue m of the Hessian of the cost function (which we call convexity) at the optimum $\boldsymbol{\theta}^*$ needs to be positive. Details on the convergence properties of NR are given in App. 6.D. In section 6.4, we show that this quadratic convergence is fast enough to keep track of the minimum with only a single step for each t-point.

6.3.3 Regularization and backtracking

To ensure that we successfully track the minimum of the cost function $\boldsymbol{\theta}_t^*$, the estimates $\tilde{\boldsymbol{\theta}}_t$ need to remain in the strongly convex region of optimization space. The existence of such strongly convex region is not always guaranteed since it depends on the ansatz. A common cause of failure of this requirement is exemplified for ansätze with (local) over-parametrization of the state manifold. In fact, if the ansatz has redundant parameters the Hessian of the cost function Eq. (6.8) will always be singular ($m = 0$). Then, arbitrarily small perturbations of the cost function can then cause $m < 0$. When this occurs, the inversion of the Hessian needed for the Newton-Raphson step is ill-defined.

The most direct approach to solve this issue is to select an ansatz with no degeneracies, facilitating a strongly convex cost function at its minima. Nevertheless, this is only possible for very simple problems, and even in this case, quasi-degeneracies can make the convergence region extremely small. Since this is a well-known problem, many alternative solutions have been proposed in the literature. In this subsection, we will explore and implement two of them – back-tracking and regularization.

Back-tracking — Small positive eigenvalues of the Hessian can cause the standard Newton-Raphson step to overshoot along the relative parameter eigenmodes. This effectively reduces the size of the neighborhood of the minimum θ^* in which quadratic convergence is granted (see Appendix 6.D). Since for positive convexity the direction provided by the Newton-Raphson step is guaranteed to be a descent direction, we can mitigate this overshoot by implementing line-search of the minimum on the segment defined by

6

the NR step. In the common variant of back-tracking, the Newton step is iteratively damped by a constant $\beta \in (0, 1)$. At each iteration, the cost function in the new point is measured, until the cost function is reduced enough (the detailed condition is given in Algorithm 6.1). While some additional evaluations of the cost function are needed, the (more expensive) gradient and Hessian are only calculated once. Due to the repeated evaluation, one needs to consider extending line-search methods to cost functions evaluated with sampling noise.

Regularization — For realistic ansätze and Hamiltonians, it is difficult to avoid (quasi-) redundancies in some regions of parameter space. In this case, the cost function might not be strongly convex around the minimum, or the convexity might be too small to ensure a sufficiently-large convergence region. Furthermore, even for an ideally-convex cost function, noisy evaluation on a quantum device might result in a distorted Hessian with non-positive eigenvalues. To mitigate this issue, we can use a regularization technique that penalizes the change in parameters along the quasi-redundant directions. We propose to use augmentation of the Hessian to regularize the NR step, obtaining a so-called quasi-Newton optimizer. Hessian augmentation is a common practice in quantum chemistry methods that feature orbital optimization, such as self-consistent field methods [219]. If the smallest eigenvalue of the Hessian λ_0 is smaller than a positive threshold convexity m_{thr}, we construct the augmented Hessian as follows

$$\mathcal{B} = \mathcal{H} + \nu \mathbb{1}, \tag{6.14}$$

where we add a constant $\nu > |\lambda_0|$. The augmented Hessian is then positive, and we can realize the NR update as

$$d\boldsymbol{\theta}^{\text{NR}} = -\beta \mathcal{B}^{-1} \boldsymbol{\mathcal{G}}. \tag{6.15}$$

Here, β is the damping constant from back-tracking line search and $\mathcal{G}$ is the gradient as in eq. (6.12). Regularization and back-tracking are typically used in tandem, as quadratic convergence is harder to guarantee when using regularization. The choice of ν is non-trivial: we want it to be large enough to suppress parameter changes along quasi-redundant directions, but we need to avoid exaggerating the damping along relevant directions. Common solutions include choosing $\nu = \rho|\lambda_0| + \mu$ with fixed positive constants ρ and μ, or using a trust-region method [229, 230] where the Newton step is constrained to lie within a ball of some radius h (such that $||d\boldsymbol{\theta}^{\text{NR}}||_2 \leq h$).

The augmented Hessian method does not require further evaluations of

the cost function, although the damping of the parameter updates might imply that more t-steps are needed to successfully resolve the Berry phase.

Algorithm 6.1: NR step subroutine with regularization and back-tracking

---

Input: Initial estimated parameters $\tilde{\boldsymbol{\theta}}_\tau$
Estimated Hessian $\tilde{\mathcal{H}}$
Estimated gradient $\tilde{\boldsymbol{\mathcal{G}}}$
Cost function $E(\theta)$ given by eq. (6.8)
Convexity requirement $m_{\text{thr}} > 0$
Positive constants μ, ρ, α, β
Output: Updated estimated parameters $\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}$.

1 $\lambda_0 \leftarrow$ lowest eigenvalue of $\tilde{\mathcal{H}}$;
  // Regularization
2 **if** $\lambda_0 < m$ **then**
3 $\quad$ $\mathcal{B} \leftarrow \tilde{\mathcal{H}} + (\rho|\lambda_0| + \mu)\mathbb{1}$;
4 $\quad$ $d\boldsymbol{\theta} \leftarrow \mathcal{B}^{-1}\tilde{\boldsymbol{\mathcal{G}}}$
5 **else**
6 $\quad$ $d\boldsymbol{\theta} \leftarrow \tilde{\mathcal{H}}^{-1}\tilde{\boldsymbol{\mathcal{G}}}$
  // Backtracking
7 **while** $E(\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}) > E(\tilde{\boldsymbol{\theta}}_\tau) + \alpha(\tilde{\boldsymbol{\mathcal{G}}} \cdot d\boldsymbol{\theta})$ **do**
8 $\quad$ $d\boldsymbol{\theta} \leftarrow \beta d\boldsymbol{\theta}$
9 **return** $\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}$

6.3.4 Measuring the final overlap

For a real ansatz, the overlap of the tracked states at $t = 0$ and $t = 1$ must be real and it can be rewritten as

$$\langle \psi(\tilde{\boldsymbol{\theta}}_0) | \psi(\tilde{\boldsymbol{\theta}}_1) \rangle = \text{Re} \left[\langle 0 | U^\dagger(\tilde{\boldsymbol{\theta}}_0) U(\tilde{\boldsymbol{\theta}}_1) | 0 \rangle \right]. \tag{6.16}$$

This quantity can readily be measured by a Hadamard test, implemented as
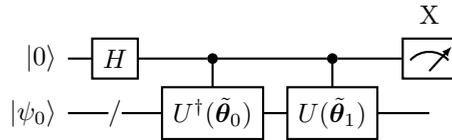
The required number of samples is small, as we only need to resolve whether the sign of the overlap is $+1$ (trivial $\Pi_{\mathcal{C}} = 0$) or -1 (nontrivial $\Pi_{\mathcal{C}} = \pi$).

Implementing the circuit above requires to realize controlled-$U(\boldsymbol{\theta})$, which might increase significantly the depth of the compiled quantum circuit and make the implementation unfeasible on near-term hardware. However, this requirement can be bypassed by using the control-free echo verification technique [156, 196], in place of the standard Hadamard test, to sample Eq. (6.16). This technique requires access to a reference state $|\psi_{\mathrm{ref}}\rangle$ orthogonal to $|\psi_0\rangle$, which should acquire a known eigenphase φ under the action of the PQC, $U(\boldsymbol{\theta}) |\psi_{\mathrm{ref}}\rangle = e^{i\varphi} |\psi_{\mathrm{ref}}\rangle$. As most of the PQC ansätze used for electronic structure states preserve the total number of electrons (including UCC and NPF), the fully unoccupied state $|0...0\rangle$ can be used as reference. Control-free echo verification circuits only require implementing the non-controlled $U(\boldsymbol{\theta})$, and furthermore provide built-in error mitigation power.

6.3.5 Overview of the algorithm

We are now ready to formalize the proposed algorithm for resolving Berry phases, Algorithm 6.2. The formalization we present here will allow us to bound the number of steps and the sampling cost in the following Section 6.4. Given a path $\mathcal{C}$ and a number of steps $1/\Delta t$, the algorithm attempts to calculate $\Pi_{\mathcal{C}}$ yielding either $\Pi_{\mathcal{C}} = 0$, $\Pi_{\mathcal{C}} = \pi$, or a FAIL state. Again, in Section 6.4 we will bound the probability of the FAIL state occurring. Additional features that extend the practicality of the algorithm and mitigate the failure cases are presented in Sec. 6.5 and later implemented in Sec. 6.6.

If the algorithm fails, it can be re-run with a larger number of steps N and thus a smaller step size Δt. A smaller step size decreases additive NR error bound (the error per step scales as Δt^2, the total bound thus scales as Δt).

Algorithm 6.2: Resolve quantized Berry phase

Input: Family of Hamiltonians $H(\mathbf{R})$, $\mathbf{R} \in \mathcal{R}$
　　　Real ansatz PQC $|\psi(\boldsymbol{\theta})\rangle$
　　　Set of initial optimal $\boldsymbol{\theta}_0^*$
　　　Closed path $\mathcal{C} \in \mathcal{R}$, $\mathbf{R}(t) : [0, 1] \mapsto \mathcal{C}$
　　　Number of steps N to discretize $\mathcal{C}$
　　　Precision requirements $\sigma_{\mathcal{G}} > 0$ and $\sigma_{\mathcal{H}} > 0$
　　　Convexity requirement $m_{\mathrm{thr}} > 0$
　　　Regularization reg $\in \{\text{False}, \text{True}\}$
　　　Final fidelity requirement $F \in (0, 1)$
Output: $\Pi_{\mathcal{C}} = 0$ or $\Pi_{\mathcal{C}} = \pi$ or *FAIL*.

1 $\Delta t = 1/N$;
2 $\tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_0^*$;
3 for $\tau \in \{0, \Delta t, 2\,\Delta t, ..., 1 - \Delta t\}$ **do**
4 　 $E(t, \theta) \leftarrow$ define cost function as in Eq. (6.8);
5 　 $\tilde{\mathcal{G}}_j \leftarrow$ sample the gradient to precision $\sigma_{\mathcal{G}}$
6 　　 $\left[\tilde{\mathcal{G}}_j = \frac{\partial E}{\partial \theta_j}(\tau + \Delta t, \tilde{\boldsymbol{\theta}}_\tau)\right]$;
7 　 $\tilde{\mathcal{H}}_{jk} \leftarrow$ sample the Hessian to precision $\sigma_{\mathcal{H}}$
8 　　 $\left[\mathcal{H}_{jk} = \frac{\partial E}{\partial \theta_j \partial \theta_k}(\tau + \Delta t, \tilde{\boldsymbol{\theta}}_\tau)\right]$;
9 　 **if** *reg = False* **then**
10 　　 $\lambda_0 \leftarrow$ lowest eigenvalue of $\tilde{\mathcal{H}}$;
11 　　 **if** $\lambda_0 < m_{thr}$ **then return** *FAIL* and exit;
12 　　 $d\boldsymbol{\theta}^{\mathrm{NR}} \leftarrow -\tilde{\mathcal{H}}^{-1}\tilde{\mathcal{G}}$ (see Eq. (6.12)); $\tilde{\boldsymbol{\theta}}_{\tau+\Delta t} \leftarrow \tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}^{\mathrm{NR}}$;
13 　 **if** *reg = True* **then**
14 　　 $\tilde{\boldsymbol{\theta}}_{\tau+\Delta t} \leftarrow$ **Subroutine 1** $(\tilde{\boldsymbol{\theta}}_\tau, \tilde{\mathcal{G}}, \tilde{\mathcal{H}})$;

15 $f \leftarrow$ final overlap as in Eq. (6.16) to precision F;
16 if $f^2 < F$ **then return** *FAIL* and exit;
17 return $\Pi_{\mathcal{C}} = \arg\{f\}$

6

6.4 Error analysis and bounding

In this section, we find analytic upper bounds on the cost of estimating a quantized Berry phase $\Pi_{\mathcal{C}}$ on a fixed curve $\mathcal{C}$ using Algorithm 6.2. To simplify the treatment regularization and back-tracking are not considered: instead, we require each estimate $\tilde{\boldsymbol{\theta}}_{j\,\Delta t}$ at the j-th step to be within the region of quadratic convergence of the cost function at the next t-step $E((j{+}1)\Delta t, \boldsymbol{\theta})$. We prove that this translates to a guarantee of convergence of the algorithm, under three conditions:

1. At the local minimum $\boldsymbol{\theta}_t^*$, where $\psi(\boldsymbol{\theta}_t^*)$ approximates the ground state, the cost function $E(t, \boldsymbol{\theta})$ is strongly convex [as described in Eq. (6.13)];

2. The number of discretization steps N is sufficiently large;

3. The sampling noise on each of the Hessian and gradient elements ($\sigma_{\mathcal{H}}$ and $\sigma_{\mathcal{G}}$ respectively) is sufficiently small.

The first point entails a requirement on the cost function, defined by the family of Hamiltonians $H(\mathbf{R})$ and the choice of ansatz $|\psi(\boldsymbol{\theta})\rangle$. This requirement is not satisfied if the ansatz state is defined with redundant parameters. We contend that, while strong convexity is a significant assumption, incorporating regularization (or one of the other techniques suggested in the outlook) can alleviate the necessity for such an assumption in practical applications. Our proof provides upper bounds on N and lower bounds on $\sigma_{\mathcal{H}}$ and $\sigma_{\mathcal{G}}$, which suffice to grant convergence. However, these are not to be considered practical prescriptions, as we do not believe them to be optimal; rather they show which are the relevant factors playing a role in the convergence of the algorithm. As the sampled gradient and Hessian are random variables, the guarantee of convergence for bounded error is to be understood in a probabilistic sense.

We first clarify natural assumptions and notation used in the calculation of the basin of convergence of Newton's method. We require the cost function $E(t, \boldsymbol{\theta})$ to be twice-differentiable by $\boldsymbol{\theta}$, for all t, in a region around the true minima $\boldsymbol{\theta}_t^*$. We require the Hessian to be Lipschitz continuous across this region,

$$\|\mathcal{H}(t, \boldsymbol{\theta}) - \mathcal{H}(t, \boldsymbol{\theta} + d\boldsymbol{\theta})\| < L\|d\boldsymbol{\theta}\|. \tag{6.17}$$

(Here, the Lipschitz constant L can be considered a bound $\|\mathcal{T}\| \leq L$ on the norm of the tensor of third derivatives $T(t, \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta})$.) We also require that the gradient of the t-derivative $\dot{\boldsymbol{\mathcal{G}}} = \nabla_{\boldsymbol{\theta}} \frac{dE}{dt}$ is bounded

by $\dot{\mathcal{G}}_{\max}$ in 2-norm. These regularity conditions are satisfied for the PQC ansätze we consider in Sec. 6.5, and in App. 6.B we argue for bounds on L and $\dot{\mathcal{G}}_{\max}$. The strong convexity assumption described in the previous paragraph entails a constant lower bound m_{thr} on the smallest eigenvalue of the Hessian $\mathcal{H}(t, \boldsymbol{\theta}_t^*)$ at the minimizer $\boldsymbol{\theta}_t^*$ for all t.

6.4.1 Bounding the NR error

We will first calculate a lower bound on Δt that ensures the error $\delta\tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*$ is bounded by a constant for all values of t. (as shown in Appendix 6.A, a bound on $\|\delta\tilde{\boldsymbol{\theta}}_1\|_1$ is sufficient to ensure $\Pi_{\mathcal{C}}$ can be accurately resolved.) In this calculation, we will allow for an additive error $\sigma_{\boldsymbol{\theta}}$ on $\tilde{\boldsymbol{\theta}}_t$ due to sampling noise; we will simultaneously calculate an upper bound on $\sigma_{\boldsymbol{\theta}}$. We sketch the calculation here and defer details to App. 6.D.

Firstly, it can be shown (Theorem 6.1 in Appendix 6.D) that the Newton-Raphson step Eq. (6.12) with cost function $E(t + \Delta t, \boldsymbol{\theta})$ is guaranteed to converge quadratically [231] to the minimizer $\boldsymbol{\theta}_{t+\Delta t}^*$ as long as the initial guess $\tilde{\boldsymbol{\theta}}_t$ is within a ball centred in the minimizer of radius $\frac{m_{\mathrm{thr}}}{4L}$. Quadratic convergence means that the distance of the updated guess from the minimizer will scale as the square of the distance of the initial guess from the minimizer,

$$\|\delta\tilde{\boldsymbol{\theta}}_{t+\Delta t}\| \leq \frac{L}{m_{\mathrm{thr}}}\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{t+\Delta t}^*\|^2. \tag{6.18}$$

The right-hand side of this equation can be bounded through the triangle inequality as

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\| \leq \|\delta\tilde{\boldsymbol{\theta}}_t\| + \|\boldsymbol{\theta}_t^* - \boldsymbol{\theta}_{t+\Delta t}^*\|. \tag{6.19}$$

We can bound the second term in this equation by taking the total t-derivative of the optimality condition $\boldsymbol{\mathcal{G}}(t, \boldsymbol{\theta}_t^*) = 0$, yielding

$$\|\boldsymbol{\theta}_{t+\Delta t}^* - \boldsymbol{\theta}_t^*\| \leq m_{\mathrm{thr}}^{-1}\dot{\mathcal{G}}_{\max}\Delta t. \tag{6.20}$$

If at step t we are within the radius given by Theorem 6.1, the NR step will quadratically converge, suppressing also the error from the previous step, and yielding $\|\delta\tilde{\boldsymbol{\theta}}_t\| \leq \frac{m_{\mathrm{thr}}^2}{16L^2}$.

To account for sampling noise effect, we then consider a small additive error $\sigma_{\boldsymbol{\theta}}$ to $\tilde{\boldsymbol{\theta}}_t$. Maximising this allowed sampling noise at each step (as we will see, this becomes the bottleneck in our method) then yields the

6

two bounds

$$\sigma_{\boldsymbol{\theta}} \leq \frac{\sqrt{2}-1}{4}\frac{m_{\mathrm{thr}}}{L}, \quad \Delta t \leq \frac{m_{\mathrm{thr}}^2}{8L\dot{\mathcal{G}}_{\mathrm{max}}}. \tag{6.21}$$

When both bounds are satisfied, we are guaranteed single steps of Newton's method will maintain convergence around the path $\mathcal{C}$.

6.4.2 Bounding the sampling noise

We now translate the bound on $\sigma_{\boldsymbol{\theta}}$ to bounds on the variance of estimates of each element of the gradient and Hessian ($\sigma_{\mathcal{G}}^2$ and $\sigma_{\mathcal{H}}^2$ respectively). This proceeds by simple propagation of variance through Eq. (6.12). We find

$$\sigma_{\boldsymbol{\theta}}^2 := \mathrm{Var}[d\boldsymbol{\theta}_{t,\Delta t}^{\mathrm{NR}}] \leq \quad \|\mathcal{H}^{-1}\|^2 \mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] + \\ + \|\mathcal{H}^{-1}\|^2 \mathbb{E}\left[\|\delta\mathcal{H}\|^2\right]\|d\boldsymbol{\theta}_{t,\Delta t}^{\mathrm{NR}}\|^2, \tag{6.22}$$

where $\delta\boldsymbol{\mathcal{G}}$ and $\delta\mathcal{H}$ are the random variables representing the errors on gradient and Hessian. (a more detailed calculation is given in App. 6.E.) Assuming $\boldsymbol{\theta}$ has n_p elements, each element of the gradient is i.i.d. with variance $\sigma_{\mathcal{G}}^2$, we get

$$\mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] = n_p\,\sigma_{\mathcal{G}}^2. \tag{6.23}$$

As $\delta\mathcal{H}$ is a $n_p \times n_p$ real symmetric matrix, assuming its elements are i.i.d. with variance $\sigma_{\mathcal{H}}^2$, we can invoke Wigner's semicircle law [232] to approximate its norm by $\sqrt{n_p}\,\sigma_{\mathcal{H}}$, thus

$$\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right] \approx n_p\,\sigma_{\mathcal{H}}^2. \tag{6.24}$$

Combining these with Eq. (6.22), and requiring the resulting variance to be small compared to the square allowed additive error $\sigma_{\boldsymbol{\theta}}^2$ we obtain the bound

$$\sigma_{\mathcal{G}}^2 + \sigma_{\mathcal{H}}^2\|d\boldsymbol{\theta}_{t,\Delta t}^{\mathrm{NR}}\|^2 \ll \frac{3-2\sqrt{2}}{16}\frac{m_{\mathrm{thr}}^4}{n_p L^2}. \tag{6.25}$$

We can then bound the norm of the NR update as $\|d\boldsymbol{\theta}_{t,\Delta t}^{\mathrm{NR}}\| \leq m_{\mathrm{thr}}^{-1}\|\boldsymbol{\mathcal{G}}\|$. Splitting the error budget in half we obtain.

$$\sigma_{\mathcal{G}}^2 \ll \frac{3-2\sqrt{2}}{32}\frac{m_{\mathrm{thr}}^4}{n_p L^2} \tag{6.26}$$

$$\sigma_{\mathcal{H}}^2 \ll \frac{3-2\sqrt{2}}{32}\frac{m_{\mathrm{thr}}^6}{n_p L^2\|\boldsymbol{\mathcal{G}}\|^2}. \tag{6.27}$$

Note that these bounds are not tight. For instance, by applying Cauchy-Schwartz inequality to bound $\|\mathcal{H}^{-1} \cdot \mathcal{G}\| \leq \|\mathcal{H}^{-1}\|\|\mathcal{G}\|$, we overlook the fact that the gradient will change more slowly along a lower-eigenvalue eigenmode of the Hessian. We believe further work might allow to define tighter bounds.

