

Quasiprobabilistic Imaginary-Time Evolution on Quantum Computers

Annie Ray^{1,2,3}, Esha Swaroop^{1,2,3}, Ningping Cao^{1,3,*}, Michael Vasmer^{1,3,4} and Anirban Chowdhury⁵

¹ Institute for Quantum Computing, University of Waterloo, Waterloo, ON N2L 0A4, Canada

² Department of Physics and Astronomy, University of Waterloo, Canada

³ Perimeter Institute for Theoretical Physics, 31 Caroline St N, Waterloo, ON N2L 2Y5, Canada

⁴ Inria Paris, 48 rue Barrault, 75013 Paris, France

⁵ IBM Quantum, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA; anirban.chowdhury@ibm.com

* Corresponding author: ncao@uwaterloo.ca

Received date: 2 June 2025; Accepted date: 25 December 2025; Published online: 15 January 2026

Abstract: Imaginary-time evolution plays an important role in algorithms for computing ground-state and thermal equilibrium properties of quantum systems, but can be challenging to simulate on classical computers. Many quantum algorithms for imaginary-time evolution have resource requirements that are prohibitive for current quantum devices and face performance issues due to noise. Here, we propose a new algorithm for computing imaginary-time evolved expectation values on quantum computers, inspired by probabilistic error cancellation, an error-mitigation technique. Our algorithm works by decomposing a Trotterization of imaginary-time evolution into a probabilistic linear combination of operations, each of which is then implemented on a quantum computer. The measurement data is then classically post-processed to obtain the expectation value of the imaginary-time evolved state. Our algorithm requires no ancillary qubits and can be made noise-resilient without additional error-mitigation. It is well-suited for estimating thermal expectation values by making use of the notion of a thermal pure quantum state. We demonstrate our algorithm by performing numerical simulations of thermal pure quantum state preparation for the 1D Heisenberg Hamiltonian on 8 qubits, and by using an IBM quantum computer to estimate the energy of the same Hamiltonian on 2 qubits. We observe promising results compared to the exact values, illustrating the potential of our algorithm for probing the physics of quantum many-body systems on current hardware.

Keywords: quantum algorithms; imaginary-time evolution; quantum error mitigation; thermal expectation values

1. Introduction

Quantum computers are expected to have wide-ranging applications in various domains [1–4] but perhaps most significantly in the simulation of quantum physical systems. Important problems in this arena concern the estimation of ground-state and thermal equilibrium properties of quantum many-body systems, problems which have applications in chemistry and material science [2,5–7]. Consequently, there has been an enormous amount of research [8–11] directed towards the development of quantum algorithms for these problems. Although computational complexity results limit the possibility for dramatic quantum speed-ups for these problems [12–14], there is also theoretical evidence that quantum computers potentially provide an advantage over classical algorithms [15].

The idea of *imaginary-time evolution* plays a key role in many algorithms, both classical and quantum, for simulating ground-state and equilibrium properties [16–24]. Given a Hamiltonian H and an inverse temperature $\beta \geq 0$, imaginary-time evolution (ITE) under H is described by the exponential operator $e^{-\beta H}$. Performing ITE with sufficiently high β will drive a quantum system to its ground state, making it a useful primitive in algorithms for ground state preparation. ITE is also used as a subroutine in quantum algorithms for preparing thermal equilibrium or Gibbs states [25], which find application in solving optimization problems. Performing ITE of quantum Hamiltonians using classical methods can be computationally prohibitive due to the exponential increase in Hilbert space dimension and the sign problem [26,27]. Many existing quantum algorithms such as those based on block-encoding [28–32] require fault-tolerant quantum computers and are beyond the reach of current hardware [33]. Several recent works

have therefore proposed quantum algorithms that are potentially more suitable for near-term to intermediate quantum devices [34–40].