6.4.3 Scaling of the total cost

To give an estimate on how many measurements we need to sample gradient and Hessian to sufficient precision, we need to recast the quantities in Eq. (6.27) (the dominant term of the sampling variance) in terms of parameters of the problem. If we use an ansatz without redundancies (or if we can get rid of redundancies through e.g. regularization), and assuming we approximate the ground state well enough, the convexity m_{thr} will be larger than the ground state gap Δ, as every parameterized rotation in the PQC ansatz will introduce a state orthogonal to the ground state. The norm of the gradient and the Lipschitz constant can be bound proportionally to their max norm, as shown in Appendix 6.B, thus $\|\mathcal{G}\| \leq \sqrt{n_p}\|H\|$, $\|\dot{\mathcal{G}}\| \leq \sqrt{n_p}\|\dot{H}\|$ and $L \leq n_p^{3/2}\|H\|$ (where $\|H\|$ is the spectral norm of the Hamiltonian). The number of measurements to sample the Hessian to precision Eq. (6.27) are proportional to the inverse of the bound, with proportionality constant $M_{\mathcal{H}}$ indicating the number of shots required to sample a single element of the Hessian to unit variance (this depends on details such as the decomposition taken to measure the Hamiltonian, and the specifics of the derivative estimation method). Multiplying this by the number of steps $\frac{1}{dt}$ [Eq. (6.21)] gives us the total number of shots required for convergence

$$M_{\mathrm{tot}} = 10^3 \frac{n_p L^3 \|\mathcal{G}\|^2 \dot{\mathcal{G}}_{\max}}{\Delta^8} M_{\mathcal{H}} \tag{6.28}$$

$$< 10^3 \frac{n_p^4 \|H\|^7 \|\dot{H}\|}{\Delta^8} M_{\mathcal{H}}. \tag{6.29}$$

6.5 Adapting to an orbital-optimized PQC ansatz

To achieve a good representation of the ground state character while minimizing depth and number of evaluations of quantum circuits, we employ a hybrid ansatz composed of classical orbital rotations and a parameterized quantum circuit (PQC) to represent correlations within an

active space. The concept of an orbital-optimized variational quantum eigensolver (OO-VQE) is explored in [233, 234] In this section, we introduce the construction of the OO-PQC ansatz and discuss its specific use in our algorithm, where the orbitals need to continuously track a changing active space depending on the nuclear geometry.

6.5.1 An OO-PQC ansatz with geometric continuity

To represent the electronic structure state, we start by choosing an atomic orbital basis, i.e. a discretization of space defined by a set of N non-orthogonal atomic orbitals $\chi_\mu(\mathbf{R}, \boldsymbol{x})$ (functions of the electronic coordinate $\boldsymbol{x} \in \mathbb{R}^3$, where we make explicit the parametric dependence on the nuclear coordinates $\mathbf{R}$); these orbitals define the overlap matrix

$$S_{\mu\nu}(\mathbf{R}) = (\chi_\mu(\mathbf{R})|\chi_\nu(\mathbf{R})) \tag{6.30}$$

$$:= \int_{\mathbb{R}^3} \chi_\mu^*(\mathbf{R}, \boldsymbol{x})\chi_\nu(\mathbf{R}, \boldsymbol{x}) \, d^3\boldsymbol{x}. \tag{6.31}$$

The atomic orbitals (AOs), along with the overlap matrix, depend on the geometry of the molecule specified by the nuclear coordinates $\mathbf{R}$. (For the sake of simplicity, we limit to considering real AOs.) From these, we could define a set of parameterized orthonormal molecular orbitals (MO)

$$\phi_p(\mathbf{R}, C^{\text{AO}}) = \sum_\mu \chi_\mu(\mathbf{R})C_{\mu p}^{\text{AO}}, \tag{6.32}$$

which would allow for the definition of a parameterized active space. The downside of this parametrization is that, to ensure MO orthonormality, we need C^{AO} to satisfy the constraint

$$C^{\text{AO}\dagger}S(\mathbf{R})C^{\text{AO}} = \mathbb{1}, \tag{6.33}$$

which depends nontrivially from $\mathbf{R}$. This implies that we cannot trivially use the same C^{AO} for different geometries $\mathbf{R}$.

In order to address this problem, we have opted to use orthonormalized atomic orbitals (OAO) that are derived from the AOs through symmetric Löwdin orthogonalization [235] as reference in the definition of parameterized MOs. The OAOs are defined as $\phi_p^{\text{OAO}}(\mathbf{R}) = \sum_\mu \chi_\mu(\mathbf{R})S_{\mu p}^{-1/2}(\mathbf{R})$. Building on these, we can define the MOs as

$$\phi_q(\mathbf{R}, C) = \sum_{\mu,p} \chi_\mu(\mathbf{R})S_{\mu p}^{-1/2}(\mathbf{R})C_{pq}, \tag{6.34}$$

where $C = S^{1/2}(\mathbf{R}) \cdot C^{\mathrm{AO}}$. The orthonormality constraint Eq. (6.33) then reduces to requiring C to be orthogonal, and it is independent on $\mathbf{R}$. In summary, Eq. (6.34) defines a set of orthonormal molecular orbitals parameterized by C, well-defined and continuous for changing $\mathbf{R}$.

To start up our algorithm, the matrix C^{AO} can be initialized by a Hartree-Fock (or any other molecular coefficient matrix, e.g. coming from a small CASSCF calculation) at some initial geometry $\mathbf{R}(0)$. From this we recover $C = S^{1/2}(\mathbf{R}(0)) \cdot C^{\mathrm{AO}}$, which is then treated as a variational parameter of the ansatz. Using the parameterized MO Eq. (6.34) we construct the electronic structure Hamiltonian $H(R, C)$ in the (parameterized) molecular basis.

Based on the initial Hartree-Fock orbital energies, we split the N orbitals into a core set with N_{O} doubly-occupied orbitals, an active set with N_{A} orbitals, and a virtual set with N_{V} empty orbitals. Although the split of the orbital indices remains constant throughout the algorithm, the orbitals themselves continuously change through their dependence on $\mathbf{R}$ and C. The correlations are treated only within an active space of η_{A} electrons in N_{A} orbitals. Tracing out the core and virtual orbitals yields the active-space Hamiltonian $H_{\mathrm{A}}(\mathbf{R}, C)$.

The correlated active-space state $|\psi(\boldsymbol{\theta})\rangle$ is represented on a quantum device, using a PQC ansatz of the form Eq. (6.10). The cost function then becomes

$$E(\mathbf{R}, C, \boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta})| H_{\mathrm{A}}(\mathbf{R}, C) |\psi(\boldsymbol{\theta})\rangle, \qquad (6.35)$$

and it can be evaluated by sampling the 1- and 2-electron reduced density matrix (RDM) of the state [67]. (Other efficient sampling schemes, e.g. based on double factorization [73, 236], can be used.)

6.5.2 Measuring boundary terms with the OO-PQC ansatz

When evaluating the final overlap [Eq. (6.16)] with an orbital-optimized ansatz, we have to consider that the states $|\psi(\boldsymbol{\theta}_0)\rangle$ and $|\psi(\boldsymbol{\theta}_1)\rangle$ are defined on different active space orbitals, determined by the MO matrices C_0 and C_1 respectively. The transformation between the two sets of orbitals, $\phi(\mathbf{R}, C_1) = \phi(\mathbf{R}, C_0) \cdot C_{0 \to 1}$, is represented by the orthogonal matrix

$$C_{0 \to 1} = C_0^{\dagger} C_1. \qquad (6.36)$$

If the algorithms successfully tracked the lowest-energy active space state of the system, the Hilbert spaces spanned by the active orbitals defined

by C_0 and C_1 should match. (The same is true for the core space and the virtual space.) This implies the matrix $C_{0\to1}$ will have block structure, with $[C_{0\to1}]_{pq} \neq 0$ only if p, q are both in the same set of orbitals (core, active or virtual). The orbital rotations within the core (virtual) subspace will not generate any phase on the state of the system, as all orbitals are doubly-occupied (doubly-unoccupied). The orbital rotation within the active space can then be translated to a unitary transformation on the state by a Bogoliubov transformation

$$G_{0\to1} = \exp\left\{ \sum_{\{p,q\}\in\text{AS}} [\log(C_{0\to1})]_{pq} c_p^\dagger c_q \right\}, \tag{6.37}$$

with $c_p^\dagger, c_p$ fermionic creation and annihilation operators on the p orbital. The final overlap

$$\text{Re}\left[\langle 0| U^\dagger(\tilde{\boldsymbol{\theta}}_0) G_{0\to1} U(\tilde{\boldsymbol{\theta}}_1) |0\rangle \right] =: \omega_{\mathcal{C}} \tag{6.38}$$

can then be sampled with a Hadamard test, given an quantum circuit implementing the (eventually controlled) operation $G_{0\to1}$. Under Jordan-Wigner encoding, a quantum circuit for $G_{0\to1}$ can be implemented as a fabric of parameterized fermionic swap gates of depth N_A following a QR decomposition of the orbital rotation generator $[\log(C_{0\to1})]_{pq}$, also known as a givens rotation fabric [103]. These gates preserve the zero-electrons reference state, allowing to employ the ancilla-free echo verification technique mentioned in section 6.3.4 to measure the final overlap.

6.5.3 Newton-Raphson updates of the OO-PQC ansatz

The proposed OO ansatz has two sets of parameters, C and $\boldsymbol{\theta}$. As the MO matrix C is subject to the constraint Eq. (6.33), its elements cannot be freely updated with NR. Instead, for each NR update with initial MO matrix C, we reparametrize the MOs with a unitary transformation: $C \leftarrow C e^{-\kappa}$, where κ is any antisymmetric matrix. The derivatives of the energy with respect to any element κ_{pq} can be evaluated analytically (see Appendix 6.C). Furthermore, under this parametrization it can be shown that κ_{pq} where p, q are both core indices or both virtual indices are *redundant* [219] in the definition of the active space orbitals; these $N_\text{O}^2 + N_\text{V}^2$ parameters are set to zero without reducing the expressivity of the ansatz. We call the unraveled set of remaining parameters $\boldsymbol{\kappa}$. To implement the Newton-Raphson step, the gradient and Hessian with respect to the combined set of parameters $(\boldsymbol{\theta}, \boldsymbol{\kappa})$ is computed. In this

manner the gradient splits into two components, and the Hessian into three components

$$\nabla_{(\boldsymbol{\theta},\boldsymbol{\kappa})}E = (\nabla_{\boldsymbol{\theta}}E, \nabla_{\boldsymbol{\kappa}}E) \tag{6.39}$$

$$\nabla^2_{(\boldsymbol{\theta},\boldsymbol{\kappa})}E = \begin{bmatrix} \nabla^2_{\boldsymbol{\theta}}E & \nabla_{\boldsymbol{\kappa}}\nabla_{\boldsymbol{\theta}}E \\ (\nabla_{\boldsymbol{\kappa}}\nabla_{\boldsymbol{\theta}}E)^{\intercal} & \nabla^2_{\boldsymbol{\kappa}}E \end{bmatrix}. \tag{6.40}$$

The PQC parameter derivatives $\nabla_{\boldsymbol{\theta}}E$ and $\nabla^2_{\boldsymbol{\theta}}E$ can be evaluated through parameter-shift rule [228, 237] or finite difference, by sampling on the quantum device. The derivatives with respect to the OO parameters $\nabla_{\boldsymbol{\kappa}}E$ and $\nabla^2_{\boldsymbol{\kappa}}E$ are linear functions of the 2-electron RDM, whose coefficients can be computed analytically [219]. The remaining component, the mixed Hessian $\nabla_{\boldsymbol{\kappa}}\nabla_{\boldsymbol{\theta}}E$ can be similarly evaluated as a linear function of $\boldsymbol{\theta}$-derivatives of the RDM; for this, we can use the same data sampled from the quantum device to evaluate $\nabla_{\boldsymbol{\theta}}E$. We detail the procedure of estimating these Hessian components in App. 6.C. Thus, evaluating the derivatives with respect to the OO parameters does not require extra sampling on the quantum device.

6.6 Numerical results

In this section, we demonstrate the application of our method to a small model system: the formaldimine molecule $H_2C{=}NH$, an established model in the context of quantum algorithms for excited states in [233, 234, 238]. This molecule is known to have a conical intersection between the singlet ground state and first excited state potential energy surfaces [239]. This CI plays an important role in the photoisomerization process of formaldimine, which in turn can be considered a minimal model for the photoisomerization of the rhodopsin protonated Schiff-base (a key step in the visual cycle process [240, 241]). We consider geometries obtained from the equilibrium configuration by varying the direction of the N–H bond, defined by the bending angle α and the dihedral angle ϕ (see Fig. 6.1d). Varying these angles defines the considered plane in nuclear configuration space $\mathcal{R}$. First, we consider a minimal model of formaldimine (within the minimal basis and a small active space), on which we can test the properties of the algorithm 6.2. Then, we investigate the effects of sampling noise on these results. Finally, we study a more complex model of the same molecule (with a larger basis set and active space), and show that we can achieve similar results by employing regularization and backtracking to deal with the degeneracies of the ansatz manifold.

6

6.6.1 Numerical simulation details

For our simulations, we use the PennyLane [171] package with the Py-Torch backend to construct the hybrid quantum-classical cost function of Eq. (6.35) supporting automatic differentiation (AD) with respect to the PQC parameters. To achieve this, we implement the transformation of the one- and two-body AO basis integrals to the parameterized MO basis [Eq. (6.34)] with AD support. The transformed integrals are projected onto the active space and contracted with the active space one- and two-electron RDM. The RDM elements and their derivatives with respect to the PQC parameters are obtained using PennyLane and its built-in AD scheme based on the parameter shift rule. In summary, gradients and Hessians of the cost function with respect to the PQC parameters are obtained using AD, the orbital gradient and Hessian components are estimated analytically (see Appendix 6.C), and the off-diagonal block of the composite Hessian [Eq. (6.40)] is retrieved by automatic differentiation of the analytical orbital gradient. We use PySCF [242] to generate the molecular integrals (i.e. the full space one- and two-electron integrals and overlap matrices) in the atomic orbital basis.

The core code developed for this project is made available as a python package in the GitHub repository [216]. This code provides a flexible implementation of the orbital-optimized PQC ansatz, which can find many applications in VQAs for chemistry. A tutorial Jupyter notebook showcasing a calculation of Berry phase in the minimal model of Formaldimine is provided in the `examples` folder in the repository.

6.6.2 Minimal model with an degeneracy-free ansatz

We first demonstrate the application of our algorithm to a minimal model of formaldimine, for which we can approximate the ground state with a simple ansatz with no degeneracies. The molecule is described in a minimal STO-3G basis-set, and we select an active space of $\eta_A = 2$ electrons in $N_A = 2$ spatial orbitals [i.e. CAS(2,2)]. As the orbital optimization already allows (spin-adapted) single excitations within the active space, the only parameterized gate we can include in our PQC ansatz is the double-excitation $U(\theta) = e^{\theta(c_0^\dagger c_1^\dagger c_2 c_3 - c_2^\dagger c_3^\dagger c_0 c_1)}$; this corresponds to the unitary coupled-cluster doubles [UCC(S)D] ansatz, where the singles (S) are not explicitly included because they would be redundant with the orbital optimization. This is enough to describe exactly any active space state compatible with the symmetries of the model, without over-parametrizing the ansatz state.
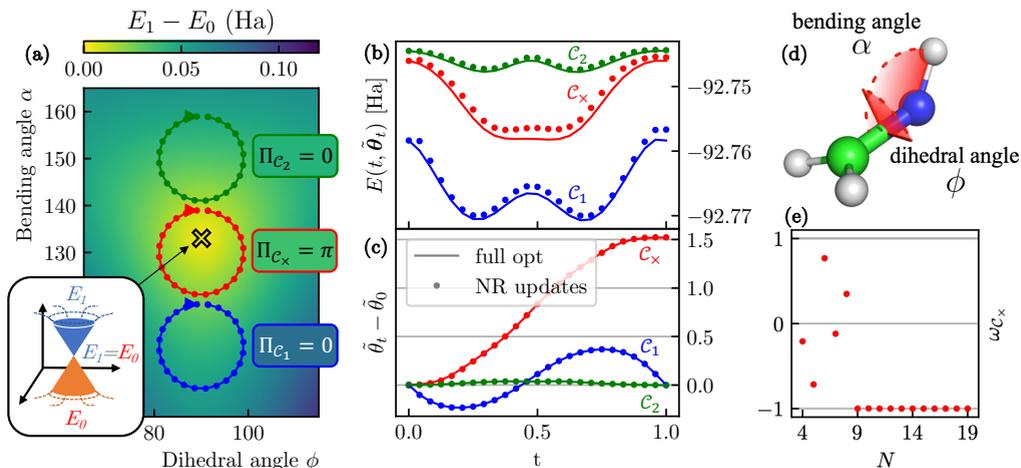
Figure 6.1: **(a)** Three loops in the nuclear configuration space of formaldimine; $\mathcal{C}_1$ (green), $\mathcal{C}_\times$ (red) and $\mathcal{C}_2$ (blue). $\mathcal{C}_\times$ encircles a CI, resulting in a non-trivial Berry phase. In this representation, the loops are discretized by $N = 25$ points. The color plot indicates the energy gap at the full configuration interaction (FCI) level. **(b)** Energy and **(c)** change in PQC parameter around the three loops, with the same color coding as in (a) and the same $N = 25$. These results refer to a (2,2) active space, and a minimal (STO-3G) basis set description of Formaldimine, with a OO-UCCD ansatz that has a single θ parametrizing the only double excitation. The continuous lines show the true optimum $\boldsymbol{\theta}_t^*$ and the relative energy (obtained by full optimization), while the markers show the progress of the estimate $\tilde{\boldsymbol{\theta}}_t$ from Algorithm 6.2, in absence of sampling noise. The Hessian stays positive throughout the path, no regularization is needed. **(e)** Final overlap computed by Algorithm 6.2 for the red loop containing a CI, for a varying number of total discretization points N. For $N < 9$, the Hessian is not always positive and regularization is needed to invert the Hessian, but no backtracking is used. **(d)** Schematic representation of Formaldimine, indicating the parameters used to define the nuclear geometries in this work.

In Fig. 6.1 we demonstrate the application of our algorithm to this model. The minimal basis set is small enough that we can run a full configuration interaction (FCI) calculation to exactly resolve the ground and first excited state energies $E_0(\mathbf{R})$ and $E_1(\mathbf{R})$. Observing the gap

$E_1(\mathbf{R}) - E_0(\mathbf{R})$ (portrayed in Fig. 6.1a), we can determine the location of the conical intersection $\phi^\times = 90°$, $\alpha^\times \approx 132°$. We then define three loops in the configuration space $\mathcal{R}$, one loop $\mathcal{C}_\times$ containing the CI and two "trivial" loops $\mathcal{C}_1, \mathcal{C}_2$. These loops are centered around $\phi = 90°$ and $\alpha = 110°$ ($\mathcal{C}_1$), $\alpha = 130°$ ($\mathcal{C}_\times$) and $\alpha = 150°$ ($\mathcal{C}_2$), and all have a radius of 10°. Fig. 6.1c shows the progress of the estimate $\tilde{\theta}_t$ (shifted by $\tilde{\theta}_0$) of the optimal PQC parameter θ_t^* throughout the $N = 25$ single-NR-update t-steps. Note that, while we only plot the single PQC parameter relative to the parameterized double excitation, the algorithm updates the MO coefficients $\tilde{C}_t$ as well. We observe that $\tilde{\boldsymbol{\theta}}_1 = \tilde{\boldsymbol{\theta}}_0$ for the trivial loops $\mathcal{C}_1, \mathcal{C}_2$, while $\tilde{\boldsymbol{\theta}}_1 \neq \tilde{\boldsymbol{\theta}}_0$ for the loop $\mathcal{C}_\times$ containing the CI. This is an indication of the effect of the Berry phase, but not the result of the algorithm yet; measuring the overlap Eq. (6.38) yields the correct Berry phase $\Pi_\mathcal{C} = \arg[\omega_\mathcal{C}]$. The estimated energy $E(t, \tilde{\boldsymbol{\theta}}_t)$ and the optimal $E(t, \boldsymbol{\theta}_t^*)$ (obtained by full local optimization) are shown in Fig. 6.1b. We can observe a small deviation from the optimal energy in the region where the character of the state changes faster (along the line $\phi = \phi^\times$, $\alpha < \alpha^\times$), but this does not disrupt the tracking of the minimum. Finally, Fig. 6.1e shows that a number of discretization points $N \geq 9$ is needed to correctly resolve $\Pi_{\mathcal{C}_\times} = \pi$, through the evaluation of the overlap [Eq. (6.38)] $\omega_{\mathcal{C}_\times} = -1$.

6.6.3 Sampling noise

In this section, we explore the robustness of our algorithm with respect to the sampling noise characteristic of VQAs. To avoid defining a specific sampling strategy and keep our results general, we directly add a proxy of sampling noise η to each element of the gradient and Hessian. Each η is an independent Gaussian random variable with variance σ^2; a different η is added to each element of the gradient of (6.39) and Hessian of Eq. (6.40) to get the noisy estimates $\tilde{\mathcal{G}}$ and $\tilde{\mathcal{H}}$. In Fig. 6.2 we show the energy profile of the three loops whose geometry is represented in Fig. 6.1a, for one such random realisation of the sampling noise. (The plotted energy expectation is evaluated exactly, noise is only added to the gradient and Hessian used in the NR updates.) These three loops yield the same Berry phase results as the noiseless case.

The probability P_{success} of Algorithm 6.2 correctly resolving the Berry phase $\Pi_{\mathcal{C}_\times}$ on the nontrivial loop $\mathcal{C}_\times$ is reported in Fig. 6.3, as a function of the number of discretization steps N and of the variance of the added noise on each sampled quantity σ^2. The expected final overlap Eq. (6.38) is -1 for this case, as the loop contains a CI. For each value of N and σ^2, we simulate 100 noisy runs of the algorithm and we declare as successful
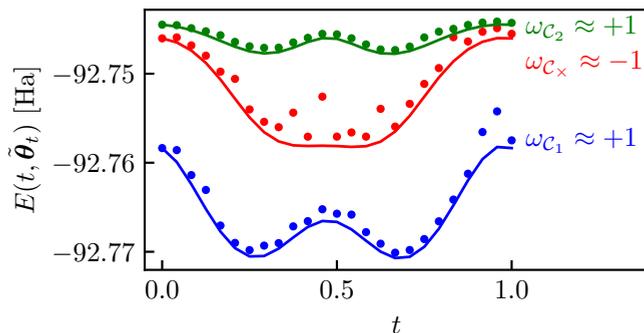
Figure 6.2: Energies throughout the loops in Fig. 6.1 (discretized by $N = 25$ steps), in the presence of sampling noise. Each element of the composite gradient Eq. (6.39) and Hessian Eq. (6.40) are perturbed by random gaussian noise with variance $\sigma^2 = 5 \times 10^{-6}$. In this instance the Hessian stays positive around the path, so no regularization is needed. The full lines, like in Fig. 6.1b, indicate the true optimum $E(t, \boldsymbol{\theta}_t^*)$ obtained by full optimization.

the ones that yield a negative final overlap (implying $\Pi_{C_\times} = \pi$). Finally we average these outcomes to retrieve the succes probabilities P_{succes}.

From these simulations we conclude that sampling noise reduces the accuracy of tracking the ground state and thus increases the probability of obtaining inaccurate energies. Nevertheless, for a moderate amount of sampling noise our algorithm still resolves the Berry phase correctly. We observe that an error on each gradient and Hessian element with variance of $\sigma^2 = 10^{-5}$ (or smaller) does not compromise the resolution of the Berry phase, as long as the number of discretization points is sufficiently high ($N > 10$, very close to the noiseless case portrayed in Fig. 6.1e). On the other hand, a large enough sampling error ($\sigma^2 \geq 5 \times 10^{-5}$) produces essentially random results ($P_{\text{success}} \approx 50\%$).

6.6.4 Larger basis and active space

To test convergence for a more realistic case where the cost function is not always strongly convex at its minima, we simulate the algorithm on a more challenging model of formaldimine. The model is constructed employing the cc-pVDZ basis set (43 atomic orbitals), and an active space of four electrons in four spatial orbitals [CAS(4,4)]. As a PQC ansatz
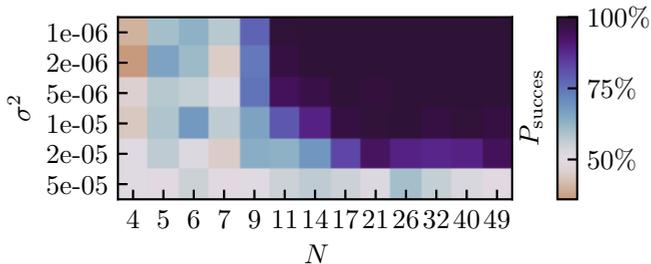
Figure 6.3: Success probability P_{succes} of the algorithm as a function of the number of discretization points N and the sampling noise variance σ^2 for the loop $\mathcal{C}_\times$ containing a CI. The details of the model and geometry of the loop $\mathcal{C}_\times$ match Fig. 6.1 (red loop). Success is defined by resolving a final overlap $\langle \psi(\tilde{\boldsymbol{\theta}}_0) | \psi(\tilde{\boldsymbol{\theta}}_1) \rangle < 0$, which returns $\Pi_{\mathcal{C}_\times} = \pi$. The success probability is computed over 100 simulated runs. A $P_{\text{success}} \approx 50\%$ indicates the algorithm returns random outcomes. Regularization by Hessian augmentation is enabled in these calculations, while no backtracking is used.

for the active space state, we use the number-preserving fabric (NPF) ansatz introduced in [57], consisting of a fabric of spin-adapted orbital rotations and double excitations on sets of two spatial orbitals (four spin-orbitals). Four layers of this ansatz are enough to recover the exact CASCI ground state energy inside the active space of 4 orbitals, resulting in 20 PQC parameters. This is an overparameterization of the ground state, implying a global redundancy in the ansatz and resulting in a singular hessian at every point. The goal of this numerical demonstration is to show that Algorithm 6.2 can still recover the Berry phase, in this case using regularization and backtracking.

In Fig. 6.4 the energies throughout two loops are shown. The location of the conical intersection $(\alpha^\times, \phi^\times)$ in the larger basis set moves compared the case shown in Fig. 6.1 (this is to be expected, as the cc-pVDZ and STO-3G models are effectively different); the basis is now too large to attempt an FCI calculation that would resolve the gap exactly. One could instead resort to a State-Averaged CASSCF calculation to resolve the location of the CI, however, the state-average approach might bias the location of the CI. For this demonstration, we manually select two loops with a slightly larger radius of 15°, centered around $\alpha = 113°$ ($\mathcal{C}_\times$, red line) and around
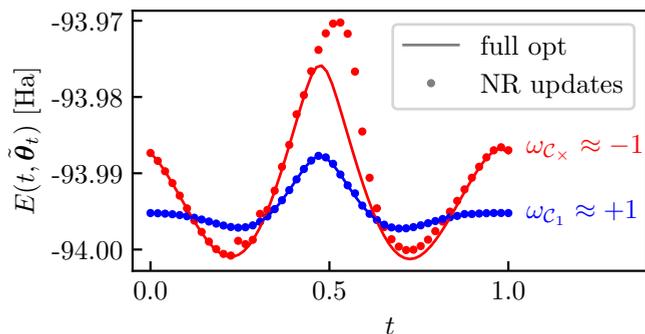
Figure 6.4: Energies throughout two loops, for a more challenging model of Formalimine, described in the cc-pVDZ basis set and with a (4,4) active space. The (red) blue loop indicates a (non-)trivial Berry phase, centered around ($\alpha = 113°$) $\alpha = 145°$ and $\phi = 90°$ (both), with a radius of $15°$, and discretized by $N = 50$ points (note these loops are different from Fig. 6.1). The basis set is too large to run a FCI calculation to locate the CI, and using another approximate method (e.g. state-average CASSCF) might bias the CI location. Instead, we manually choose larger loop geometries and use Algorithm 6.2 to find the value of $\Pi_{\mathcal{C}}$.

$\alpha = 145°$ ($\mathcal{C}_1$, blue line). These values are chosen based on the location of the CI at the level of theory of large state-average CAS(14,14)SCF calculation, which returns $\alpha^\times \approx 113°$. We choose $\phi = 90°$, as the CI is forced to lie on the $\phi^\times = 90$ hyperplane due to the Cs reflection point-group symmetry. Indeed, we can resolve the correct Berry phase with only $N = 50$ discretiation points, accumulating a small error around the loop which has a minor effect on the final overlap [Eq. (6.38)] of $\omega_{\mathcal{C}_\times} = -0.9994$ (loop containing the CI), and $\omega_{\mathcal{C}_1} = 0.99998$ (trivial loop).

6.7 Conclusion and outlook

In this work, we introduced a hybrid algorithm to resolve conical intersections through the Berry phase they induce. This is achieved by tracking the ground state with a variational quantum ansatz, along a closed path $\mathcal{C}$ in nuclear configuration space. This algorithm only requires approximating the ground state (in contrast to e.g. the state-average VQE [233]) for one nuclear geometry $\mathbf{R}$ at a time, reducing the expressivity requirements

of the ansatz. The key requirement of the algorithm is that the ansatz parameters are changed smoothly, tracking a local minimum of the cost function (6.8) and ensuring the $U(1)$-gauge (global phase) of the ansatz state remains well-defined.

As the output is quantized ($\Pi_{\mathcal{C}} \in \{0, \pi\}$), the result only needs to be estimated to a constant precision, and the algorithm is robust to some amount of error; we considered optimization error (the error on estimates of a variational parameter, due to the approximate minimizer) and sampling noise on the quantities measured on the quantum device. We showed that one can update the variational parameters with a single newton step for each geometry in a discretization of the loop $\mathcal{C}$. We proved analytically that the algorithm is granted to converge for a large enough number of discretization steps N, and small enough additive error, under the assumption that the cost function is strongly convex at its minimum. (We consider sampling noise explicitly, but the robustness results extends to any additive noise, including hardware noise.) We argue this result practically extends to cases where the strong convexity assumption is not satisfied, as long as some regularization technique is employed.

This reasoning is corroborated by numerical demonstrations of CI resolution on a small example system – the formaldimine molecule. Using a minimal description of formaldimine [STO-3G basis, a CAS(2,2), UCCD ansatz], for which we have a strongly convex cost function, we show convergence of Algorithm 6.2 without using regularization for a sufficiently large $N > 11$, as we expect from our analytical results. We also demonstrate the effect of sampling noise in this setting, showing that our algorithm is robust to a sizable amount of noise, achieving convergence for N comparable to the noiseless case. Finally, we demonstrate the application of Algorithm 6.2 with regularization on a more complicated and realistic model of formaldimine [cc-pVDZ basis, CAS(4,4), NPF ansatz]. This case shows that, even with a cost function that is never convex, we can employ regularization to resolve the Berry phase correctly.

6.7.1 Paths towards improving convergence

The key step in our algorithm, where most of the cost in terms of quantum resources is concentrated, is the evaluation of the (NR) parameter update. Ensuring the parameter estimates remain within the basin of convergence of the cost function is crucial, and it is the bottleneck in terms of the cost of our algorithm. As shown in Section 6.4, the size of the basin of convergence depends on the convexity of the cost function at the minimum (6.13). Overparametrizations (local or global) of the cost function are

especially disrupting, as they produce singular Hessians ($m = 0$) even at optimal points. In this work, we proposed to use regularization by Hessian augmentation and back-tracking to solve this problem; this technique is practical and, as shown numerically in Section 6.6.4, it can produce convergent results for systems with overparametrized cost functions. However, the success of these techniques may depend on the choice of their hyperparameters (α, β, μ, ρ in Algorithm 6.1), and their application makes the analytical study of the algorithm convergence harder. In this section, we suggest alternative approaches to regularization and resource allocation for the algorithm, which would require further study.

Quantum natural gradients — Quantum natural gradient (QNG) descent is a recently-proposed parameter update technique [243, 244], which takes into account the geometry of the ground state manifold. Natural gradient techniques, already long in use in classical machine learning [245], are invariant with respect to reparametrizations of the cost function; more importantly for our case, they nullify the effect of overparametrizations [246]. The idea of QNG is to transform gradients with respect to the ansatz parameters into gradients with respect to Quantum Information Geometry. This reparametrization is achieved through the Fubini-Study metric tensor $g(\tilde{\boldsymbol{\theta}}_t)$ (to be evaluated at each update). A gradient descent step would then become:

$$\tilde{\boldsymbol{\theta}}_{t+\Delta t} = \tilde{\boldsymbol{\theta}}_t - \eta g^+(\tilde{\boldsymbol{\theta}}_t)\boldsymbol{\mathcal{G}}(t + \Delta t, \tilde{\boldsymbol{\theta}}_t), \tag{6.41}$$

where η is the learning rate and $g^+(\tilde{\boldsymbol{\theta}}_t)$ is the pseudo-inverse of the metric tensor. Here $\boldsymbol{\mathcal{G}}(t + \Delta t, \tilde{\boldsymbol{\theta}}_t)$ is just the usual gradient of the energy as in Algorithm 6.2. The resulting QNG step updates the ansatz by a fixed amount in the norm induced by the distance between quantum states, instead of a parameter-space norm, solving the issues connected to overparametrization. Another option would be to use classical natural gradients (NG) [245] defined on the 1- and 2-RDM manifold, which rely on the classical Fisher information matrix.

Adaptive step selection — Choosing a sufficient number of steps N to discretize $\mathcal{C}$ is key to the success of our algorithm, as proven in Section 6.4 and shown in Fig. 6.1 (bottom right). This is because the minimum $\boldsymbol{\theta}_t^*$, along with its basin of convergence, changes between subsequent steps by an amount proportional to the step size $\Delta t = 1/N$. In Algorithm 6.2, we propose to linearly discretize a given parametrization of $\mathcal{C}$ for the sake of simplicity. An adaptive choice of Δt could greatly reduce the cost of the algorithm, letting the steps be larger in the regions where the ground state changes the least, while concentrating more points in the regions where the

ground state character sharply shifts. An adaptive step selection technique that preserves the provable convergence could be easily implemented if the gradient and Hessian of the cost function are measured through the 1- and 2-electron RDM. In fact, the cost function can then be written as

$$E(t, \boldsymbol{\theta}) = \sum_\xi D(\boldsymbol{\theta})_\xi h(t)_\xi, \qquad (6.42)$$

where $\xi = pq, pqrs \in [N_A]$ are the one- and two-body active-space orbital indices, $D(\boldsymbol{\theta})_\xi$ are the RDM elements and $h(t)_\xi$ are the one- and two-electron integrals. The derivatives with respect to $\boldsymbol{\theta}$ are then calculated by chain rule from the derivatives of the RDMs, at the current parameter value of $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_t$. This allows to compute energy $E(t', \tilde{\boldsymbol{\theta}}_t)$, gradient $\mathcal{G}(t', \tilde{\boldsymbol{\theta}}_t)$ and Hessian $\mathcal{H}(t', \tilde{\boldsymbol{\theta}}_t)$ for any value t', without further evaluations of the PQC. The step size can be then chosen as the maximum Δt such that the Hessian convexity $m(t + \Delta t, \tilde{\boldsymbol{\theta}}_t)$ remains above some positive threshold value. Further research could quantify the improvement that adaptive step selection would bring to our algorithm, and identify a method to implement this for optimized energy derivatives sampling techniques, such as those using double factorization [247].

6.7.2 Potential applications

The algorithm we propose resolves the Berry phase along a given path $\mathcal{C}$; the description of the loop is an input of the algorithm. This loop construction will depend on the details of the considered application, and might involve chemical intuition and the consideration symmetries of the molecule where present. In realistic applications, we conceive our algorithm as a tool that can help to (1) certify CIs proposed by other methods, (2) determine whether a CI plays a role in a certain reaction, and/or (3) locate a point of the CI manifold in parameter space. In either case, an initial proposal of a path $\mathcal{C}$ that might contain the CI is necessary.

The case 1 is the most direct application of our algorithm. The loop $\mathcal{C}$ is chosen to surround a quasi-degeneracy previously identified by an approximate classical method. The result of our algorithm could then confirm or disprove the presence of the CI. In case 2, given a photochemical reaction whose geometry is approximately known, we can use our algorithm on a set of loops to understand wether a CI plays a role in the reaction. These loops can be constructed by variations of the reaction path along perpendicular coordinates, focusing on the modes that influence the orbitals involved in the reaction, thus greatly reducing the search space for the

CI. Finally, to locate the CI (case 3) various search approaches can be considered. For example, starting from a large loop $\mathcal{C}$ that is known to contain the CI, binary-search can be used to iteratively shrink the loop and locate the CI to the desired precision. The considered loops could be defined on a plane, if there is heuristic information about the direction along which the potential energy surfaces split. In alternative, a mesh of orthogonal loops in a subspace of the nuclear configuration space $\mathcal{R}$ can be tested. This, in combination with the data about the ground state energy collected by running the optimization in our algorithm, could also be used to determine the approximate location of the minimal energy crossing point [i.e. the point $\mathbf{R}_{\mathrm{MECP}}$ on the CI manifold with minimum $E_0(\mathbf{R}_{\mathrm{MECP}}) = E_1(\mathbf{R}_{\mathrm{MECP}})$]. Further work is needed to explore these problems, develop procedures to solve them and define and test practical application cases in the three categories.

6.7.3 Outlook

The bounds presented in Section 6.4 have been calculated to provide a guarantee of convergence for our method, which is an atypical feature for variational quantum algorithms. These are not supposed to be tight bounds or resource estimates for a realistic application of our algorithm. Further research is needed to define better bounds. This, along with the choice of a specific method to extract the energy and its derivatives (e.g. RDM sampling and parameter shift rule) could allow estimating the cost of a practical application of this algorithm. Furthermore, a study of the errors due to the ansatz not perfectly reproducing the ground state, and those induced by circuit noise, could help to understand the practical limitations of the algorithm.

The computation of Berry phases is also central when characterizing topological phases of matter [248]. For the specific case of non-interacting Hamiltonians with chiral or inversion symmetry, the winding number is analytically computed through the Zak phase [249], which is related to the Berry phase that is accumulated after a closed loop through the Brillouin Zone. For interacting systems, in which one cannot access momentum space, there is a mechanism in which one introduces an external periodic perturbation to the Hamiltonian [250, 251]. As long as the perturbation does not close the gap and respects the symmetries of the system, the Berry phase can be computed by considering a closed loop in parameter space, similar to what is proposed in this work. As an outlook, one could consider extending the VQE approach to detect topological phases of matter through the computation of the Zak phase. Furthermore, VQA

6

approaches to other topological invariants, such as the Chern number [252] could be considered (possibly inspired by methods related to their experimental detection, such as Thouless pumping [253]).

Finally, classical algorithms to resolve conical intersections from Berry phases inspired by this approach could be designed. If the approximate ground state is represented by a variational classical ansatz that fixes the $U(1)$ gauge, a simple extension of our method could be achievable. For example, this could be achieved with an extension of CASSCF (essentially a CASCI solver on top of an orbital optimization) that implements continuous local optimization of the SCF matrix, to allow enforcing the smoothness constrains that are crucial to keep the gauge of the state fixed.

6.A Bounding overlaps by change in ansatz parameters

The variational Berry phase for a real ansatz state (as introduced in section 6.3.1) is resolved as the argument of the boundary term $\arg[\langle\psi(\tilde{\boldsymbol{\theta}}(0))|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle]$. To obtain a nontrivial Berry phase, the initial and final parameters $\tilde{\boldsymbol{\theta}}(0)$ and $\tilde{\boldsymbol{\theta}}(1)$ need to be far enough to allow $\langle\psi(\tilde{\boldsymbol{\theta}}(0))|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle = -1$. This implies that the optimal parameters need to change enough along the parametrization of the path (for t going from 0 to 1). In this appendix, we translate this into a lower bound on the one-norm-distance between initial and final parameters $\|\tilde{\boldsymbol{\theta}}(1) - \tilde{\boldsymbol{\theta}}(0)\|_1$. As a consequence, we also find a lower bound on how much error can be allowed without on the final parameter $\tilde{\boldsymbol{\theta}}(1)$ compromising the Berry phase measurement.