In this work, we add to the quantum toolkit for imaginary-time evolution and develop a new near-term friendly algorithm using ideas from quantum error mitigation (QEM), a collection of classical post-processing techniques that aim to mitigate the effects of noise without requiring the full machinery of quantum error correction and fault-tolerance [41,42]. These techniques are usually not scalable, but have been shown to significantly improve the performance of current devices [42]. In particular, our work is inspired by the error-mitigation technique of probabilistic error cancellation (PEC) [41,43–45]. The original motivation of PEC as a technique is to invert a noise channel on noisy quantum devices [46]. Probabilistic error cancellation simulates the action of an ideal, unitary operation from a linear combination of native channels implementable on a quantum device aided by post-processing. The basic idea underlying PEC is as follows. Given a quantum device with a set of known native (and possibly noisy) gates, one decomposes a target unitary U into a weighted sum of the native gates. The coefficients in the linear combination will be real numbers that sum to 1, although they may or may not be positive and hence form a *quasi-probability distribution* (QPD). Nevertheless, the action of U can be effected by a protocol that probabilistically samples a native gate from the linear combination. With a sufficient number of samples and simple classical post-processing, the expectation value of an observable after the action of the target unitary U can be reconstructed. Recovering the ideal expectation value comes at an increased cost related to the negativity of the quasi-probability distribution. The success of the PEC method relies on an accurate characterization of the native gates in the quantum device. This can be a very challenging task in practice, although recent work has shown that tailoring the device noise to Pauli noise (via randomized compiling [47]) allows efficient and accurate noise characterization [48]. We note that related QPD ideas have also been used in other quantum information contexts, such as classical simulation of quantum circuits [49–53], circuit cutting [54–58], and for mitigating noise in non-Clifford gates [59].

Here, we design a quasi-probabilistic algorithm inspired by PEC to perform imaginary-time evolution for a given Hamiltonian H . The starting point of our algorithm is the observation that we can implement non-unitary and even non-trace-preserving operations as a linear combination of native gates and measurement channels. We use our algorithm to estimate the expectation value of a given observable for imaginary-time evolved states, and we analyze its cost, specifically for the case of Trotterized imaginary time evolution. In addition, we show that our algorithm is suitable for preparing *thermal pure quantum* (TPQ) states, pure states which can be used to estimate the thermal expectation value of observables; see Figure 1. We demonstrate the efficacy of our approach through classical simulations and demonstrations on quantum hardware.

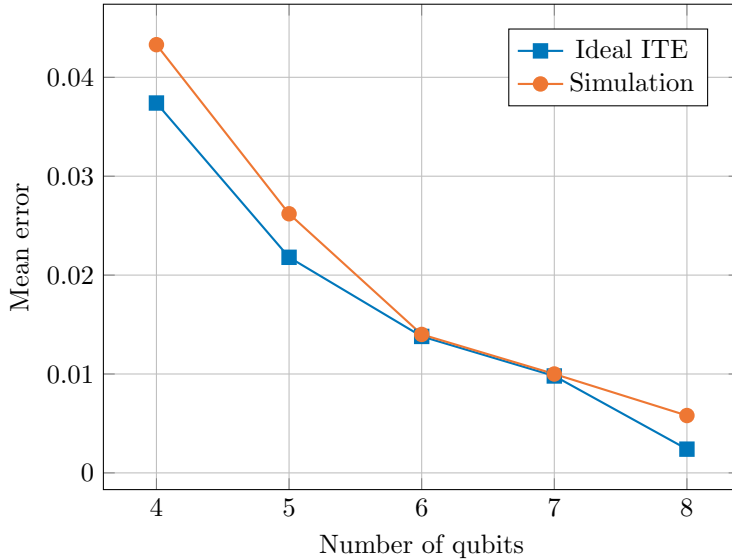


Figure 1. Thermal expectation value estimation using ITE. For each data point, we average the error in the thermal expectation value $|\langle \psi_i | O_j | \psi_i \rangle - \text{tr}(|\psi_\beta\rangle\langle\psi_\beta| O_j)|$ over 30 randomly chosen Pauli observables O_j and 10 random TPQ states $|\psi_i\rangle = e^{-0.02H} U_i |0\rangle^{\otimes n}$, where H is the 1D Heisenberg Hamiltonian on n qubits and U_i is a randomly chosen Clifford unitary. The blue and orange points show the values for exact and simulated ITE, respectively, where the simulation implements Algorithm 1.