We first state a lemma which will be useful in the proof:

Lemma 5

Given two unitary operators U, U', each decomposed as a product of $N \geq 2$ unitary operators $U = U_{N-1}...U_1 U_0$, there holds the bound $\|U - U'\| \leq \sum_{j\in[N]}\|U_j - U_j'\|$.

We prove this by induction. For $N = 2$, the proof is by the triangle inequality

$$\|U_1' U_0' - U_1 U_0\| = \|U_1'(U_0' - U_0) + (U_1' - U_1)U_0\| \tag{6.43}$$

$$\leq \|U_1'\|\|(U_0' - U_0)\| + \|(U_1' - U_1)\|\|U_0\| \tag{6.44}$$

$$= \|(U_0' - U_0)\| + \|(U_1' - U_1)\|. \tag{6.45}$$

The proof for $N + 1$ can be similarly be reduced to the proof for N:

$$\|U'_N U'_{N-1}...U'_0 - U_N U_{N-1}...U_0\| \leq \|U'_N - U_N\| + \|U'_{N-1}...U'_0 - U_{N-1}...U_0\|. \tag{6.46}$$

Let us consider the case of real Ansatz Eq. (6.10). We remind $U_i = e^{A_i \theta_i}$ for antisymmetric real A_i. We can then bound the overlap between ansatz states using Lemma 5,

$$1 - \langle\psi(\boldsymbol{\theta})|\psi(\boldsymbol{\theta}')\rangle = 1 - \langle\psi_0| U^\dagger(\boldsymbol{\theta})U(\boldsymbol{\theta}') |\psi_0\rangle \tag{6.47}$$

$$= \langle\psi_0| U^\dagger(\boldsymbol{\theta})[U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}')] |\psi_0\rangle \tag{6.48}$$

$$\leq \|U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}')\| \tag{6.49}$$

$$\leq \sum_{j \in [N]} \|U_j(\theta_j) - U_j(\theta'_j)\| \tag{6.50}$$

$$= \sum_{j \in [N]} \|e^{A_j \theta_j} - e^{A_j \theta'_j}\| \tag{6.51}$$

$$= 2 \sum_{j \in [N]} \left\| \sin\left[\frac{A_j}{2}(\theta_j - \theta'_j)\right] \right\| \leq \sum_{j \in [N]} \|A_j\| |\theta_j - \theta'_j| \tag{6.52}$$

The difference of parametrers $|\theta_j - \theta'_j|$ is always rescaled by the respective $\|A_j\|$; we can interpret this by considering θ_j and A_j as dimensionful quantities, with inverse dimension to each other. We can always redefine units rescaling θ_j and A_j – without loss of generality we choose units for which $\|A_j\| = 1$ (the only assumption being the boundedness of A_j). Under this choice,

$$\langle\psi(\tilde{\boldsymbol{\theta}}(0))|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle = -1 \implies \|\tilde{\boldsymbol{\theta}}(0) - \tilde{\boldsymbol{\theta}}(1)\|_1 \geq 2 \tag{6.53}$$

Thus the parameters need to change (in 1-norm) by at least 2 along the path to achieve the same state and nontivial Berry phase. By the same reasoning, an error on the final parameters $\delta\tilde{\boldsymbol{\theta}}(1)$ bounded by $\|\delta\tilde{\boldsymbol{\theta}}(1)\|_1 < 1$ will not change the argument (i.e. the sign) of the overlap, thus allowing to resolve the correct Berry phase.

6

6.B Bounding the norm of energy derivatives

Suppose we have a variational state parameterized by $\boldsymbol{\theta}$

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle = \prod_{k=0}^{n_p-1} U_k(\theta_k)V_k|0\rangle \tag{6.54}$$

on a n_p-dimensional manifold (within a larger Hilbert space), where the product is assumed to be taken in unitary composition order (right to left). Assume each U_k is generated by anti-hermitian operators iA_k with unit norm

$$U_k(\theta) = e^{iA_k\theta}, \quad A_k = A_k^\dagger, \quad \|A_k\| = 1. \tag{6.55}$$

Given H is an observable of known norm $\|H\|$, define

$$E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle. \tag{6.56}$$

Define the tensors of derivatives,

$$\mathcal{G}_j(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_j}E(\boldsymbol{\theta}) \tag{6.57}$$

$$\mathcal{H}_{jk}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_j\partial\theta_k}E(\boldsymbol{\theta}) \tag{6.58}$$

$$\mathcal{T}_{jkl}(\boldsymbol{\theta}) = \frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}E(\boldsymbol{\theta}) \tag{6.59}$$

$$\tag{6.60}$$

our goal is to bound their (vector-induced) 2-norms $\|\mathcal{G}\|, \|\mathcal{H}\|, \|\mathcal{T}\|$.

We first notice that

$$\frac{\partial}{\partial\theta_j}U(\theta) = \left(\prod_{k=j}^{n_p-1} U_k(\theta_k)V_k\right)iA_j\left(\prod_{k=0}^{j-1} U_k(\theta_k)V_k\right) = i\tilde{A}_jU(\theta) \quad (6.61)$$

where $\|\tilde{A}_j\| = \|A_j\| = 1$, as conjugation by a unitary preserves norm. Using this, we can get expressions for the tensors of derivatives in terms

of commutators of Hermitian operators of known norm

$$\mathcal{G}_j(\boldsymbol{\theta}) = i\langle\psi(\boldsymbol{\theta})|[H, \tilde{A}_j]|\psi(\boldsymbol{\theta})\rangle \tag{6.62}$$

$$\mathcal{H}_{jk}(\boldsymbol{\theta}) = -\langle\psi(\boldsymbol{\theta})|[[H, \tilde{A}_j], \tilde{A}_k]|\psi(\boldsymbol{\theta})\rangle \tag{6.63}$$

$$\mathcal{T}_{jkl}(\boldsymbol{\theta}) = -i\langle\psi(\boldsymbol{\theta})|[[[H, \tilde{A}_j], \tilde{A}_k], \tilde{A}_l]|\psi(\boldsymbol{\theta})\rangle \tag{6.64}$$

A trivial bound on this involves bounding each commutator by its norm, e.g.

$$\|\mathcal{G}(\boldsymbol{\theta})\|^2 = \sum_{k=0}^{n_p+1} \|\langle\psi(\boldsymbol{\theta})|[H, \tilde{A}_j]|\psi(\boldsymbol{\theta})\rangle\|^2 \le 4n_p\|H\|^2. \tag{6.65}$$

It is an open question wether we can improve on this bound.

We can get a similar result for the derivatives with respect to the orbital rotations, considering the reparametrization described in Sec. 6.5.3. We call $|\psi\rangle$ the PQC ansatz state in the full (active + core + virtual) space, padded with virtual (core) registers of qubits in the state $|0\rangle$ ($|1\rangle$). We drop explicit dependence on C and $\boldsymbol{\theta}$, and we make explicit the differential rotation parameters κ. The cost function is then

$$E(\kappa) = \langle\psi| e^{\sum_{pq} \kappa_{pq} E_{pq}} H e^{-\sum_{rs} \kappa_{rs} E_{rs}} |\psi\rangle, \quad E_{pq} = c_{p,\uparrow}^\dagger c_{q,\uparrow} + c_{p,\downarrow}^\dagger c_{q,\downarrow}, \tag{6.66}$$

where E_{pq} is the generator of a spin-adapted orbital rotation. Its derivatives at $\kappa = 0$ [note that the index pairs (pq) are collected in one index for the purpose of rotating higher order derivatives] are easily calculated to be

$$\mathcal{G}_{(pq)}(\kappa = 0) = \langle\psi| [H, E_{pq}] |\psi\rangle, \tag{6.67}$$

$$\mathcal{H}_{(pq),(rs)}(\kappa = 0) = \langle\psi| [[H, E_{pq}], E_{rs}] |\psi\rangle, \tag{6.68}$$

$$\mathcal{T}_{(pq),(rs),(tu)}(\kappa = 0) = \langle\psi| [[[H, E_{pq}], E_{rs}], E_{tu}] |\psi\rangle. \tag{6.69}$$

Observing that $\|E_{pq}\| = 2$ we obtain the same result as above (up to constant factors 2, 4, 8 respectively, coming from this norm).

6.C Analytical orbital gradient and Hessian

In this section, we expand on the estimation of analytic orbital gradient [right block of the vector in Eq. (6.39)] and orbital-orbital and orbital-circuit Hessian [bottom right and top right blocks of the matrix in Eq. (6.40), respectively]. We show how after the reparametrization

$C \leftarrow C \cdot e^{-\kappa}$, the κ-derivatives of the cost function $E(C \cdot e^{-\kappa}, \boldsymbol{\theta})$ can be expressed as a linear function of the 1- and 2-electron RDM, and the mixed Hessian $\nabla_{\kappa} \nabla_{\boldsymbol{\theta}} E(C \cdot e^{-\kappa}, \boldsymbol{\theta})$ can be expressed in terms of $\boldsymbol{\theta}$-derivatives of the same RDM.

We first define the 1- and 2-electron reduced density matrix (RDM) in the spin-restricted formalism

$$\gamma_{pq}(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta})| E_{pq} |\psi(\boldsymbol{\theta})\rangle \tag{6.70}$$

$$\Gamma_{pqrs}(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta})| e_{pqrs} |\psi(\boldsymbol{\theta})\rangle , \tag{6.71}$$

where $E_{pq} = \sum_{\sigma} c_{p\sigma}^{\dagger} c_{q\sigma}$ and $e_{pqrs} = \sum_{\sigma\tau} c_{p\sigma}^{\dagger} c_{r\tau}^{\dagger} c_{s\tau} c_{q\sigma} = E_{pq}E_{rs} - \delta_{qr}E_{ps}$. Here, p, q, r, s are meant to be general indices (either occupied, active or virtual), where the state $|\psi(\boldsymbol{\theta})\rangle$ is to be intended as padded by two registers of qubits in the $|0\rangle^{\otimes 2N_{\mathrm{V}}}$ ($|1\rangle^{\otimes 2N_{\mathrm{O}}}$) state for the virtual (occupied) orbitals. In the molecular orbital basis defined by C [orbitals in Eq. (6.34)], we can write the Hamiltonian as

$$H = \sum_{pq} h_{pq} E_{pq} + \frac{1}{2} \sum_{pqrs} g_{pqrs} e_{pqrs} \tag{6.72}$$

where h_{pq} and g_{pqrs} are the one- and two-electron integrals (with spatial orbital indices p, q, r, s ordered according to the chemists' convention), and they implicitly depend on C through the MOs. The expectation value of the Hamiltonian can then be written as a contraction of the integrals with the RDM,

$$E(C, \boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta})| H(C) |\psi(\boldsymbol{\theta})\rangle = \sum_{pq} [h(C)]_{pq} \gamma_{pq} + \frac{1}{2} \sum_{pqrs} [g(C)]_{pqrs} \Gamma_{pqrs} \tag{6.73}$$

where we made explicit the dependence on C.

To derive analytical orbital rotation derivatives, we closely follow Ref. [219]. We start by separating the dependence on κ of the reparametrized cost function $E(C \cdot e^{-\kappa}, \boldsymbol{\theta})$, by using the equivalent state transformation formalism provided by Thouless theorem [254]

$$E(C \cdot e^{-\kappa}, \boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta})| H(C \cdot e^{-\kappa}) |\psi(\boldsymbol{\theta})\rangle = \langle \psi(\boldsymbol{\theta})| e^{\hat{\kappa}} H(C) e^{-\hat{\kappa}} |\psi(\boldsymbol{\theta})\rangle \tag{6.74}$$

where $\hat{\kappa} = \sum_{pq} \kappa_{pq} E_{pq}$ is the operator that generates a unitary on the Hilbert state space equivalent to the orbital rotation. We know that the

rotations where p, q are both virtual indices form a redundant subgroup, so we can freeze the corresponding $\kappa_{pq} = 0$; the same is true for p, q both core space indices. In other terms, $\kappa_{pq} = 0$ if $p, q \in V$ or $p, q \in O$, with V and O the sets of virtual and core indices. The remaining elements of κ_{pq} satisfy $\kappa_{pq} = -\kappa_{qp}$. We define a vector of unique non-redundant orbital rotation parameters

$$\boldsymbol{\kappa} = \{\kappa_{pq}, \forall p \in O \cup A, \forall q \in A \cup V : q > p\}, \tag{6.75}$$

and we redefine the cost function with respect to this vector,

$$E(C, \boldsymbol{\kappa}, \boldsymbol{\theta}) \equiv E(C \cdot e^{-\kappa}, \boldsymbol{\theta}). \tag{6.76}$$

We are interested in the derivative with respect to this vector; we can always switch from the matrix κ to the unraveled vector of unique non-redundant parameters $\boldsymbol{\kappa}$, and vice versa. By comparing the Taylor series in κ with the Baker-Campbell-Hausdorff expansion:

$$E(\boldsymbol{\theta}, \boldsymbol{\kappa}) = \langle\psi(\boldsymbol{\theta})| H |\psi(\boldsymbol{\theta})\rangle + \langle\psi(\boldsymbol{\theta})| [\hat{\kappa}, H] |\psi(\boldsymbol{\theta})\rangle + \frac{1}{2} \langle\psi(\boldsymbol{\theta})| [\hat{\kappa}, [\hat{\kappa}, H]] |\psi(\boldsymbol{\theta})\rangle + \dots \tag{6.77}$$

One can readily verify that the analytical orbital derivatives at $\kappa_{pq} = 0$ are given by:

$$[\nabla_{\boldsymbol{\kappa}} E]_{pq} := \frac{\partial E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial \kappa_{pq}}\bigg|_{\kappa=0} = \langle\psi(\boldsymbol{\theta})| [E_{pq}^-, H] |\psi(\boldsymbol{\theta})\rangle \tag{6.78}$$

$$[\nabla_{\boldsymbol{\kappa}}^2 E]_{pqrs} := \frac{\partial^2 E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial \kappa_{pq} \partial \kappa_{rs}}\bigg|_{\kappa=0} = \frac{1}{2}(1 + P_{pq,rs}) \langle\psi(\boldsymbol{\theta})| [E_{pq}^-, [E_{rs}^-, H]] |\psi(\boldsymbol{\theta})\rangle \tag{6.79}$$

where $P_{pq,rs}$ permutes the pair of indices pq with rs. The calculation of the commutators in Eq. (6.78) and (6.79) can be found in common quantum chemistry textbooks [219], and they all one- or two-body operators; thus their expectation value can be written as a linear form in the RDM (γ, Γ). The gradient evaluates to

$$[\nabla_{\boldsymbol{\kappa}} E]_{pq} = 2(F_{pq}(\boldsymbol{\theta}) - F_{qp}(\boldsymbol{\theta})) \tag{6.80}$$

6

where F is the generalized Fock matrix,

$$F_{pq}(\boldsymbol{\theta}) = \sum_m \gamma_{pm}(\boldsymbol{\theta})h_{qm} + \sum_{mnk} \Gamma_{pmnk}(\boldsymbol{\theta})g_{qmnk} \qquad (6.81)$$

The Hessian evaluates to

$$[\nabla_{\boldsymbol{\kappa}}^2 E]_{pqrs} = (1 - P_{pq})(1 - P_{rs}) \left[2\gamma_{pr}h_{qs} - (F_{pr} + F_{rp})\delta_{qs} + 2Y_{pqrs}\right], \qquad (6.82)$$

where we introduced

$$Y_{pqrs} = \sum_{mn} \Gamma_{pmrn}g_{qmns} + \Gamma_{pmnr}g_{qmns} + \Gamma_{prmn}g_{qsmn} \qquad (6.83)$$

and dropped the explicit dependence on $\boldsymbol{\theta}$. For the composite Hessian, we simply take the gradient of Eq. (6.80) with respect to $\boldsymbol{\theta}$, by using the chain rule

$$[\nabla_{\boldsymbol{\kappa}}\nabla_{\boldsymbol{\theta}}E]_{pq} := \left.\frac{\partial^2 E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial \kappa_{pq}\partial\boldsymbol{\theta}}\right|_{\kappa=0} = 2\left(\frac{\partial F_{pq}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} - \frac{\partial F_{qp}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) \qquad (6.84)$$

where

$$\frac{\partial F_{pq}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \sum_m \frac{\partial \gamma_{pm}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}h_{qm} + \sum_{mnk} \frac{\partial \Gamma_{pmnk}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}g_{qmnk}. \qquad (6.85)$$

Thus, once we have the derivatives of the 1- and 2-RDM to sufficient precision, we can evaluate the orbital gradient, Hessian and composite Hessian analytically, recovering all terms in Eq. (6.39) and Eq. (6.40) without any additional quantum cost.

6.D Bounding the cumulative error due to Newton-Raphson updates

In this section, we prove that using a single Newton-Raphson (NR) parameter update per Δt-step is sufficient to achieve an error on the estimate of the minimizer scaling as $O(\Delta t^2)$ after any number of Δt-steps, as long as the cost function is strongly-convex at the minimum and Δt is small enough. First, we recall sufficient conditions for quadratic convergence of the Newton-Raphson step. We then use these to bound the error of a single NR-step when the cost function is changed from $E(t, \boldsymbol{\theta}) \to E(t + \Delta t, \boldsymbol{\theta})$.

We translate this into an upper bound on Δt which guarantees the error stays bounded throughout the optimization path. Finally, we show that we can allow a sufficiently small additive error on the Newton-Raphson update and retain the bounded error throughout the path.

Quadratic convergence of NR — Consider a cost function $E(\boldsymbol{\theta})$ with gradient $\mathcal{G}_j = \frac{\partial E}{\partial \theta_j}$ and Hessian $\mathcal{H}_{jk} = \frac{\partial E}{\partial \theta_j \partial \theta_k}$, and an initial guess of a minimizer $\boldsymbol{\theta}^{(0)}$. The Newton-Raphson step prescribes the update $\boldsymbol{\theta}^{(0)} \mapsto \boldsymbol{\theta}^{\mathrm{NR}} = \boldsymbol{\theta}^{(0)} + d\boldsymbol{\theta}^{\mathrm{NR}}$ with $d\boldsymbol{\theta}^{\mathrm{NR}} = \mathcal{H}^{-1}(\boldsymbol{\theta}^{(0)})\mathcal{G}(\boldsymbol{\theta}^{(0)})$. Theorem 3.5 from Nocedal and Wright [231] gives sufficient conditions under which quadratic convergence of the NR update is guaranteed. We simplify these conditions, and obtain the following

Theorem 6.1
Consider a cost function $E(\boldsymbol{\theta})$ with Lipschitz-continuous Hessian $\|\mathcal{H}(\boldsymbol{\theta}) - \mathcal{H}(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\| \leq L\|\delta\boldsymbol{\theta}\|$ and a local minimizer $\boldsymbol{\theta}^*$ with positive convexity $m := \|\mathcal{H}^{-1}(\boldsymbol{\theta}^*)\|^{-1} > 0$. Given an initial guess $\boldsymbol{\theta}^{(0)}$ which is close enough to the minimum, i.e.

$$\|\boldsymbol{\theta}^{(0)} - \boldsymbol{\theta}^*\| \leq \frac{m}{4L}, \tag{6.86}$$

the NR update will converge quadratically towards the minimum with

$$\|\boldsymbol{\theta}^{NR} - \boldsymbol{\theta}^*\| \leq \frac{L}{m}\|\boldsymbol{\theta}^{(0)} - \boldsymbol{\theta}^*\|^2. \tag{6.87}$$

To prove this, we only need to show that a strong convexity condition is satisfied within a r-ball centered in $\boldsymbol{\theta}^*$ including all close-enough possible initial guesses ($r = \frac{m}{4L}$), i.e.

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})\| \leq 2m^{-1}, \quad \forall\|\delta\boldsymbol{\theta}\| \leq \frac{m}{4L}. \tag{6.88}$$

To prove this we expand $\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})$ using Taylor's theorem,

$$\exists 0 < s < 1: \quad \mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta}) = \mathcal{H}^{-1}(\boldsymbol{\theta}^*) + \delta\boldsymbol{\theta} \cdot \frac{\partial \mathcal{H}^{-1}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta}) \tag{6.89}$$

$$\frac{\partial \mathcal{H}^{-1}}{\partial \boldsymbol{\theta}} = -\mathcal{H}^{-1}\frac{\partial \mathcal{H}}{\partial \boldsymbol{\theta}}\mathcal{H}^{-1} \tag{6.90}$$

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})\| \leq \|\mathcal{H}^{-1}(\boldsymbol{\theta}^*)\| + \|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\|^2 L\|\delta\boldsymbol{\theta}\| \tag{6.91}$$

$$\leq m^{-1} + \|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\|^2 \frac{m}{4}, \tag{6.92}$$

where we used the Lipschitz constant L as a bound on the derivative of

6

the Hessian. This last condition holds if

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\| \leq 2m^{-1}. \tag{6.93}$$

As we can choose $s < 1$ and the result is clearly true at $\delta\boldsymbol{\theta} = 0$, the result holds recursively.

Single NR update with a changing cost function — We now consider a family of cost functions $E(t, \boldsymbol{\theta})$ continuously parameterized by t. Suppose we have an approximation $\tilde{\boldsymbol{\theta}}_t$ of the minimizer $\boldsymbol{\theta}_t^*$ of $E(t, \boldsymbol{\theta})$, with error $\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\|$. In each step of our method (Algorithm 6.2), we shift the cost function $E(t, \boldsymbol{\theta}) \to E(t + \Delta t, \boldsymbol{\theta})$ by Δt and we use the current minimizer estimate $\tilde{\boldsymbol{\theta}}_t$ as initial guess for the next step; $\boldsymbol{\theta}_{t+\Delta t}^{(0)} = \tilde{\boldsymbol{\theta}}_t$. We can bound the error of this initial guess by the triangle inequality,

$$\|\boldsymbol{\theta}_{t+\Delta t}^{(0)} - \boldsymbol{\theta}_{t+\Delta t}^*\| \leq \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\| + \|\boldsymbol{\theta}_{t+\Delta t}^* - \boldsymbol{\theta}_t^*\|. \tag{6.94}$$

While the first term is the (given) error on the estimate, the second can be obtained by taking the total t-derivative of the minimum condition $\mathcal{G}(t, \boldsymbol{\theta}^*(t)) = 0$, and applying Taylor's theorem

$$\exists \tau \in [t, t+\Delta t]: \quad \|\boldsymbol{\theta}_{t+\Delta t}^* - \boldsymbol{\theta}_t^*\| = \left\| \frac{d\boldsymbol{\theta}_t^*}{dt} \Big|_{t=\tau} \right\| \Delta t = \|\mathcal{H}^{-1}(\tau, \boldsymbol{\theta}_\tau^*)\dot{\mathcal{G}}(\tau, \boldsymbol{\theta}_\tau^*)\|\Delta t \tag{6.95}$$

with $\dot{\mathcal{G}} = \nabla \frac{\partial}{\partial t} E(\boldsymbol{\theta}, t)$.

We now assume that the convexity at the minimum is bounded from below throughout the whole t-path by a constant $m \geq 0$,

$$m(t, \boldsymbol{\theta}_t^*) := \|\mathcal{H}^{-1}(t, \boldsymbol{\theta}_t^*)\|^{-1} > m \quad \forall t \in [0, 1], \tag{6.96}$$

and that the gradient of the change is never larger than $\dot{\mathcal{G}}_{\max}$. (while the first assumption imposes the nontrivial condition of strong convexity at the minimum, the second is always granted for cost functions from a continuous family of bounded Hamiltonians). We can then write

$$\|\boldsymbol{\theta}_{t+\Delta t}^{(0)} - \boldsymbol{\theta}_{t+\Delta t}^*\| \leq \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\| + m^{-1}\dot{\mathcal{G}}_{\max}\Delta t. \tag{6.97}$$

To ensure this initial guess is within the quadratic convergence region of the NR step, we require

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\| + m^{-1}\dot{\mathcal{G}}_{\max}\Delta t \leq \frac{m}{4L}, \tag{6.98}$$

choosing an $\alpha, \beta \in (0, 1]$ this condition can be written as

$$\Delta t = \frac{m^2}{4L\dot{\mathcal{G}}_{\max}}\alpha\beta, \tag{6.99}$$

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\| = (1 - \alpha)\beta\frac{m}{4L}. \tag{6.100}$$

This allows to apply Theorem 6.1, and bound the error after a single NR step,

$$\|\tilde{\boldsymbol{\theta}}_{t+\Delta t} - \boldsymbol{\theta}_{t+\Delta t}^*\| \leq \frac{L}{m}\left[\frac{\beta m}{4L}\right]^2 = \beta^2\frac{m}{16L}. \tag{6.101}$$

Multiple steps — We want to ensure the error on the minimizer estimate remains bounded for t taking subsequent values is $[0, \Delta t, 2\,\Delta t, ..., 1]$, while taking a single NR step at a time. We can do this by imposing the error after each step [Eq. (6.101)] is not larger than the error on the previous step estimate,

$$\beta^2\frac{m}{16L} \leq (1 - \alpha)\beta\frac{m}{4L}. \tag{6.102}$$

This is granted for any $\beta \in (0, 1]$ by choosing $\alpha = 1 - \frac{\beta}{4}$. The maximum $\Delta t = \frac{3}{4}\frac{m^2}{4L\dot{\mathcal{G}}_{\max}}$ is achieved by picking $\beta = 1$, and yields an error bounded by the constant $\frac{m}{16L}$.

Allowing an additive error — To account for sampling noise, it is useful to consider an additive error of magnitude $\sigma_{\boldsymbol{\theta}}$ on the estimate $\tilde{\boldsymbol{\theta}}_t$ of the minimizer $\boldsymbol{\theta}_t^*$ at each t-point, modifying Eq. (6.100) into

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\| = (1 - \alpha)\beta\frac{m}{4L} + \sigma_{\boldsymbol{\theta}}. \tag{6.103}$$

This yields the condition

$$\frac{L}{m}\left[\frac{\beta m}{4L} + \sigma_{\boldsymbol{\theta}}\right]^2 \leq (1 - \alpha)\beta\frac{m}{4L}. \tag{6.104}$$

If we define γ by $\sigma_{\boldsymbol{\theta}} = \frac{\gamma\beta m}{4L}$, we can write

$$(1 + \gamma)^2\frac{m}{16L}\beta^2 \leq (1 - \alpha)\beta\frac{m}{4L}, \tag{6.105}$$

6

which is saturated by $\beta = 4\frac{1-\alpha}{(1+\gamma)^2}$. We then get

$$\Delta t = \frac{m^2}{L\dot{\mathcal{G}}_{\max}} \frac{\alpha(1-\alpha)}{(1+\gamma)^2}, \tag{6.106}$$

which is maximised (while keeping $\beta \leq 1$) by $\alpha = \max[\frac{1}{2}, 1 - \frac{(1+\gamma)^2}{4}]$. The allowed additive noise is then

$$\sigma_{\boldsymbol{\theta}} = \frac{\gamma}{(1+\gamma)^2} \frac{m}{L}(1-\alpha) = \frac{\gamma m}{2L} \min[\frac{1}{2}, \frac{1}{(1+\gamma)^2}], \tag{6.107}$$

maximised for the choice $\gamma = \sqrt{2} - 1$ yielding

$$\sigma_{\boldsymbol{\theta}} = \frac{\sqrt{2}-1}{4} \frac{m}{L}, \quad \Delta t = \frac{m^2}{8L\dot{\mathcal{G}}_{\max}}. \tag{6.108}$$

6.E Bounding the sampling cost

We call $\sigma_{\mathcal{G}}^2$ and $\sigma_{\mathcal{H}}^2$ the variances of each element of the gradient and Hessian respectively, due to sampling noise. To compute the error on the parameter updates, we propagate these variances through the definition of the NR update Eq. (6.12). The first-order differential change (here denoted with δ) of the NR update $d\boldsymbol{\theta}^{\mathrm{NR}}$ with respect to changes in the gradient and Hessian is

$$\delta[d\boldsymbol{\theta}^{\mathrm{NR}}] = \mathcal{H}^{-1} \cdot [-\delta\mathcal{G} + \delta\mathcal{H} \cdot d\boldsymbol{\theta}^{\mathrm{NR}}], \tag{6.109}$$

where we use the ordered matrix-product notation, with vectors in boldface. When $\delta\mathcal{G}$ and $\delta\mathcal{H}$ are the random variables representing the errors on the gradient and Hessian, the expected mean square error on the NR update defining the norm of the covariance matrix

$$\left\| \mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}] \right\| := \mathbb{E}\left[\|\delta[d\boldsymbol{\theta}^{\mathrm{NR}}]\|^2\right] = \mathbb{E}\left[\|\mathcal{H}^{-1}\cdot\delta\mathcal{G}\|^2\right] + \mathbb{E}\left[\|\mathcal{H}^{-1}\cdot\delta\mathcal{H}\cdot d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right], \tag{6.110}$$

where we used the zero-average property of $\delta\mathcal{G}$ and $\delta\mathcal{H}$ to drop the expectation values of mixed terms. This can further be bounded as

$$\mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}] \leq \|\mathcal{H}^{-1}\|^2 \mathbb{E}\left[\|\delta\mathcal{G}\|^2\right] + \|\mathcal{H}^{-1}\|^2 \mathbb{E}\left[\|\delta\mathcal{H}\|^2\right] \|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2. \tag{6.111}$$

Assuming the same variance $\sigma_{\mathcal{G}}^2$ on each of the n_p elements $\mathcal{G}_j$ of the gradient, we get

$$\mathbb{E}\left[\|\delta\mathcal{G}\|^2\right] = n_p\,\sigma_{\mathcal{G}}^2. \tag{6.112}$$

As the Hessian is a random real symmetric matrix with i.i.d. elements, each with a variance $\sigma_{\mathcal{H}}$, we can invoke Wigner's semicircle law [232] to bound the spectral norm as

$$\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right] \leq (\sqrt{n_p}\,\sigma_{\mathcal{H}})^2. \tag{6.113}$$

Combining these with the strong convexity bound $\|\mathcal{H}^{-1}\| \leq m^{-1}$, we get

$$\mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}] \leq m^{-2}\left[n_p\,\sigma_{\mathcal{G}}^2 + n_p\,\sigma_{\mathcal{H}}^2\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right] \tag{6.114}$$

The calculations in Appendix 6.D conclude that, for each Δt-step, we can afford an additive error on the NR update of at most $\sigma_{\boldsymbol{\theta}} \leq \frac{\gamma}{4}\frac{m}{L}$ with $\gamma = \sqrt{2} - 1$. Comparing this result to the variance just calculated, we can formulate the requirement on the variance

$$m^{-2}n_p\left[\sigma_{\mathcal{G}}^2 + \sigma_{\mathcal{H}}^2\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right] \ll \frac{\gamma^2}{16}\frac{m^2}{L^2}. \tag{6.115}$$

This gives conditions on the elementary sampling variances

$$\sigma_{\mathcal{G}}^2 \ll \frac{\gamma^2}{16}\frac{m^4}{L^2 n_p} \tag{6.116}$$

$$\sigma_{\mathcal{H}}^2 \ll \frac{\gamma^2}{16}\frac{m^4}{L^2 n_p\,\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2} < \frac{\gamma^2}{16}\frac{m^6}{L^2 n_p\|\dot{\mathcal{G}}_{\mathrm{max}}\|^2\Delta t^2} \tag{6.117}$$

To recast this bound in terms of variables of the problem, we use the following relations derived in Appendix 6.B: $L = \max\|\mathcal{T}\| < n_p^{3/2}\|H\|$ (the norm of the third derivative tensor $\mathcal{T}$ is bounded by $n_p^{3/2}$ times by its infinity norm), $\|\dot{\mathcal{G}}_{\mathrm{max}}\|\Delta t < \sqrt{n_p}\|\frac{dH}{dt}\|\Delta t \approx \sqrt{n_p}\|H\|$ (same infinity norm bound). Furthermore, we assume the convexity is larger than the ground state gap, $m > \Delta$; this holds if the ansatz approximates the ground state well enough, and changes in any ansatz parameter θ_k introduce a different excited state. Substituting these relations we obtain

$$\sigma_{\mathcal{G}}^2 \ll 0.01\frac{\Delta^4}{\|H\|^2 n_p^4}, \tag{6.118}$$

$$\sigma_{\mathcal{H}}^2 \ll 0.01\frac{\Delta^6}{\|H\|^4 n_p^5}. \tag{6.119}$$

The number of total required shots to sample the Hessian (gradient) for all the N steps will thus scale as $\sigma_{\mathcal{H}}^{-2}\Delta t^{-1}$ ($\sigma_{\mathcal{G}}^{-2}\Delta t^{-1}$). Picking the maximal

$\Delta t = \frac{m^2}{8L\dot{\mathcal{G}}_{\max}}$, and considering only the dominant term (relative to sampling the Hessian) we can write

$$\#\text{shots} \propto n_p^7 \frac{\|H\|^7 \|\frac{dH}{dt}\|}{\Delta^8} \qquad (6.120)$$

Bibliography

[1] P. W. ANDERSON. More Is Different. *Science*, **177**, 393–396 (1972).

[2] R. P. FEYNMAN. Simulating physics with computers. *International Journal of Theoretical Physics*, **21**, 467–488 (1982).

[3] F. ARUTE, K. ARYA, R. BABBUSH, D. BACON, J. C. BARDIN, R. BARENDS ET AL. Quantum supremacy using a programmable superconducting processor. *Nature*, **574**, 505–510 (2019).

[4] Y. WU, W.-S. BAO, S. CAO, F. CHEN, M.-C. CHEN, X. CHEN ET AL. Strong Quantum Computational Advantage Using a Superconducting Quantum Processor. *Physical Review Letters*, **127**, 180501 (2021).

[5] Q. ZHU, S. CAO, F. CHEN, M.-C. CHEN, X. CHEN, T.-H. CHUNG ET AL. Quantum Computational Advantage via 60-Qubit 24-Cycle Random Circuit Sampling, arXiv:2109.03494 (2021).

[6] L. S. MADSEN, F. LAUDENBACH, M. F. ASKARANI, F. RORTAIS, T. VINCENT, J. F. F. BULMER ET AL. Quantum computational advantage with a programmable photonic processor. *Nature*, **606**, 75–81 (2022).

[7] H.-S. ZHONG, Y.-H. DENG, J. QIN, H. WANG, M.-C. CHEN, L.-C. PENG ET AL. Phase-Programmable Gaussian Boson Sampling Using Stimulated Squeezed Light. *Physical Review Letters*, **127**, 180502 (2021).

[8] P. A. M. Dirac. *The Principles of Quantum Mechanics.* Clarendon Press (1981). ISBN 978-0-19-852011-5.

[9] A. S. Holevo. Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel. *Problemy Peredachi Informatsii*, **9**, 3–11 (1973).

[10] M. Freedman, M. Larsen & Z. Wang. A modular functor which is universal for quantum computation, arXiv:quant-ph/0001108 (2000).

[11] D. Aharonov, M. Ben-Or, R. Impagliazzo & N. Nisan. Limitations of Noisy Reversible Computation, arXiv:quant-ph/9611028 (1996).

[12] D. Aharonov. A Quantum to Classical Phase Transition in Noisy Quantum Computers. *Physical Review A*, **62**, 062311 (2000).

[13] A. Kitaev. Anyons in an exactly solved model and beyond. *Annals of Physics*, **321**, 2–111 (2006).

[14] C. Gidney & M. Ekerå. How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. *Quantum*, **5**, 433 (2021).

[15] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker & M. Troyer. Elucidating reaction mechanisms on quantum computers. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 7555–7560 (2017).

[16] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute et al. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, **614**, 676–681 (2023).

[17] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas et al. Fault-tolerant operation of a quantum error-correction code. *ArXiv:2009.11482* (2020).

[18] J. Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, **2**, 79 (2018).

[19] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik & J. L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, **5**, 4213 (2014).

[20] K. G. H. Vollbrecht & J. I. Cirac. Quantum Simulators, Continuous-Time Automata, and Translationally Invariant Systems. *Physical Review Letters*, **100**, 010501 (2008).

[21] S. Lloyd. Universal Quantum Simulators. *Science*, **273**, 1073–1078 (1996).

[22] H. F. Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, **10**, 545–545 (1959).

[23] M. Suzuki. Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems. *Communications in Mathematical Physics*, **51**, 183–190 (1976).

[24] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin & F. Verstraete. Quantum Metropolis sampling. *Nature*, **471**, 87–90 (2011).

[25] G. Verdon, J. Marks, S. Nanda, S. Leichenauer & J. Hidary. Quantum Hamiltonian-Based Models and the Variational Quantum Thermalizer Algorithm, arXiv:1910.02071 (2019).

[26] A. N. Chowdhury, G. H. Low & N. Wiebe. A Variational Quantum Algorithm for Preparing Quantum Gibbs States, arXiv:2002.00055 (2020).

[27] S. Lu, M. C. Bañuls & J. I. Cirac. Algorithms for Quantum Simulation at Finite Energies. *PRX Quantum*, **2**, 020321 (2021).

[28] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandao & G. K.-L. Chan. Quantum imaginary time evolution, quantum lanczos, and quantum thermal averaging. *arXiv:1901.07653* (2019).

[29] S. McArdle, T. Jones, S. Endo, Y. Li, S. Benjamin & X. Yuan. Variational ansatz-based quantum simulation of imaginary time evolution. *npj Quantum Information*, **5**, 75 (2019).

[30] J. Kempe, A. Kitaev & O. Regev. The Complexity of the Local Hamiltonian Problem. *SIAM Journal on Computing*, **35**, 1070–1097 (2006).

[31] D. AHARONOV, D. GOTTESMAN, S. IRANI & J. KEMPE. The Power of Quantum Systems on a Line. *Communications in Mathematical Physics*, **287**, 41–65 (2009).

[32] B.-X. ZHENG, C.-M. CHUNG, P. CORBOZ, G. EHLERS, M.-P. QIN, R. M. NOACK, H. SHI, S. R. WHITE, S. ZHANG & G. K.-L. CHAN. Stripe order in the underdoped region of the two-dimensional Hubbard model. *Science*, **358**, 1155–1160 (2017).

[33] A. BAIARDI & M. REIHER. The Density Matrix Renormalization Group in Chemistry and Molecular Physics: Recent Developments and New Challenges. *The Journal of Chemical Physics*, **152**, 040903 (2020).

[34] L.-H. FRAHM & D. PFANNKUCHE. Ultrafast ab Initio Quantum Chemistry Using Matrix Product States. *Journal of Chemical Theory and Computation*, **15**, 2154–2165 (2019).

[35] P. HOHENBERG & W. KOHN. Inhomogeneous Electron Gas. *Physical Review*, **136**, B864–B871 (1964).

[36] W. KOHN. Nobel Lecture: Electronic structure of matter—wave functions and density functionals. *Reviews of Modern Physics*, **71**, 1253–1266 (1999).

[37] E. Y. LOH, J. E. GUBERNATIS, R. T. SCALETTAR, S. R. WHITE, D. J. SCALAPINO & R. L. SUGAR. Sign problem in the numerical simulation of many-electron systems. *Physical Review B*, **41**, 9301–9307 (1990).

[38] M. TROYER & U.-J. WIESE. Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations. *Physical Review Letters*, **94**, 170201 (2005).

[39] E. CAMPBELL. Random Compiler for Fast Hamiltonian Simulation. *Physical Review Letters*, **123**, 070503 (2019).

[40] M. C. TRAN, A. Y. GUO, Y. SU, J. R. GARRISON, Z. ELDREDGE, M. FOSS-FEIG, A. M. CHILDS & A. V. GORSHKOV. Locality and digital quantum simulation of power-law interactions. *Physical Review X*, **9**, 031006 (2019).

[41] A. M. CHILDS, Y. SU, M. C. TRAN, N. WIEBE & S. ZHU. A Theory of Trotter Error. *ArXiv* (2019).

[42] E. H. Lieb & D. W. Robinson. The finite group velocity of quantum spin systems. *Communications in Mathematical Physics*, **28**, 251–257 (1972).

[43] A. M. Childs & N. Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Information and Computation*, **12**, 901–924 (2012).

[44] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari & R. D. Somma. Exponential improvement in precision for simulating sparse Hamiltonians. *Proc. ACM STOC*, pages 283–282 (2014).

[45] D. W. Berry, A. M. Childs & R. Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 792–809 (2015).

[46] G. H. Low & I. L. Chuang. Hamiltonian Simulation by Qubitization. *Quantum*, **3**, 163 (2019).

[47] G. H. Low & I. L. Chuang. Optimal Hamiltonian simulation by quantum signal processing. *Physical review letters*, **118**, 010501 (2017).

[48] A. Gilyén, Y. Su, G. H. Low & N. Wiebe. Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 193–204. ACM, New York, NY, USA (2019). ISBN 978-1-4503-6705-9.

[49] E. Farhi, J. Goldstone, S. Gutmann & M. Sipser. Quantum Computation by Adiabatic Evolution, arXiv:quant-ph/0001106 (2000).

[50] M. Born & V. Fock. Beweis des Adiabatensatzes. *Zeitschrift für Physik*, **51**, 165–180 (1928).

[51] S. Jansen, M.-B. Ruskai & R. Seiler. Bounds for the adiabatic approximation with applications to quantum computation. *Journal of Mathematical Physics*, **48**, 102111 (2007).

[52] M. H. S. Amin. Consistency of the adiabatic theorem. *Physical Review Letters*, **102**, 220401 (2009).