Our method of simulating ITE in expectation using quasiprobability decompositions has a few features that make it amenable for implementation on near-term devices where qubits are few, and multi-qubit entangling gates are noisier than single-qubit gates. First, our method does not need auxiliary qubits beyond those of the system described by a given Hamiltonian. Second, we avoid the problem of implementing multi-qubit-controlled operations between auxiliary and data qubits [25,60], which can require large SWAP networks for implementation on devices that only permit entangling operations between nearest-neighbors (such as present-day superconducting devices).

The remainder of the paper is organized as follows. In Section 2, we review the quasiprobability decomposition of linear maps and discuss our extension to nonlinear maps, including our algorithm for estimating rescaled expectation values. Following this, in Section 3, we use our algorithm for the specific case of imaginary time evolution and combine the protocol with various choices of initial states, focusing on the thermal pure quantum state. In this section, we also present the results of our demonstrations conducted using IBM’s quantum computing hardware. Finally, we provide a summary and discuss future work in Section 5.

2. Implementing Linear Maps via Quasiprobability Decompositions

We consider the problem of implementing Hermiticity-preserving linear maps—which may not be trace-preserving in general—on a quantum device with a (basis) set of completely positive (CP) native operations. In Section 2.1, we present our algorithm for estimating rescaled expectation values of observables with respect to quantum states evolving under the linear map of interest. In Section 2.2, we analyze the sampling cost of our algorithm. Our method requires, as input, a decomposition of the linear map of interest into a linear combination of the set of available operations. In Section 2.4, we discuss the conditions under which a linear map can be decomposed as such, and we briefly review the method to obtain the linear decomposition. And in Section 2.3, we generalize our algorithm to the case of implementing a sequence of linear maps.

2.1. Procedure for Estimating Rescaled Expectation Values of Observables

Suppose we have a linear map $\mathcal{T}$ over the 2^n -dimensional Hilbert space, and a set $\{\mathcal{B}_i\}$ of native operations on a quantum device such that for some real coefficients q_i ,

$$\mathcal{T}(\rho) = \sum_i q_i \mathcal{B}_i(\rho) \quad (1)$$

for all input states ρ . Using Eq. (1), we can define parameters γ and p_i such that

$$\gamma := \sum_i |q_i|, \quad (2)$$

$$p_i := \frac{|q_i|}{\gamma}. \quad (3)$$

We refer to such a linear combination as a quasiprobability decomposition (QPD). Note that since the q_i are real, $\gamma \geq 0$ and the p_i form a probability distribution. Moreover, Eq. (1) can be rewritten as

$$\mathcal{T}(\rho) = \gamma \sum_i \text{sgn}(q_i) \cdot p_i \cdot \mathcal{B}_i(\rho). \quad (4)$$

This immediately suggests a way to probabilistically build the map $\mathcal{T}$ by sampling from the set $\{\mathcal{B}_i\}$ according to the probability distribution p_i (see Figure 2).

In particular, we consider the task of estimating the expectation value of some observable A (having operator norm $\|A\| \leq 1$) with respect to the state $\mathcal{T}(\rho)$. Since $\mathcal{T}$ may be a non-trace-preserving map, the state $\mathcal{T}(\rho)$ is unnormalized in general. Thus, the rescaled expectation value of the observable A is

$$\langle A \rangle_{\mathcal{T}(\rho)} := \frac{\text{tr}[A\mathcal{T}(\rho)]}{\text{tr}[\mathcal{T}(\rho)]}. \quad (5)$$

In order to correctly estimate the rescaled expectation value of A , we require estimates of the numerator and denominator of Eq. (5) which are obtained using Eq. (4)

$$\text{tr}[A\mathcal{T}(\rho)] = \gamma \sum_i \text{sgn}(q_i) \cdot p_i \cdot \text{tr}[A\mathcal{B}_i(\rho)], \quad (6)$$

$$\text{tr}[\mathcal{T}(\rho)] = \gamma \sum_i \text{sgn}(q_i) \cdot p_i \cdot \text{tr}[\mathcal{B}_i(\rho)]. \quad (7)$$