[53] N. WIEBE & N. S. BABCOCK. Improved error-scaling for adiabatic quantum evolutions. *New Journal of Physics*, **14**, 013024 (2012).

[54] D. WECKER, M. B. HASTINGS & M. TROYER. Progress towards practical quantum variational algorithms. *Physical Review A*, **92**, 042303 (2015).

[55] E. FARHI, J. GOLDSTONE & S. GUTMANN. A Quantum Approximate Optimization Algorithm, arXiv:1411.4028 (2014).

[56] S. HADFIELD, Z. WANG, B. O'GORMAN, E. G. RIEFFEL, D. VENTURELLI & R. BISWAS. From the Quantum Approximate Optimization Algorithm to a Quantum Alternating Operator Ansatz. *Algorithms*, **12**, 34 (2019).

[57] G.-L. R. ANSELMETTI, D. WIERICHS, C. GOGOLIN & R. M. PARRISH. Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems. *New Journal of Physics*, **23**, 113010 (2021).

[58] H. R. GRIMSLEY, S. E. ECONOMOU, E. BARNES & N. J. MAYHALL. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications*, **10**, 3007 (2019).

[59] A. KANDALA, A. MEZZACAPO, K. TEMME, M. TAKITA, M. BRINK, J. M. CHOW & J. M. GAMBETTA. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, **549**, 242–246 (2017).

[60] J. VON NEUMANN. *Mathematical Foundations of Quantum Mechanics.* Princeton University Press (1955). ISBN 978-0-691-02893-4.

[61] A. Y. KITAEV. Quantum measurements and the Abelian Stabilizer Problem, arXiv:quant-ph/9511026 (1995).

[62] M. NIELSEN & I. CHUANG. *Quantum Computation and Quantum Information.* Cambridge Series on Information and the Natural Sciences. Cambridge University Press (2000). ISBN 978-0-521-63503-5.

[63] M. BORN & R. OPPENHEIMER. Zur Quantentheorie der Molekeln. *Annalen der Physik*, **389**, 457–484 (1927).

[64] B. BAUER, S. BRAVYI, M. MOTTA & G. K.-L. CHAN. Quantum algorithms for quantum chemistry and quantum materials science. *ArXiv:2001.03685* (2020).

[65] S. McArdle, S. Endo, A. Aspuru-Guzik, S. Benjamin & X. Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, **92**, 015003 (2020).

[66] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley & R. Babbush. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Information*, **7**, 1–9 (2021).

[67] X. Bonet-Monroig, R. Babbush & T. E. O'Brien. Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States. *Physical Review X*, **10**, 031064 (2020).

[68] V. Verteletskyi, T.-C. Yen & A. F. Izmaylov. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *Journal of Chemical Physics*, **152**, 224109 (2020).

[69] T. E. O'Brien, G. Anselmetti, F. Gkritis, V. E. Elfving, S. Polla, W. J. Huggins et al. Purification-based quantum error mitigation of pair-correlated electron simulations. *Nature Physics*, **19**, 1787–1792 (2023).

[70] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love & M. Head-Gordon. Simulated Quantum Computation of Molecular Energies. *Science*, **309**, 1704–1707 (2005).

[71] D. W. Berry, C. Gidney, M. Motta, J. R. McClean & R. Babbush. Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization. *Quantum*, **3** (2019).

[72] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler & H. Neven. Encoding Electronic Spectra in Quantum Circuits with Linear T Complexity. *Physical Review X*, **8**, 041015 (2018).

[73] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler & M. Troyer. Quantum computing enhanced computational catalysis. *Physical Review Research*, **3**, 033055 (2021).

[74] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe & R. Babbush. Even more efficient quantum computations of chemistry through tensor hypercontraction. *PRX Quantum*, **2**, 030305 (2021).

[75] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar et al. Is there evidence for exponential quantum advantage in quantum chemistry?, arXiv:2208.02199 (2022).

[76] Y. Su, D. W. Berry, N. Wiebe, N. Rubin & R. Babbush. Fault-Tolerant Quantum Simulations of Chemistry in First Quantization. *PRX Quantum*, **2**, 040332 (2021).

[77] T. E. O'Brien, L. B. Ioffe, Y. Su, D. Fushman, H. Neven, R. Babbush & V. Smelyanskiy. Quantum Computation of Molecular Structure Using Data from Challenging-To-Classically-Simulate Nuclear Magnetic Resonance Experiments. *PRX Quantum*, **3**, 030345 (2022).

[78] K. Seetharam, D. Biswas, C. Noel, A. Risinger, D. Zhu, O. Katz et al. Digital quantum simulation of NMR experiments, arXiv:2109.13298 (2021).

[79] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean & T. E. O'Brien. Quantum Error Mitigation. Technical report, arXiv (2022).

[80] A. Goussev, R. A. Jalabert, H. M. Pastawski & D. A. Wisniacki. Loschmidt echo. *Scholarpedia*, **7**, 11687 (2012).

[81] P. Wocjan & S. Zhang. Several natural BQP-Complete problems, arXiv:quant-ph/0606179 (2006).

[82] J. R. McClean, J. Romero, R. Babbush & A. Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, **18**, 023023 (2016).

[83] J.-S. Xu, M.-H. Yung, X.-Y. Xu, S. Boixo, Z.-W. Zhou, C.-F. Li, A. Aspuru-Guzik & G.-C. Guo. Demon-like algorithmic quantum cooling and its realization with quantum optics. *Nature Photonics*, **8**, 113–118 (2014).

[84] D. Poulin & P. Wocjan. Preparing ground states of quantum many-body systems on a quantum computer. *Physical Review Letters*, **102**, 130503 (2009).

[85] J. Lemieux, G. Duclos-Cianci, D. Sénéchal & D. Poulin. Resource estimate for quantum many-body ground-state preparation on a quantum computer. *Physical Review A*, **103**, 052408 (2021).

[86] L. Lin & Y. Tong. Near-optimal ground state preparation. *Quantum*, **4**, 372 (2020).

[87] S. McArdle, X. Yuan & S. Benjamin. Error-mitigated digital quantum simulation. *Physical Review Letters*, **122**, 180501 (2019).

[88] O. Kyriienko. Quantum inverse iteration algorithm for programmable quantum simulators. *npj Quantum Information*, **6**, 7 (2020).

[89] Y. Ge, J. Tura & J. I. Cirac. Faster ground state preparation and high-precision ground energy estimation with fewer qubits. *Journal of Mathematical Physics*, **60**, 22202 (2019).

[90] P. O. Boykin, T. Mor, V. Roychowdhury, F. Vatan & R. Vrijen. Algorithmic cooling and scalable NMR quantum computers. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 3388–3393 (2002).

[91] D. Kielpinski, B. E. King, C. J. Myatt, C. A. Sackett, Q. A. Turchette, W. M. Itano, C. Monroe, D. J. Wineland & W. H. Zurek. Sympathetic cooling of trapped ions for quantum logic. *Physical Review A: Atomic, Molecular, and Optical Physics*, **61**, 032310 (2000).

[92] H. Wang. Quantum algorithm for preparing the ground state of a system via resonance transition. *Scientific Reports*, **7**, 16342 (2017).

[93] M. Popp, J.-J. Garcia-Ripoll, K. G. Vollbrecht & J. I. Cirac. Ground-state cooling of atoms in optical lattices. *Physical Review A: Atomic, Molecular, and Optical Physics*, **74**, 013622 (2006).

[94] M. Raghunandan, F. Wolf, C. Ospelkaus, P. O. Schmidt & H. Weimer. Initialization of quantum simulators by sympathetic cooling. *Science Advances*, **6**, eaaw9268 (2020).

[95] M. Metcalf, J. E. Moussa, W. A. de Jong & M. Sarovar. Engineered thermalization and cooling of quantum many-body systems. *Physical Review Research*, **2**, 023214 (2020).

[96] S. Kretschmer, K. Luoma & W. Strunz. Collision model for non-Markovian quantum dynamics. *Physical Review A: Atomic, Molecular, and Optical Physics*, **94**, 012106 (2016).

[97] D. W. Berry, G. Ahokas, R. Cleve & B. C. Sanders. Efficient quantum algorithms for simulating sparse hamiltonians. *Communications in Mathematical Physics*, **270**, 359–371 (2007).

[98] Z.-C. Yang, A. Rahmani, A. Shabani, H. Neven & C. Chamon. Optimizing Variational Quantum Algorithms Using Pontryagin's Minimum Principle. *Physical Review X*, **7**, 021027 (2017).

[99] A. Bapat & S. Jordan. Bang-bang control as a design principle for classical and quantum optimization algorithms, arXiv:1812.02746 (2019).

[100] F. G. Brandao & K. M. Svore. Quantum speed-ups for solving semidefinite programs. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, volume 555, pages 415–426. IEEE (2017).

[101] S. Polla. Qdclib: A python package to simulate QDC protocols (2020). URL https://github.com/aQaLeiden/QuantumDigitalCooling.

[102] The cirq contributors. Quantumlib/Cirq (2023). URL https://github.com/quantumlib/Cirq.

[103] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *ArXiv:2004.04197* (2020).

[104] G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess et al. Quantum Approximate Optimization of the Long-Range Ising Model with a Trapped-Ion Quantum Simulator. *Proceedings of the National Academy of Sciences*, **117**, 25396–25401 (2020).

[105] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Physical Review X*, **8**, 031022 (2018).

[106] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom et al. Unsupervised Machine Learning on a Hybrid Quantum Computer, arXiv:1712.05771 (2017).

[107] D. Gottesman. Theory of fault-tolerant quantum computation. *Physical Review A: Atomic, Molecular, and Optical Physics*, **57** (1998).

[108] A. Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, **303**, 2–30 (2003).

[109] A. G. Fowler, M. Mariantoni, J. M. Martinis & A. N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A: Atomic, Molecular, and Optical Physics*, **86**, 032324 (2012).

[110] Y. R. Sanders, D. W. Berry, P. C. S. Costa, L. W. Tessler, N. Wiebe, C. Gidney, H. Neven & R. Babbush. Compilation of Fault-Tolerant Quantum Heuristics for Combinatorial Optimization. *PRX Quantum*, **1**, 020312 (2020).

[111] E. Campbell, A. Khurana & A. Montanaro. Applying quantum algorithms to constraint satisfaction problems. *Quantum*, **3** (2019).

[112] A. J. da Silva, T. B. Ludermir & W. R. de Oliveira. Quantum perceptron over a field and neural network architecture selection in a quantum computer. *Neural Networks*, **76**, 55–64 (2016).

[113] H. Chen, L. Wossnig, S. Severini, H. Neven & M. Mohseni. Universal discriminative quantum neural networks. *Quantum Machine Intelligence*, **3**, 1 (2020).

[114] S. Bravyi, D. Gosset & R. Movassagh. Classical algorithms for quantum mean values. *Nature Physics*, **17**, 337–341 (2021).

[115] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. Rol, C. Bultink, X. Fu et al. Error mitigation by symmetry verification on a variational quantum eigensolver. *Physical Review A: Atomic, Molecular, and Optical Physics*, **100**, 010302 (2019).

[116] J. R. McClean, M. E. Schwartz, J. Carter & W. A. de Jong. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Physical Review A: Atomic, Molecular, and Optical Physics*, **95**, 042308 (2017).

[117] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao et al. OpenFermion: The electronic

structure package for quantum computers. *Quantum Science and Technology*, **5**, 034014 (2020).

[118] K. TEMME, S. BRAVYI & J. M. GAMBETTA. Error mitigation for short-depth quantum circuits. *Physical Review Letters*, **119**, 180509 (2017).

[119] A. KANDALA, K. TEMME, A. D. CÓRCOLES, A. MEZZACAPO, J. M. CHOW & J. M. GAMBETTA. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, **567**, 491–495 (2019).

[120] S. ENDO, S. C. BENJAMIN & Y. LI. Practical quantum error mitigation for near-future applications. *Physical Review X*, **8**, 031027 (2018).

[121] N. C. RUBIN, R. BABBUSH & J. MCCLEAN. Application of fermionic marginal constraints to hybrid quantum algorithms. *New Journal of Physics*, **20**, 053020 (2018).

[122] X. BONET-MONROIG, R. SAGASTIZABAL, M. SINGH & T. E. O'BRIEN. Low-cost error mitigation by symmetry verification. *Physical Review A*, **98**, 62339 (2018).

[123] Z. JIANG, J. MCCLEAN, R. BABBUSH & H. NEVEN. Majorana loop stabilizer codes for error correction of fermionic quantum simulations. *Physical Review Applied*, **12**, 064041 (2019).

[124] T. E. O'BRIEN, B. TARASINSKI & B. M. TERHAL. Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments. *New Journal of Physics*, **21**, 023022 (2019).

[125] R. D. SOMMA. Quantum eigenvalue estimation via time series analysis. *New Journal of Physics*, **21**, 123025 (2019).

[126] A. E. RUSSO, K. M. RUDINGER, B. C. A. MORRISON & A. D. BACZEWSKI. Evaluating energy differences on a quantum computer with robust phase estimation. *Physical Review Letters*, **126**, 210501 (2021).

[127] D. AHARONOV & A. TA-SHMA. Adiabatic Quantum State Generation and Statistical Zero Knowledge, arXiv:quant-ph/0301023 (2003).

[128] J. D. WHITFIELD, J. BIAMONTE & A. ASPURU-GUZIK. Simulation of electronic structure hamiltonians using quantum computers. *Molecular Physics*, **109**, 735–750 (2011).

[129] B. L. HIGGINS, D. W. BERRY, S. D. BARTLETT, M. W. MITCHELL, H. M. WISEMAN & G. J. PRYDE. Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements. *New Journal of Physics*, **11**, 073023 (2009).

[130] N. WIEBE & C. GRANADE. Efficient bayesian phase estimation. *Physical Review Letters*, **117**, 010503 (2016).

[131] K. M. SVORE, M. B. HASTINGS & M. FREEDMAN. Faster Phase Estimation, arXiv:1304.0741 (2013).

[132] A. ROGGERO. Spectral density estimation with the Gaussian Integral Transform. *Physical Review A*, **102**, 022409 (2020).

[133] E. MAGESAN, J. M. GAMBETTA & A. J. EMERSON. Scalable and robust randomized benchmarking of quantum processes. *Physical Review Letters*, **106**, 180504 (2011).

[134] N. WIEBE, C. GRANADE, C. FERRIE & D. G. CORY. Hamiltonian learning and certification using quantum resources. *Physical Review Letters*, **112**, 190501 (2014).

[135] J. J. WALLMAN & J. EMERSON. Noise tailoring for scalable quantum computation via randomized compiling. *Physical Review A*, **94**, 052325 (2016).

[136] I. ZINTCHENKO & N. WIEBE. Randomized gap and amplitude estimation. *Physical Review A: Atomic, Molecular, and Optical Physics*, **93**, 062306 (2016).

[137] S. KIMMEL, G. H. LOW & T. J. YODER. Robust calibration of a universal single-qubit gate-set via robust phase estimation. *Physical Review A: Atomic, Molecular, and Optical Physics*, **92**, 062315 (2015).

[138] J. COTLER & F. WILCZEK. Quantum overlapping tomography. *Physical Review Letters*, **124** (2020).

[139] D. WANG, O. HIGGOTT & S. BRIERLEY. Accelerated variational quantum eigensolver. *Physical Review Letters*, **122**, 140504 (2019).

[140] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean et al. Witnessing eigenstates for quantum simulation of Hamiltonian spectra. *Science Advances*, **4** (2018).

[141] Y. Atia & D. Aharonov. Fast-forwarding of Hamiltonians and exponentially precise measurements. *Nature Communications*, **8**, 1572 (2017).

[142] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush & G. K.-L. Chan. Low rank representations for quantum simulation of electronic structure. *npj Quantum Information*, **7**, 83 (2021).

[143] G. Uhrig. Keeping a quantum bit alive by optimized π-Pulse sequences. *Physical Review Letters*, **98**, 100504 (2007).

[144] P. O'Malley, R. Babbush, I. Kivlichan, J. Romero, J. McClean, R. Barends et al. Scalable quantum simulation of molecular energies. *Physical Review X*, **6**, 031007 (2016).

[145] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari & W. J. Zeng. Digital zero noise extrapolation for quantum error mitigation. *IEEE Int. Conf. Quant. Comp. Eng.* (2020).

[146] X. Mi, P. Roushan, C. Quintana, S. Mandrà, J. Marshall, C. Neill et al. Information scrambling in quantum circuits. *Science (New York, N.Y.)*, **374**, 1479–1483 (2021).

[147] A. Montanaro & S. Stanisic. Error mitigation by training with fermionic linear optics, arXiv:2102.02120 (2021).

[148] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush & J. R. McClean. Virtual distillation for quantum error mitigation. *Physical Review X*, **11**, 041036 (2021).

[149] I. D. Kivlichan, J. McClean, N. Wiebe, A. A.-G. Craig Gidney, G. K.-L. Chan & R. Babbush. Quantum simulation of electronic structure with linear depth and connectivity. *Physical Review Letters*, **120**, 110501 (2018).

[150] L. G. Valiant. Quantum computers that can be simulated classically in polynomial time. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, STOC '01, pages 114–123.

Association for Computing Machinery, New York, NY, USA (2001). ISBN 978-1-58113-349-3.

[151] W. W. Ho & T. H. Hsieh. Efficient variational simulation of non-trivial quantum states. *SciPost Physics*, **6** (2019).

[152] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau et al. SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, **17**, 261–272 (2020).

[153] S. L. Braunstein & C. M. Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, **72**, 3439–3443 (1994).

[154] V. Giovannetti, S. Lloyd & L. Maccone. Quantum-enhanced measurements: Beating the standard quantum limit. *Science (New York, N.Y.)*, **306**, 1330–1336 (2004).

[155] E. Knill, G. Ortiz & R. D. Somma. Optimal quantum measurements of expectation values of observables. *Physical Review A*, **75**, 012328 (2007).

[156] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean & R. Babbush. Error Mitigation via Verified Phase Estimation. *PRX Quantum*, **2**, 020317 (2021).

[157] Z. Cai. Resource-efficient Purification-based Quantum Error Mitigation. Technical report, arXiv (2021).

[158] M. Huo & Y. Li. Dual-state purification for practical quantum error mitigation. *Physical Review A*, **105**, 022427 (2022).

[159] D. Wierichs, J. Izaac, C. Wang & C. Y.-Y. Lin. General parameter-shift rules for quantum gradients. *Quantum*, **6**, 677 (2022).

[160] G. G. Guerreschi & M. Smelyanskiy. Practical optimization for hybrid quantum-classical algorithms, arXiv:1701.01450 (2017).

[161] Y. Li & S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X*, **7**, 021050 (2017).

[162] V. Giovannetti, S. Lloyd & L. Maccone. Quantum metrology. *Physical Review Letters*, **96**, 010401 (2006).

[163] J. Saunders & J.-F. Van Huele. Qubit Quantum Metrology with Limited Measurement Resources, arXiv:2108.02876 (2021).

[164] C. Di Franco, M. Paternostro & M. S. Kim. Hamiltonian tomography in an access-limited setting without state initialization. *Physical Review Letters*, **102**, 187203 (2009).

[165] D. Burgarth, K. Maruyama & F. Nori. Indirect quantum tomography of quadratic Hamiltonians. *New Journal of Physics*, **13**, 013019 (2011).

[166] E. Knill & R. Laflamme. Power of one bit of quantum information. *Physical Review Letters*, **81**, 5672–5675 (1998).

[167] Y. Gu, Y. Ma, N. Forcellini & D. E. Liu. Noise-resilient phase estimation with randomized compiling. *Physical Review Letters*, **130**, 250601 (2023).

[168] K. Mitarai & K. Fujii. Methodology for replacing indirect measurements with direct measurements. *Physical Review Research*, **1**, 013006 (2019).

[169] A. Harrow & J. Napp. Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms. *Physical Review Letters*, **126**, 140502 (2021).

[170] G. Ortiz, J. E. Gubernatis, E. Knill & R. Laflamme. Quantum algorithms for fermionic simulations. *Physical Review A*, **64**, 022319 (2001).

[171] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith et al. PennyLane: Automatic differentiation of hybrid quantum-classical computations, arXiv:1811.04968 (2022).

[172] A. F. Izmaylov, T.-C. Yen & I. G. Ryabinkin. Revising measurement process in the variational quantum eigensolver: Is it possible to reduce the number of separately measured operators? *Chemical Science*, **10**, 3746–3755 (2019).

[173] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell & S. Brierley. Efficient quantum measurement of Pauli operators in the presence of finite sampling error. *Quantum*, **5** (2021).

[174] T.-C. YEN, A. GANESHRAM & A. F. IZMAYLOV. Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates. *npj Quantum Information*, **9**, 14 (2023).

[175] R. BARENDS, A. SHABANI, L. LAMATA, J. KELLY, A. MEZZACAPO, U. L. HERAS ET AL. Digitized adiabatic quantum computing with a superconducting circuit. *Nature*, **534**, 222–226 (2016).

[176] S. EBADI, A. KEESLING, M. CAIN, T. T. WANG, H. LEVINE, D. BLUVSTEIN ET AL. Quantum optimization of maximum independent set using Rydberg atom arrays. *Science (New York, N.Y.)*, **376**, 1209–1215 (2022).

[177] Y. KIM, A. EDDINS, S. ANAND, K. X. WEI, E. VAN DEN BERG, S. ROSENBLATT ET AL. Evidence for the utility of quantum computing before fault tolerance. *Nature*, **618**, 500–505 (2023).

[178] D. AHARONOV, W. VAN DAM, J. KEMPE, Z. LANDAU, S. LLOYD & O. REGEV. Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation. *SIAM Journal on Computing*, **37**, 166–194 (2007).

[179] A. MESSIAH. *Quantum Mechanics: Volume II.* North-Holland Publishing Company Amsterdam (1962).

[180] C. GROSS & I. BLOCH. Quantum simulations with ultracold atoms in optical lattices. *Science (New York, N.Y.)*, **357**, 995–1001 (2017).

[181] P. SCHOLL, M. SCHULER, H. J. WILLIAMS, A. A. EBERHARTER, D. BARREDO, K.-N. SCHYMIK ET AL. Quantum simulation of 2D antiferromagnets with hundreds of Rydberg atoms. *Nature*, **595**, 233–238 (2021).

[182] D. BLUVSTEIN, H. LEVINE, G. SEMEGHINI, T. T. WANG, S. EBADI, M. KALINOWSKI ET AL. A quantum processor based on coherent transport of entangled atom arrays. *Nature*, **604**, 451–456 (2022).

[183] A. D. KING, J. RAYMOND, T. LANTING, R. HARRIS, A. ZUCCA, F. ALTOMARE ET AL. Quantum critical dynamics in a 5,000-qubit programmable spin glass. *Nature*, **617**, 61–66 (2023).

[184] B. KOCZOR. Exponential Error Suppression for Near-Term Quantum Devices. *Physical Review X*, **11**, 031057 (2021).

[185] K. C. YOUNG, M. SAROVAR & R. BLUME-KOHOUT. Error Suppression and Error Correction in Adiabatic Quantum Computation: Techniques and Challenges. *Physical Review X*, **3**, 041013 (2013).

[186] M. C. TRAN, Y. SU, D. CARNEY & J. M. TAYLOR. Faster Digital Quantum Simulation by Symmetry Protection. *PRX Quantum*, **2**, 010323 (2021).

[187] S. BOIXO, E. KNILL & R. D. SOMMA. Eigenpath traversal by phase randomization. *Quantum Information & Computation*, **9**, 833–855 (2009).

[188] S. G. JOHNSON. Saddle-point integration of C_∞ "bump" functions, arXiv:1508.04376 (2015).

[189] S. SACHDEV. Quantum phase transitions. *Physics World*, **12**, 33 (1999).

[190] A. T. REZAKHANI, A. K. PIMACHEV & D. A. LIDAR. Accuracy versus run time in an adiabatic quantum search. *Physical Review A*, **82**, 052305 (2010).

[191] S. J. EVERED, D. BLUVSTEIN, M. KALINOWSKI, S. EBADI, T. MANOVITZ, H. ZHOU ET AL. High-fidelity parallel entangling gates on a neutral atom quantum computer. Technical report, arXiv (2023).

[192] J. ROLAND & N. J. CERF. Quantum Search by Local Adiabatic Evolution. *Physical Review A*, **65**, 042308 (2002).

[193] B. F. SCHIFFER, J. TURA & J. I. CIRAC. Adiabatic spectroscopy and a variational quantum adiabatic algorithm. *PRX Quantum*, **3**, 020347 (2022).

[194] M. CEREZO, A. ARRASMITH, R. BABBUSH, S. C. BENJAMIN, S. ENDO, K. FUJII ET AL. Variational quantum algorithms. *Nature Reviews Physics*, **3**, 625–644 (2021).

[195] Y. YANG, A. CHRISTIANEN, S. COLL-VINENT, V. SMELYANSKIY, M. C. BAÑULS, T. E. O'BRIEN, D. S. WILD & J. I. CIRAC. Simulating prethermalization using near-term quantum computers, arXiv:2303.08461 (2023).

[196] S. POLLA, G.-L. R. ANSELMETTI & T. E. O'BRIEN. Optimizing the information extracted by a single qubit measurement. *Physical Review A*, **108**, 012403 (2023).

[197] A. K. GEIM & K. S. NOVOSELOV. The rise of graphene. *Nature Materials*, **6**, 183–191 (2007).

[198] M. V. BERRY & M. WILKINSON. Diabolical Points in the Spectra of Triangles. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **392**, 15–43 (1984).

[199] W. DOMCKE, D. YARKONY & H. KÖPPEL, editors. *Conical Intersections: Theory, Computation and Experiment.* Number v. 17 in Advanced Series in Physical Chemistry. World Scientific, Singapore ; Hackensack, NJ (2011). ISBN 978-981-4313-44-5.

[200] D. R. YARKONY. Nonadiabatic Quantum Chemistry—Past, Present, and Future. *Chemical Reviews*, **112**, 481–498 (2012).

[201] D. POLLI, P. ALTOÈ, O. WEINGART, K. M. SPILLANE, C. MAN-ZONI, D. BRIDA ET AL. Conical intersection dynamics of the primary photoisomerization event in vision. *Nature*, **467**, 440–443 (2010).

[202] G. OLASO-GONZÁLEZ, M. MERCHÁN & L. SERRANO-ANDRÉS. Ultrafast Electron Transfer in Photosynthesis: Reduced Pheophytin and Quinone Interaction Mediated by Conical Intersections. *The Journal of Physical Chemistry B*, **110**, 24734–24739 (2006).

[203] H. E. ZIMMERMAN. Molecular Orbital Correlation Diagrams, Mobius Systems, and Factors Controlling Ground- and Excited-State Reactions. II. *Journal of the American Chemical Society*, **88**, 1566–1567 (1966).

[204] F. BERNARDI, M. OLIVUCCI & M. A. ROBB. Potential energy surface crossings in organic photochemistry. *Chemical Society Reviews*, **25**, 321–328 (1996).

[205] L. GONZÁLEZ, D. ESCUDERO & L. SERRANO-ANDRÉS. Progress and Challenges in the Calculation of Electronic Excited States. *ChemPhysChem*, **13**, 28–51 (2012).

[206] J. R. MCCLEAN, S. BOIXO, V. N. SMELYANSKIY, R. BABBUSH & H. NEVEN. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, **9**, 4812 (2018).

[207] S. TAMIYA, S. KOH & Y. O. NAKAGAWA. Calculating nonadiabatic couplings and Berry's phase by variational quantum eigensolvers. *Physical Review Research*, **3**, 023244 (2021).

[208] X. XIAO, J. K. FREERICKS & A. F. KEMPER. Robust measurement of wave function topology on NISQ quantum computers, arXiv:2101.07283 (2022).

[209] B. MURTA, G. CATARINA & J. FERNÁNDEZ-ROSSIER. Berry phase estimation in gate-based adiabatic quantum simulation. *Physical Review A*, **101**, 020302 (2020).

[210] H. C. LONGUET-HIGGINS, U. ÖPIK, M. H. L. PRYCE & R. A. SACK. Studies of the Jahn-Teller effect .II. The dynamical problem. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **244**, 1–16 (1958).

[211] C. A. MEAD. The "noncrossing" rule for electronic potential energy surfaces: The role of time-reversal invariance. *The Journal of Chemical Physics*, **70**, 2276–2283 (1979).

[212] I. G. RYABINKIN, L. JOUBERT-DORIOL & A. F. IZMAYLOV. Geometric Phase Effects in Nonadiabatic Dynamics near Conical Intersections. *Accounts of Chemical Research*, **50**, 1785–1793 (2017).

[213] J. WHITLOW, Z. JIA, Y. WANG, C. FANG, J. KIM & K. R. BROWN. Simulating conical intersections with trapped ions, arXiv:2211.07319 (2023).

[214] C. H. VALAHU, V. C. OLAYA-AGUDELO, R. J. MACDONELL, T. NAVICKAS, A. D. RAO, M. J. MILLICAN ET AL. Direct observation of geometric phase in dynamics around a conical intersection, arXiv:2211.07320 (2022).

[215] C. S. WANG, N. E. FRATTINI, B. J. CHAPMAN, S. PURI, S. M. GIRVIN, M. H. DEVORET & R. J. SCHOELKOPF. Observation of wave-packet branching through an engineered conical intersection. *Physical Review X*, **13**, 011008 (2023).

[216] E. KORIDON. Auto_oo (2023). URL https://github.com/Emieeel/auto_oo.

[217] E. TELLER. The Crossing of Potential Surfaces. *The Journal of Physical Chemistry*, **41**, 109–116 (1937).

[218] G. Herzberg & H. C. Longuet-Higgins. Intersection of potential energy surfaces in polyatomic molecules. *Discussions of the Faraday Society*, **35**, 77–82 (1963).

[219] T. Helgaker, P. Jørgensen & J. Olsen. *Molecular Electronic-Structure Theory*. Wiley, 1 edition (2000). ISBN 978-0-471-96755-2 978-1-119-01957-2.

[220] R. Broer, L. Hozoi & W. C. Nieuwpoort. Non-orthogonal approaches to the study of magnetic interactions. *Molecular Physics*, **101**, 233–240 (2003).

[221] V. Veryazov, P. Å. Malmqvist & B. O. Roos. How to select active space for multiconfigurational quantum chemistry? *International Journal of Quantum Chemistry*, **111**, 3329–3338 (2011).

[222] D. R. Yarkony. Diabolical conical intersections. *Reviews of Modern Physics*, **68**, 985–1013 (1996).

[223] C. Alden Mead. The molecular Aharonov—Bohm effect in bound states. *Chemical Physics*, **49**, 23–32 (1980).

[224] S. M. Harwood, D. Trenev, S. T. Stober, P. Barkoutsos, T. P. Gujarati, S. Mostame & D. Greenberg. Improving the Variational Quantum Eigensolver Using Variational Adiabatic Quantum Computing. *ACM Transactions on Quantum Computing*, **3**, 1:1–1:20 (2022).

[225] C. A. Mead & D. G. Truhlar. On the determination of Born–Oppenheimer nuclear motion wave functions including complications due to conical intersections and identical nuclei. *The Journal of Chemical Physics*, **70**, 2284–2296 (1979).

[226] R. J. Bartlett, S. A. Kucharski & J. Noga. Alternative coupled-cluster ansätze II. The unitary coupled-cluster method. *Chemical Physics Letters*, **155**, 133–140 (1989).

[227] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love & A. Aspuru-Guzik. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology*, **4**, 014008 (2018).

[228] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac & N. Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, **99**, 032331 (2019).

[229] H. J. A. Jensen & P. Jorgensen. A direct approach to second-order MCSCF calculations using a norm extended optimization scheme. *The Journal of Chemical Physics*, **80**, 1204–1214 (1984).

[230] B. Helmich-Paris. A trust-region augmented Hessian implementation for restricted and unrestricted Hartree–Fock and Kohn–Sham methods. *The Journal of Chemical Physics*, **154**, 164104 (2021).

[231] J. Nocedal & S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer, New York, 2nd ed edition (2006). ISBN 978-0-387-30303-1.

[232] E. P. Wigner. Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics*, **62**, 548–564 (1955).

[233] S. Yalouz, B. Senjean, J. Günther, F. Buda, T. E. O'Brien & L. Visscher. A state-averaged orbital-optimized hybrid quantum–classical algorithm for a democratic description of ground and excited states. *Quantum Science and Technology*, **6**, 024004 (2021).

[234] S. Yalouz, E. Koridon, B. Senjean, B. Lasorne, F. Buda & L. Visscher. Analytical Nonadiabatic Couplings and Gradients within the State-Averaged Orbital-Optimized Variational Quantum Eigensolver. *Journal of Chemical Theory and Computation*, **18**, 776–794 (2022).

[235] P.-O. Löwdin. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics*, **18**, 365–375 (1950).

[236] J. Cohn, M. Motta & R. M. Parrish. Quantum Filter Diagonalization with Compressed Double-Factorized Hamiltonians. *PRX Quantum*, **2**, 040352 (2021).

[237] P. Huembeli & A. Dauphin. Characterizing the loss landscape of variational quantum circuits. *Quantum Science and Technology*, **6**, 025011 (2021).

[238] H. Hirai. Excited-state molecular dynamics simulation based on variational quantum algorithms, arXiv:2211.02302 (2022).

[239] V. Bonačić-Koutecký & J. Michl. Photochemicalsyn-anti isomerization of a Schiff base: A two-dimensional description of a conical intersection in formaldimine. *Theoretica chimica acta*, **68**, 45–55 (1985).

[240] R. R. Birge. Nature of the primary photochemical events in rhodopsin and bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1016**, 293–327 (1990).

[241] M. Chahre. Trigger and Amplification Mechanisms in Visual Phototransduction. *Annual Review of Biophysics and Biophysical Chemistry*, **14**, 331–360 (1985).

[242] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt et al. Recent developments in the PySCF program package. *The Journal of Chemical Physics*, **153**, 024109 (2020).

[243] J. Stokes, J. Izaac, N. Killoran & G. Carleo. Quantum Natural Gradient. *Quantum*, **4**, 269 (2020).

[244] J. J. Meyer. Fisher Information in Noisy Intermediate-Scale Quantum Applications. *Quantum*, **5**, 539 (2021).

[245] S.-I. Amari. Natural Gradient Works Efficiently in Learning. *Neural Computation*, **10**, 251–276 (1998).

[246] T. Liang, T. Poggio, A. Rakhlin & J. Stokes. Fisher-Rao Metric, Geometry, and Complexity of Neural Networks, arXiv:1711.01530 (2019).

[247] E. G. Hohenstein, O. Oumarou, R. Al-Saadon, G.-L. R. Anselmetti, M. Scheurer, C. Gogolin & R. M. Parrish. Efficient Quantum Analytic Nuclear Gradients with Double Factorization, arXiv:2207.13144 (2022).

[248] J. K. Asbóth, L. Oroszlány & A. Pályi. *A Short Course on Topological Insulators*, volume 919 of *Lecture Notes in Physics*. Springer International Publishing, Cham (2016). ISBN 978-3-319-25605-4 978-3-319-25607-8.

[249] J. Zak. Berry's phase for energy bands in solids. *Physical Review Letters*, **62**, 2747–2750 (1989).

[250] Y. Hatsugai. Quantized Berry Phases as a Local Order Parameter of a Quantum Liquid. *Journal of the Physical Society of Japan*, **75**, 123601 (2006).

[251] T. Fukui, Y. Hatsugai & H. Suzuki. Chern Numbers in Discretized Brillouin Zone: Efficient Method of Computing (Spin) Hall Conductances. *Journal of the Physical Society of Japan*, **74**, 1674–1677 (2005).

[252] S.-s. CHERN. Characteristic Classes of Hermitian Manifolds. *Annals of Mathematics*, **47**, 85–121 (1946).

[253] R. CITRO & M. AIDELSBURGER. Thouless pumping and topology. *Nature Reviews Physics*, **5**, 87–101 (2023).

[254] D. J. THOULESS. Stability conditions and nuclear rotations in the Hartree-Fock theory. *Nuclear Physics*, **21**, 225–232 (1960).

[255] S. POLLA, Y. HERASYMENKO & T. E. O'BRIEN. Quantum digital cooling. *Physical Review A*, **104**, 012414 (2021).

[256] E. KORIDON, J. FRAXANET, A. DAUPHIN, L. VISSCHER, T. E. O'BRIEN & S. POLLA. A hybrid quantum algorithm to detect conical intersections, arXiv:2304.06070 (2023).

[257] B. F. SCHIFFER, D. VAN VREUMINGEN, J. TURA & S. POLLA. Virtual mitigation of coherent non-adiabatic transitions by echo verification, arXiv:2307.10358 (2023).

[258] P. STEINDL, H. SNIJDERS, G. WESTRA, E. HISSINK, K. IAKOVLEV, S. POLLA ET AL. Artificial Coherent States of Light by Multiphoton Interference in a Single-Photon Stream. *Physical Review Letters*, **126**, 143601 (2021).

[259] L. A. MARKOVICH, S. MALIKIS, S. POLLA & J. T. BRUGUÉS. Phase shift rule with the optimal parameter selection, arXiv:2309.07655 (2023).

Acknowledgments

Embarking on a career in scientific research is driven by various motivations, ranging from the dedication to technological progress to the fulfillment found in solving complex puzzles. Personally, I find great satisfaction in the collaborative essence of the scientific endeavor – whether engaged in lively whiteboard discussions, during conference sessions, or casual coffee breaks, the collective construction of knowledge remains a source of joy. In this limited space I can only thank a few among the many formidable individuals I could share this experience with in the past few years. All their voices, echoing more or less directly, resonate through the pages of this thesis.

Tom, thank you for leading me on this journey and teaching me to be a researcher. Five years ago, while still working on your own thesis, you took on the challenging task of constructing a research group and of supervising my PhD. You have done your work beautifully. I always felt supported yet encouraged to walk my own path; your caring personally for me and your team is the hallmark of a great supervisor.

Carlo, thank you for being the constant center of a great community, the *biggest fan* of us all. While our collaboration didn't center on scientific work, you taught me irreplaceable skills and gave me constant motivation, advice and opportunities to shine.

Luuk and Franco, thank you for your guidance and help in all the steps of (re)building the QC^2 project. You introduced me to a new field and its intriguing questions, and made me discover a passion for quantum chemistry.

devoted to solving problems and writing together. Eric, Dhruv and Victor, thank you for choosing to work with me for your Bachelor or Master projects, I am proud of your achievements.

In conclusion, I want to thank all the individuals in the QC^2 collaboration. Souloke and Edison, thank you for believing in me through your jump to a new research field. Detlef, I am grateful for the support and trust that paved the way for the success of this initiative.

I cannot end these acknowledgments without recognizing the weight that personal connections had in the years leading up to this moment. Without the unwaivering support of my friends and family it would not have been possible to pursue my PhD, especially during the crushing period marked by the COVID pandemic. This thesis is dedicated to all the people who brightened those days with their loving presence:

To my friends Dana, Annie, Mike, Ines, Katya, Margriet, Kostas, Joana, Sonakshi, who were there when I most needed them.

To Martina, Maurizio and Nicoletta, whose continuous support allowed me to get where I am and never fall.

To Jeanne and Charlie, who brighten each of my days.

Samenvatting

Quantumcomputing is een opkomende technologie die het potentieel heeft om complexe quantumsystemen te simuleren die buiten het bereik van klassieke numerieke methoden vallen. Ondanks de recente formidabele vooruitgang op het gebied van quantumhardware, blijft het bouwen van een quantumcomputer die nuttige berekeningen kan uitvoeren een uitdagende taak. Bij gebrek aan een betrouwbare quantumcomputer is de studie van potentiële toepassingen afhankelijk van wiskundige methoden, ingenieuze benaderingen en heuristieken die zijn afgeleid van de toepassingsgebieden. Dit proefschrift richt zich op de ontwikkeling van quantumalgoritmen met toepassingen in de simulatie van complexe quantumsystemen.

Het inleidende hoofdstuk schetst de uitdagingen van quantumsimulatie, identificeert en formaliseert de belangrijkste simulatiedoelstellingen, en bespreekt succesvolle quantum- en klassieke simulatiealgoritmen. Met behulp van de simulatie van chemische systemen als prototypisch voorbeeld leidt het hoofdstuk de lezer vervolgens door het proces van de ontwikkeling van quantumsimulaties.

De daaropvolgende hoofdstukken introduceren en detailleren nieuwe algoritmen voor quantumsimulatie, allemaal verbonden door de rode draad van het introduceren van een enkele hulpqubit in het simulatiealgoritme. Deze qubit (een fundamentele eenheid van quantuminformatie) speelt een actieve rol in elk algoritme en dient als sleutelelement in het constructieve ontwerp ervan. Het belang van het werken met een eenvoudig systeem, zoals een qubit, wordt door het hele proefschrift heen duidelijk en blijkt essentieel te zijn vanuit zowel fundamenteel als toepassingsgericht perspectief.