Denoting the trace-preserving and non-trace-preserving operations in $\{\mathcal{B}_i\}$ as $\mathcal{B}_{\text{TP}}$ and $\mathcal{B}_{\text{NTP}}$, respectively, the denominator in Eq. (5) can be rewritten as

$$\text{tr}[\mathcal{T}(\rho)] = \gamma \left(\sum_{\mathcal{B}_i \in \mathcal{B}_{\text{TP}}} \text{sgn}(q_i) \cdot p_i + \sum_{\mathcal{B}_i \in \mathcal{B}_{\text{NTP}}} \text{sgn}(q_i) \cdot p_i \cdot \text{tr}[\mathcal{B}_i(\rho)] \right). \quad (8)$$

In this work, we consider $\mathcal{B}_{\text{NTP}}$ to comprise rank-1 projectors $\Pi_{|\psi\rangle} = |\psi\rangle\langle\psi|$ on specific quantum states $|\psi\rangle$ (see Table A1 for example). Typically, such non-trace-preserving operations would be implemented via a POVM given by $\{|\psi\rangle\langle\psi|, |\psi^\perp\rangle\langle\psi^\perp|\}$ and post-selecting on the outcome $|\psi\rangle$. Thus, $\text{tr}[\mathcal{B}_i(\rho)]$ is the probability of post-selecting the outcome corresponding to the projector for $\mathcal{B}_i$.

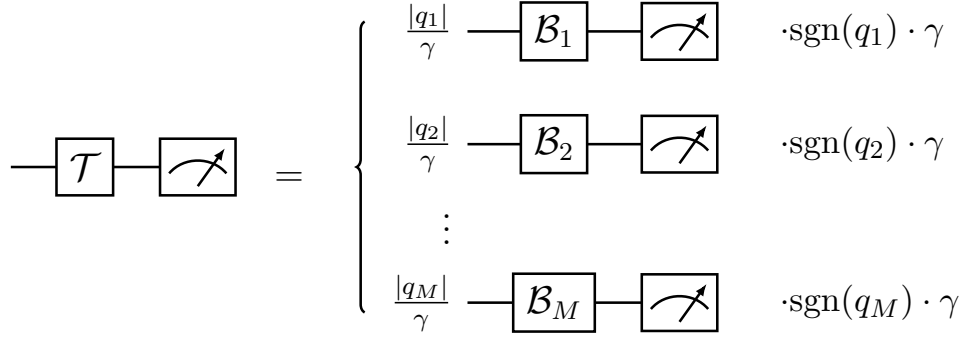


Figure 2. A schematic of a linear map $\mathcal{T}$ implemented by sampling from its quasiprobability distribution (see Eq. (4), Algorithm 1). Given a quasiprobability decomposition of the map $\mathcal{T}$, circuits are randomly sampled according to the distribution $|q_i|/\gamma$, and their measurement outcomes are weighted by the factors $\text{sgn}(q_i)\gamma$.

If we randomly sample N circuits from the probability distribution p_i in Eq. (2), the random variable $W_j := \gamma \text{sgn}(q_i) I_j$ corresponding to the j^{th} circuit is an estimator (see Eq. (9)) for $\text{tr}[\mathcal{T}(\rho)]$, where $j \in \{1, 2, \dots, N\}$. Here, the indicator variable I_j is 0 when the j^{th} circuit implements an operation $\mathcal{B}_i \in \mathcal{B}_{\text{NTP}}$ whose associated POVM outcome is discarded; otherwise $I_j = 1$.

$$\mathbb{E}[W_j] = \mathbb{E}[\gamma \text{sgn}(q_i) I_j] = \gamma \sum_{i, I_j} \text{sgn}(q_i) I_j \cdot \frac{|q_i|}{\gamma} \cdot \text{Pr}[I_j] \quad (9)$$

$$= \gamma \sum_i \text{sgn}(q_i) \cdot p_i \cdot \text{tr}[\mathcal{B}_i(\rho)] \quad (10)$$

$$= \text{tr}[\mathcal{T}(\rho)] \quad (11)$$

Similarly, we show in Appendix D that the random variable $M_j = \gamma \text{sgn}(q_i) A_j$, where A_j is the outcome of measuring the observable A for the j^{th} circuit, is an estimator of $\text{tr}[A\mathcal{T}(\rho)]$. We remark that the measurement outcome A_j is necessarily zero when the indicator variable I_j is zero, i.e., when the j^{th} sample is discarded.

$$\mathbb{E}[M_j] = \text{tr}[A\mathcal{T}(\rho)] \quad (12)$$

Algorithm 1 Estimate re-scaled expectations

Input:

QPD of $\mathcal{T}$ as $\mathcal{T}(\cdot) = \gamma \sum_i \text{sgn}(q_i) p_i \mathcal{B}_i(\cdot)$
 Initial state ρ
 Number of random samples N
 Observable A

Output: Estimate of the re-scaled expectation value $\langle A \rangle_{\mathcal{T}}$

```

1: for  $j = 1$  to  $N$  do
2:   Sample  $i$  from the distribution  $p_i$ 
3:   if  $\mathcal{B}_i$  is TP then
4:     apply  $\mathcal{B}_i$  directly to  $\rho$ 
5:      $I_j \leftarrow 1$  ▷ Indicator variable
6:   else
7:     Measure in the basis corresponding to  $\mathcal{B}_i$ 
8:     Store the outcome as  $I_j \in \{0, 1\}$ 
9:   end if
10:   $W_j \leftarrow \gamma \cdot \text{sgn}(q_i) \cdot I_j$  ▷ Weighted outcome
11:  Measure  $A$ 
12:  Store the outcome as  $A_j$ 
13:   $M_j \leftarrow \gamma \cdot \text{sgn}(q_i) \cdot A_j$ 
14: end for
15:  $\text{tr}[\mathcal{T}(\rho)] \leftarrow \sum_j W_j / N$  ▷ Empirical estimate of  $\text{tr}[\mathcal{T}(\rho)]$ 
16:  $\text{tr}[A\mathcal{T}(\rho)] \leftarrow \sum_j M_j / N$  ▷ Empirical estimate of  $\text{tr}[A\mathcal{T}(\rho)]$ 
17: return  $\text{tr}[A\mathcal{T}(\rho)] / \text{tr}[\mathcal{T}(\rho)]$ 

```

Together, the two estimators W_j, M_j allow us to estimate the rescaled expectation value in Eq. (5). We formalize the procedure described above in Algorithm 1.

2.2. Sampling Cost of the Algorithm

The crux of Algorithm 1 lies in the estimation of $\text{tr}[A\mathcal{T}(\rho)]$ and $\text{tr}[\mathcal{T}(\rho)]$ separately, in order to estimate $\langle A \rangle_{\mathcal{T}(\rho)}$ via their ratio as per Eq. (5). We use standard concentration inequalities to bound the number of samples N needed to estimate $\text{tr}[A\mathcal{T}(\rho)]$ and $\text{tr}[\mathcal{T}(\rho)]$ each within some precision ϵ , which allows us to bound the error in the estimate of $\langle A \rangle_{\mathcal{T}(\rho)}$ as per Lemma 1.