Hoofdstuk 2 onderzoekt het simuleren van koeling via een met één qubit gesimuleerde koelmachine. In de natuur koelen systemen af door interactie met grote koude omgevingen, waar ze warmte en entropie kunnen afvoeren. Hoewel het theoretisch mogelijk is om dergelijke baden te simuleren, brengt dit een aanzienlijke rekenlast met zich mee. We stellen voor om de omgeving te vervangen door een enkele hulpqubit, die periodiek wordt gereset naar zijn energiezuinige toestand, waardoor warmte en entropie aan het systeem kan worden onttrokken, analoog aan de werking van een koelkast. Ons onderzoek naar koelmachines met één qubit brengt ons ertoe een categorie algoritmen te introduceren die zijn ontworpen voor het voorbereiden van energiezuinige toestanden van gesimuleerde systemen, die we *quantum digital cooling* (quantumdigitale koeling) noemen. We beschrijven verschillende mogelijke benaderingen van deze algoritmen en karakteriseren deze met analytische en numerieke hulpmiddelen.

De volgende drie hoofdstukken hebben betrekking op *echo verification* (echoverificatie) — een nieuwe techniek voor het beperken van fouten die voor het eerst wordt geïntroduceerd in hoofdstuk 3. Quantumapparaten vormen een aanzienlijke uitdaging vanwege hun inherente ruis, die leidt tot de geleidelijke corruptie van opgeslagen en verwerkte gegevens en uiteindelijk tot rekenfouten. Quantumhardwareonderzoek heeft tot doel het ruisniveau te verminderen, met langetermijnvooruitzichten voor quantumfoutcorrectie om dit probleem op te lossen. Ondertussen moet elk effectief quantumalgoritme worden ontworpen met bestendigheid tegen ruis in overweging. Technieken voor het beperken van fouten spelen een cruciale rol bij het bieden van deze mogelijkheid.

De foutbeperkingstechniek die in hoofdstuk 3 is geïntroduceerd, schrijft voor dat een enkel bit aan informatie uit een toestand van het gesimuleerde systeem moet worden gemeten, terwijl de resterende quantuminformatie wordt gebruikt om fouten te detecteren en hun effect te verminderen. Deze aanpak wordt geïmplementeerd via een methode die doet denken aan de echo van Loschmidt, omdat de single-qubit-meting is ingeklemd tussen twee berekeningen die elkaar in de tijd weerspiegelen. Echoverificatie is gemakkelijk toepasbaar op een grote verscheidenheid aan quantumalgoritmen die zijn afgestemd op bestaande technologie. We onderzoeken de toepassing ervan op het schatten van verwachtingswaarden en het schatten van quantumfasen met één controlequbit, waarbij we numerieke benchmarks uitvoeren op simulaties van kleine quantumchemie- en magnetismemodellen. De toepassing van deze techniek resulteerde in de meest uitgebreide experimentele test van een variationeel quantumalgoritme voor de chemie.

In hoofdstuk 4 bestuderen we de theoretische grenzen van het meetmodel

opgelegd door echoverificatie, waarbij per toestandsvoorbereiding een enkel bit aan informatie wordt geëxtraheerd. Dit model sluit aan bij de algemene categorie van binaire metingen en vertegenwoordigt ja-nee-vragen die kunnen worden gesteld over een quantumtoestand, en heeft een theoretische betekenis die verder gaat dan het bereik van echoverificatie. We ontwikkelen een raamwerk om de metingen van verwachtingswaarden binnen dit model te optimaliseren en een substantiële prestatieverbetering aan te tonen vergeleken met een naïeve aanpak.

In Hoofdstuk 4 onderzoeken we het gebruik van echoverificatie in een duidelijk andere context: het beperken van algoritmische fouten in het adiabatische algoritme. Het adiabatische algoritme is een belangrijke techniek die wordt gebruikt om fundamentele toestanden van een gesimuleerd systeem voor te bereiden. Hoewel de adiabatische stellingen hun succes verzekeren binnen de limiet van de oneindige rekentijd, zijn praktische eindige-tijdberekeningen gevoelig voor systematische fouten. Deze fout is van een andere aard dan de stochastische hardwarefout die doorgaans wordt aangepakt door mitigatietechnieken. We introduceren een techniek die het mogelijk maakt om de systematische fout om te zetten in een stochastische fout, waardoor de toepassing van echoverificatie het effect ervan kan onderdrukken.

Ten slotte onderzoekt hoofdstuk 6 een mogelijke toepassing van bestaande quantumcomputers in de quantumchemie: de detectie van conische intersecties in een moleculair model. Een groot deel van de quantumchemie is gebaseerd op de Born-Oppenheimer-benadering, die de beschrijving van kernen en elektronen scheidt. Conische intersecties zijn belangrijke punten in de geometrie van een molecuul, waar de Born-Oppenheimer-benadering mislukt. Dit vergemakkelijkt processen zoals niet-stralingsrelaxatie, die van bijzonder belang zijn bij de studie van fotochemische reacties. Conische intersecties worden gekenmerkt door een eigenschap die Berry-fase wordt genoemd en die alleen de discrete waarden 0 of π kan aannemen. We ontwerpen een quantumalgoritme dat het mogelijk maakt om deze twee waarden te onderscheiden, weergegeven in een enkel stukje informatie. De inherente discretisatie van het resultaat zorgt ervoor dat het algoritme bestand is tegen een bepaalde hoeveelheid ruis. We leveren analytisch bewijs van deze robuustheid en we voeren numerieke tests uit op een moleculair model dat is ontworpen om bepaald gedrag te reproduceren van het biochemische systeem dat verantwoordelijk is voor lichtperceptie.

Summary

Quantum computing is an emerging technology, holding the potential to simulate complex quantum systems beyond the reach of classical numerical methods. Despite recent formidable advancements in quantum hardware, constructing a quantum computer capable of performing useful calculations remains a challenging task. In the absence of a reliable quantum computer, the study of potential applications relies on mathematical methods, ingenious approximations, and heuristics derived from the application fields. This thesis focuses on the development of quantum algorithms with applications in the simulation of complex quantum systems.

The introductory chapter outlines the challenges of quantum simulation, identifies and formalizes key simulation targets, and reviews successful quantum and classical simulation algorithms. Using the simulation of chemical systems as a prototypical example, the chapter then guides the reader through the quantum simulation development pipeline.

Subsequent chapters introduce and detail novel algorithms for quantum simulation, all connected by the common thread of introducing a single auxiliary qubit into the simulation algorithm. This qubit (a fundamental unit of quantum information) plays an active role in each algorithm, serving as a key element in their constructive design. The significance of working with a simple system, such as a qubit, becomes evident throughout the thesis, proving essential from both fundamental and applicative perspectives.

Chapter 2 explores simulating cooling through a single-qubit emulated fridge. In Nature, systems cool down by interacting with large cold environ-

ments, where they can dissipate heat and entropy. While simulating such baths is theoretically possible, it comes with a significant computational burden. We propose to substitute the environment with a single auxiliary qubit, which is periodically reset to its low-energy state allowing to extract heat and entropy from the system, analogous to the functioning of a fridge. Our investigation of single-qubit fridges leads us to introduce a category of algorithms designed for preparing low-energy states of simulated systems, which we name quantum digital cooling. We describe various possible approaches to quantum digital cooling, characterizing them with analytical and numerical tools.

The following three chapters relate to Echo Verification – a novel error mitigation technique first introduced in chapter 3. Quantum devices present a significant challenge due to their inherent noise, leading to the gradual corruption of stored and processed data and ultimately causing computation errors. Quantum hardware research aims to reduce noise levels, with long-term prospects for quantum error correction to solve this problem. Meanwhile, any effective quantum algorithm must be designed with resilience to noise in consideration. Error mitigation techniques play a crucial role in providing this resilience.

The error mitigation technique introduced in chapter 3 prescribes to measure a single bit of information from a state of the simulated system, while using the remaining quantum information to detect errors and contrast their effect. This approach is implemented through a method reminiscent of Loschmidt's echo, as the single-qubit measurement is sandwiched between two computations which mirror each other in time. Echo verification is readily applicable to a wide variety of quantum algorithms tailored to near-term devices. We explore its application to expectation value estimation and single-control quantum phase estimation, conducting numerical benchmarks on simulations of small quantum chemistry and magnetism models. The adoption of this technique resulted in the most extensive experimental test of a variational quantum algorithm for chemistry.

In chapter 4, we study the theoretical limits of the measurement model imposed by echo verification, wherein a single bit of information is extracted per state preparation. This model aligns to the general category of binary measurements, representing yes-no questions that can be posed about a quantum state, holding theoretical significance beyond the scope of echo verification. We develop a framework to optimize measurements of expectation values within this model, and demonstrate a substantial improvement in performance compared to a naive approach.

In Chapter 4, we explore the utilization of echo verification in a distinctly

different context: mitigating algorithmic errors in the adiabatic algorithm. The adiabatic algorithm is an important technique utilized to prepare fundamental states of a simulated system. While the adiabatic theorems ensure its success in the limit of infinite computation time, practical finite-time computations are susceptible to systematic errors. This error is of different nature from the stochastic hardware error typically addressed by mitigation techniques. We introduce a technique that allows to convert the systematic error into a stochastic one, enabling the application of echo verification to suppress its effect.

Finally, chapter 6 explores a potential application of near-term quantum computers in quantum chemistry: the detection of conical intersections in a molecular model. Much of quantum chemistry relies on the Born-Oppenheimer approximation, which separates the description of nuclei and electrons. Conical intersections are significant points in the geometry of a molecule, where the Born-Oppenheimer approximation breaks down. This facilitates processes like non-radiative relaxation, holding particular significance in the study of photochemical reactions. Conical intersections are characterized by a property called Berry phase, which can only take the discrete values of 0 or π. We design a quantum algorithm that allows to discern these two values, represented in a single bit of information. The inherent discreteness of the result renders the algorithm resilient to a certain amount of noise. We provide analytical proof of this resilience, and we conduct numerical testing on a molecular toy model designed to reproduce some behaviors of the biochemical system responsible for light perception.

Sinossi

La computazione quantistica è una tecnologia emergente che ha il potenziale di permettere la simulazione di sistemi quantistici complessi, oltre la portata dei metodi numerici classici. Nonostante gli ultimi anni abbiano visto formidabili progressi nello sviluppo dell'hardware quantistico, costruire un computer quantistico capace di eseguire calcoli utili rimane una sfida onerosa. In assenza di un calcolatore quantistico affidabile, lo studio delle possibili applicazioni si basa su metodi matematici, astute approssimazioni e strategie euristiche derivanti dai vari campi di applicazione. Questa tesi si concentra sullo sviluppo di algoritmi quantistici applicati alla simulazione di sistemi quantistici complessi.

Il capitolo introduttivo delinea le sfide della simulazione quantistica, individua e formalizza gli obiettivi principali ed esamina gli algoritmi (classici e quantistici) di maggior successo. Utilizzando la simulazione di sistemi chimici come esempio prototipico, il capitolo guida il lettore attraverso il percorso che porta allo sviluppo di un'applicazione della simulazione quantistica.

I capitoli successivi presentano e approfondiscono nuovi algoritmi per la simulazione quantistica, tutti accomunati dalla caratteristica di introdurre un singolo qubit ausiliario nell'algoritmo di simulazione. Questo qubit (un'unità fondamentale di informazione quantistica) gioca un ruolo attivo in ciascun algoritmo, fungendo da elemento chiave nel processo di sviluppo del metodo. La rilevanza del lavorare con un sistema semplice, come un qubit, si manifesta nel corso dell'intera tesi, dimostrandosi essenziale sia dal punto di vista fondamentale che applicativo.

Il capitolo 2 esplora la simulazione del processo di raffreddamento attraverso un frigorifero emulato a singolo qubit. In natura, i sistemi fisici si raffreddano interagendo con un ambiente esteso e freddo, nel quale possono dissipare calore ed entropia. Anche se la simulazione di tali ambienti termodinamici è possibile a livello teorico, essa comporta un notevole onere computazionale. Proponiamo di sostituire l'ambiente con un singolo qubit ausiliario, che viene periodicamente azzerato al suo stato di minima energia, consentendo quindi la ripetuta estrazione di calore ed entropia dal sistema simulato, in maniera analoga ad un frigorifero. La nostra indagine sui frigoriferi a singolo qubit ci porta a introdurre una categoria di algoritmi quantistici progettati per preparare stati fondamentali di sistemi simulati, che chiamiamo *quantum digital cooling* (raffreddamento quantistico digitale). Descriviamo vari approcci per la realizzazione di questi algoritmi, caratterizzandoli con strumenti analitici e numerici.

I successivi tre capitoli sono collegati alla tecnica di mitigazione degli errori chiamata *echo verification* (verifica dell'eco), sviluppata ed introdotta nel capitolo 3. Una sfida significativa del calcolo quantistico è posta dal rumore intrinseco ai dispositivi quantistici, che porta al graduale degrado dei dati memorizzati e processati e, in definitiva, provoca errori di calcolo. La ricerca sull'hardware quantistico mira a ridurre i livelli di rumore, ed a lungo termine la correzione quantistica degli errori promette di risolvere questo problema. Nel frattempo, qualsiasi algoritmo quantistico efficace deve essere progettato tenendo in considerazione la resilienza al rumore. Le tecniche di mitigazione degli errori svolgono un ruolo cruciale nel fornire questa resilienza.

La tecnica di mitigazione degli errori introdotta nel capitolo 3 prescrive di misurare un singolo bit di informazione da uno stato del sistema simulato, utilizzando il resto dell'informazione quantistica per rilevare gli errori e contrastarne gli effetti. Quest'approccio è implementato attraverso un metodo che ricorda l'eco di Loschmidt, poiché la misurazione di un singolo qubit è interposta tra due computazioni speculari nel tempo. La verifica dell'eco è facilmente applicabile a una vasta gamma di algoritmi quantistici progettati per i dispositivi del prossimo futuro. Esploriamo la sua applicazione alla stima del valore atteso e alla stima quantistica della fase a singolo controllo, conducendo test numerici su simulazioni di piccoli modelli chimica quantistica e magnetismo. L'adozione di questa tecnica ha portato al test sperimentale di un algoritmo quantistico variazionale per la chimica di massima dimensione.

Nel capitolo 4, studiamo i limiti teorici del modello di misurazione imposto dalla verifica dell'eco, in cui viene estratto un singolo bit di informazione da ogni stato preparato. Questo modello corrisponde alla

categoria generale delle misurazioni con esito binario, che ha rilevanza teorica al di là dello specifico campo della mitigazione degli errori. Sviluppiamo una metodologia per ottimizzare la misurazione del valore atteso all'interno di questo modello e dimostriamo un miglioramento sostanziale delle prestazioni rispetto a un approccio elementare.

Nel capitolo 4, esploriamo l'utilizzo della verifica dell'eco in un contesto nettamente diverso: la mitigazione degli errori algoritmici nell'algoritmo adiabatico. L'algoritmo adiabatico è un'importante tecnica utilizzata per preparare stati fondamentali di un sistema simulato. Seppure i teoremi adiabatici ne garantiscono il successo nel limite di tempo di calcolo infinito, in pratica la sua approssimazione a tempo finito è suscettibile ad errori sistematici. Questo errore ha una natura diversa dall'errore stocastico causato dal rumore dei dispositivi quantistici, sul quale le tecniche di mitigazione si concentrano. Introduciamo una tecnica che consente di convertire l'errore sistematico in uno stocastico, consentendo l'applicazione della verifica dell'eco per ridurre i suoi effetti.

Infine, il capitolo 6 esplora una potenziale applicazione dei computer quantistici del prossimo futuro in chimica quantistica: la identificazione di intersezioni coniche in un modello molecolare. Gran parte della chimica quantistica si basa sull'approssimazione di Born-Oppenheimer, che separa la descrizione di nuclei ed elettroni. Le intersezioni coniche sono punti significativi nella geometria di una molecola, dove l'approssimazione di Born-Oppenheimer cessa di essere valida. Questo permette processi come il rilassamento non radiativo, e riveste particolare importanza nello studio delle reazioni fotochimiche. Le intersezioni coniche sono caratterizzate da una proprietà chiamata fase di Berry, che può assumere solo i valori discreti di 0 o π. Progettiamo un algoritmo quantistico che consente di distinguere questi due valori, una distinzione rappresentabile in un singolo bit di informazione. La natura discreta del risultato rende l'algoritmo resiliente a una certa quantità di rumore; forniamo una prova matematica di questa resilienza. Conduciamo test numerici su un modello molecolare giocattolo, progettato per riprodurre alcuni comportamenti del sistema biochimico responsabile della percezione della luce.

Curriculum Vitæ

I was born on the 26th of January 1994 in Bergamo, Italy. From a young age I was fascinated by magnets, gyroscopes and symbolic languages. While attending secondary education at LSS Filippo Lussana, I developed an interest for electronics, music and philosophy.

Guided by my passions, I began my Bachelor's studies in Physics at Università degli Studi di Milano in 2013 and graduated *cum laude* in 2016. During bachelor thesis, I engaged in an internship at LENS in Florence, bridging the research of M. Bellini's experimental quantum optics group and M. G. A. Paris's quantum information theory group in Milan. Throughout my studies, I played an active role in the Italian Association of Physics Students (AISF), organizing physics events at local, national, and international levels.

I got to know Leiden through a LION summer school, and decided to move there to pursue a Master's in physics in 2017. Admitted to the Casimir Master Track, I had the opportunity to work on short research projects in experimental quantum optics (Löffler group) and spin qubits (Vandersypen group in TU Delft). My final thesis on quantum digital cooling under the supervision of T. E. O'Brien and Y. Herasymenko, led to my graduation *summa cum laude* in 2019. During this period, I received the Casimir prize for academic performance and the Lorentz prize for the best master thesis in theoretical physics in the Netherlands. As part of the Casimir track, I composed a research proposal on *Non-Unitary Algorithms for Near-Term Quantum Computing*, which was awarded with funding for a Ph.D. position and forms the base of this thesis.

Supervised by T. E. O'Brien and C. W. J. Beenakker, my doctoral research centered on advancing quantum algorithms for the simulation of complex quantum systems. During my time at Leiden, I found great satisfaction in contributing to the instruction of Master courses in condensed matter physics and applied quantum algorithms. I actively contributed to the development of course content for the latter, taking charge of sections covering numerics for quantum simulation algorithms and applications to quantum chemistry.

In the final two years of my doctoral program, I concurrently held a part-time student researcher role with the Google Quantum AI team, which allowed me to absorb invaluable knowledge and a perspective on research oriented to ambitious goals.

During my doctoral studies, I witnessed the birth and growth of the Applied Quantum Algorithms Leiden $\langle \text{aQa}^\text{L} \rangle$ group, actively contributing by fostering collaborations both within and outside Leiden University. In 2022, I authored a research proposal on quantum algorithms for quantum chemistry, securing continued funding from Shell to support the research in collaboration with L. Visscher's Theoretical Chemistry group at VU Amsterdam. This funding facilitated the hiring of two postdoctoral researchers, kicking off the QC^2 (quantum computational quantum chemistry) project. In the upcoming period, I am dedicated to continuing this research initiative in Leiden. As a part of this, I am co-organizing a Lorentz Center workshop, which will bring together world-class experts in the simulation of quantum systems in Leiden.

List of publications

[255] S. Polla, Y. Herasymenko & T. E. O'Brien. Quantum digital cooling. *Physical Review A*, **104**, 012414 (2021).

[Chapter 2 is based on this publication.]

[156] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean & R. Babbush. Error Mitigation via Verified Phase Estimation. *PRX Quantum*, **2**, 020317 (2021).

[Chapter 3 is based on this publication.]

[196] S. Polla, G.-L. R. Anselmetti & T. E. O'Brien. Optimizing the information extracted by a single qubit measurement. *Physical Review A*, **108**, 012403 (2023).

[Chapter 4 is based on this publication.]

[257] B. F. Schiffer, D. van Vreumingen, J. Tura & S. Polla. Virtual mitigation of coherent non-adiabatic transitions by echo verification, arXiv:2307.10358 (2023).

[Chapter 5 is based on this preprint.]

[256] E. Koridon, J. Fraxanet, A. Dauphin, L. Visscher, T. E. O'Brien & S. Polla. A hybrid quantum algorithm to detect conical intersections, arXiv:2304.06070 (2023).

[Chapter 6 is based on this preprint.]

[258] P. Steindl, H. Snijders, G. Westra, E. Hissink, K. Iakovlev, S. Polla, J. A. Frey, J. Norman, A. C. Gossard, J. E. Bowers, D. Bouwmeester, and W. Löffler Artificial Coherent States of Light by Multiphoton Interference in a Single-Photon Stream. *Physical Review Letters*, **126**, 143601 (2021).

[69] T. E. O'Brien, G. Anselmetti, F. Gkritsis, V. E. Elfving, S. Polla, W. J. Huggins et al. Purification-based quantum error mitigation of pair-correlated electron simulations. *Nature Physics*, **19**, 1787–1792 (2023).

[259] L. A. Markovich, S. Malikis, S. Polla & J. T. Brugués. Phase shift rule with the optimal parameter selection, arXiv:2309.07655 (2023).