Lemma 1 (Algorithm 1 cost). *Given a decomposition of a linear map $\mathcal{T}$ into a set of operations as in Eq. (1) and a Hermitian observable A with $\|A\| \leq 1$, Algorithm 1 uses $N = \mathcal{O}\left(\frac{\gamma^2}{\epsilon^2}\right)$ random circuits drawn from the distribution p_i in Eq. (2), to return an estimate of the re-scaled expectation value $\langle A \rangle_{\mathcal{T}}$ within error*

$$\frac{2\epsilon}{|\text{tr}[\mathcal{T}(\rho)]|} \left(1 + \frac{\epsilon}{|\text{tr}[\mathcal{T}(\rho)]|} \right). \quad (13)$$

Proof sketch of Lemma 1. We note that for an observable A with operator norm $\|A\| \leq 1$, the random variables W_j, M_j are such that

$$-\gamma \leq W_j, M_j \leq \gamma. \quad (14)$$

For N circuits (enumerated by the index j) sampled independently from p_i , we can define the following:

$$\overline{W} = \frac{\sum_j W_j}{N}, \quad (15)$$

$$\overline{M} = \frac{\sum_j M_j}{N} \quad (16)$$

For some $\epsilon \geq 0$, Hoeffding's inequality implies

$$\Pr\left[|\overline{W} - \mathbb{E}[W_j]| \geq \epsilon\right] \leq \exp\left(-\frac{N\epsilon^2}{2\gamma^2}\right) \quad (17)$$

$$\Pr\left[|\overline{M} - \mathbb{E}[M_j]| \geq \epsilon\right] \leq \exp\left(-\frac{N\epsilon^2}{2\gamma^2}\right). \quad (18)$$

We can bound the RHS of Eq. (17) and (18) by $0 \leq \delta < 1$ by choosing $N = \frac{2\gamma^2}{\epsilon^2} \log(1/\delta)$. Then with probability at least $1 - \delta$, we have

$$\left|\frac{\overline{M}}{\overline{W}} - \frac{\mathbb{E}[M_j]}{\mathbb{E}[W_j]}\right| \leq \left|\frac{\mathbb{E}[W_j](\overline{M} - \mathbb{E}[M_j])}{\overline{W}\mathbb{E}[W_j]}\right| + \left|\frac{\mathbb{E}[M_j](\mathbb{E}[W_j] - \overline{W})}{\overline{W}\mathbb{E}[W_j]}\right| \quad (19)$$

$$\leq \frac{\epsilon}{|\overline{W}|} \left(1 + \left|\frac{\mathbb{E}[M_j]}{\mathbb{E}[W_j]}\right|\right) \quad (20)$$

$$\leq \frac{2\epsilon}{|\overline{W}|} \quad (21)$$

In the above, we have used Eqs (9) and (12), and that $\frac{\mathbb{E}[M_j]}{\mathbb{E}[W_j]} \leq 1$. Additionally, we have

$$\frac{1}{|\overline{W}|} \leq \frac{1}{||\mathbb{E}[W_j]| - \epsilon|} \leq \frac{1}{|\mathbb{E}[W_j]|} \left(1 + \frac{\epsilon}{|\mathbb{E}[W_j]|}\right). \quad (22)$$

Using Eqs (21) and (22), and recalling Eq. (5), we see that for the chosen N , with probability at least $(1 - \delta)$, we have

$$\left|\frac{\overline{M}}{\overline{W}} - \langle A \rangle_{\mathcal{T}}\right| \leq \frac{2\epsilon}{|\text{tr}[\mathcal{T}(\rho)]|} \left(1 + \frac{\epsilon}{|\text{tr}[\mathcal{T}(\rho)]|}\right). \quad (23)$$

□

Algorithm 1 requires a decomposition of the form Eq. (1) as input. Given a basis, such a decomposition can be obtained classically by a linear program [41,61]. The cost of finding the QPD scales polynomially with the Hilbert-space dimension and is thus tractable only for operations on k -qubit subsystems with $k \ll n$. Implementing an n -qubit operation therefore requires us to first approximate it as a sequence of k -qubit operations, and then use the QPD method for each operation in the sequence. This protocol and an analysis of its cost are the focus of the next section. We defer a detailed discussion of how to find a suitable quasi-probability decomposition to Section 2.4. There we will also discuss some possible choices of QPD basis, and their effect on the cost of our algorithms through the factor γ .

2.3. Implementing a Sequence of Operations

We now explain how to implement a sequence of linear maps $\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_R$ where each map $\mathcal{T}_i$ acts non-trivially on a constant k number of qubits (see Figure 3). Later, we will seek to implement an imaginary-time evolution by approximating it through a Trotter-Suzuki formula that gives rise to a sequence of k -local maps. We restrict ourselves to the case where each $\mathcal{T}_i$ has the form $\mathcal{T} \otimes \mathcal{I}$ for some k -qubit map $\mathcal{T}$. That is to say, $\mathcal{T}_i$ acts as $\mathcal{T}$ on a k -qubit subsystem and as the identity on all other qubits; note that each $\mathcal{T}_i$ may act non-trivially on a different set of k qubits. We emphasize that our algorithm and its analysis do not require the $\mathcal{T}_i$'s to have this form. We make this assumption because it helps simplify notation and is sufficient for implementing imaginary-time evolution of k -local Hamiltonians such as the Heisenberg model. We present the protocol for applying a sequence of $\mathcal{T}$ maps in Algorithm 2, and in Lemma 2, we analyze its cost.

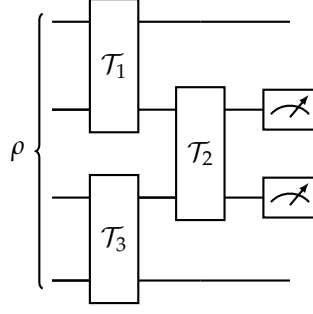


Figure 3. Example of an operation $\mathcal{T}' = \mathcal{T}_1 \mathcal{T}_2 \mathcal{T}_3$ that we implement. Here $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ are 2-qubit operations applied on nearest-neighbor qubits at different locations. In the analysis of Lemma 2, we assume for simplicity that $\mathcal{T}_1 = \mathcal{T}_2 = \mathcal{T}_3 = \mathcal{T}$.

Algorithm 2 Estimate re-scaled expectations for repeated applications of $\mathcal{T}$

Input:

Operator $\mathcal{E}$ consisting of R applications of $\mathcal{T}$
 QPD of $\mathcal{T}$ as $\mathcal{T}(\cdot) = \gamma \sum_i \text{sgn}(q_i) p_i \mathcal{B}_i(\cdot)$
 Initial state ρ
 Number of random samples N
 Observable A

Output: Estimate of the re-scaled expectation value $\langle A \rangle_{\mathcal{E}}$

```

1: for  $j = 1$  to  $N$  do
2:   Sample  $(i_1, \dots, i_R)$  from the joint (independent) distribution  $p_i^R$ 
3:   for  $k = 1$  to  $R$  do
4:     if  $\mathcal{B}_{i_k}$  is TP then
5:       Apply  $\mathcal{B}_{i_k}$ 
6:        $I_{j_k} \leftarrow 1$  ▷ Indicator variable
7:     else
8:       Measure in the basis corresponding to  $\mathcal{B}_{i_k}$ 
9:       Store the outcome as  $I_{j_k} \in \{0, 1\}$ 
10:    end if
11:     $W_{j_k} \leftarrow \gamma \cdot \text{sgn}(q_{i_k}) \cdot I_{j_k}$  ▷ Weighted outcome
12:  end for
13:  Measure  $A$ 
14:  Store the outcome as  $A_j$ 
15:   $M_j \leftarrow \gamma^R \cdot \text{sgn}(q_{i_1}) \dots \text{sgn}(q_{i_R}) \cdot A_j$ 
16: end for
17:  $\text{tr}[\widetilde{\mathcal{E}(\rho)}] \leftarrow \sum_j (\prod_k W_{j_k}) / N$  ▷ Empirical estimate of  $\text{tr}[\mathcal{E}(\rho)]$ 
18:  $\text{tr}[\widetilde{A\mathcal{E}(\rho)}] \leftarrow \sum_j M_j / N$  ▷ Empirical estimate of  $\text{tr}[A\mathcal{E}(\rho)]$ 
19: return  $\text{tr}[\widetilde{A\mathcal{E}(\rho)}] / \text{tr}[\widetilde{\mathcal{E}(\rho)}]$ 

```

Lemma 2 (Algorithm 2 cost). *Suppose we are given the QPD of a linear map $\mathcal{T}$, acting non-trivially on a constant number of qubits. Define the n -qubit map $\mathcal{E}$ as the composition of R applications of $\mathcal{T}$, where each $\mathcal{T}$ acts non-trivially on at most k -qubits (see Figure 3). Then, for a Hermitian observable A with $\|A\| \leq 1$, Algorithm 2 uses $N = \mathcal{O}\left(\frac{\gamma^{2R}}{\epsilon^2}\right)$ random circuits drawn from the distribution p_i in Eq. (2), to return an estimate of the rescaled expectation value $\langle A \rangle_{\mathcal{E}}$ within error*

$$\frac{2\epsilon}{|\text{tr}[\mathcal{E}(\rho)]|} \left(1 + \frac{\epsilon}{|\text{tr}[\mathcal{E}(\rho)]|} \right). \quad (24)$$

The proof of Lemma 2 is similar to that of Lemma 1 and can be found in Appendix D.

2.4. Quasiprobability decomposition of linear maps

We now consider the task of obtaining the linear decomposition in Eq. (1), which is required as an input to Algorithms 1 and 2. Given a k -qubit map $\mathcal{T}$ and a complete basis (consisting of 16^k linearly independent operators), the coefficients q_i in Eq. (1) can be directly solved for via a system of 16^k independent equations. The tensor product of the Endo-Benjamin-Li (EBL) basis [43] is an example of a complete basis for all linear maps. It is composed of single-qubit Clifford operations (TP) and projectors (non-TP); see Table A1. It may be possible for a basis with fewer elements than 16^k to be used to obtain a QPD for a given map $\mathcal{T}$. We note that a basis containing only trace-preserving (TP) elements can only be used to express maps $\mathcal{T}$ that are scalar multiples of TP operations. However, operations such as the ITE operator are not simply scalar multiples of TP operations,¹ and therefore require the QPD basis to have at least one non-TP operation.

Different choices of basis will affect the values of the coefficients q_i and hence the value of γ . From the analysis of our algorithm in Section 2.2, we know that the number of samples required for estimating expectation values scales as $\mathcal{O}(\gamma^2)$. Consequently, the sampling cost of our method can be reduced by minimizing the value of γ in the QPD of the map of interest. In the context of PEC, the QPD of a k -qubit map $\mathcal{T}$ over a basis $\{\mathcal{B}_i\}$ can be obtained by solving a linear program [41,61] that minimizes γ . It is known that a lower bound of γ for a HPTP map $\mathcal{T}$ is given by its diamond norm $\|\mathcal{T}\|_\diamond$ [62,63].²

As an example of the impact of the choice of basis, we consider the ITE operator $e^{-\beta H}$ of the 2-qubit Heisenberg Hamiltonian $H = -XX - YY - ZZ + II$.³ In Figure 4, we show the trends of γ as a function of β for two different choices of basis. We find that the ITE operator for the 2-qubit Heisenberg Hamiltonian can be decomposed into the Takagi basis (Table A2) with a lower value of γ than the EBL basis, where the Takagi basis value is closer to the lower bound given by the diamond norm (see Figure 4). This reduced value of γ can be attributed to the presence of entangling operations in the Takagi basis. Since the EBL basis for 2-qubits is simply the (uncorrelated) Cartesian product of two single-qubit basis sets, it has a high sampling overhead for simulating entangling operations in expectation.

This is apparent in the case of the CNOT operation [43,64], which has $\gamma_{\text{CNOT}} = 9$ for the 2-qubit EBL basis but $\gamma_{\text{CNOT}} = 1$ when decomposed into the Takagi basis. We remark that although the Takagi basis permits a decomposition of the ITE operator in this case, it is not a complete basis for general HP maps on a 2-qubit system. The Takagi basis is only known to be complete for the set of all 2-qubit CPTP maps [64]. A more detailed analysis of QPD and choices of basis can be found in Appendix A.

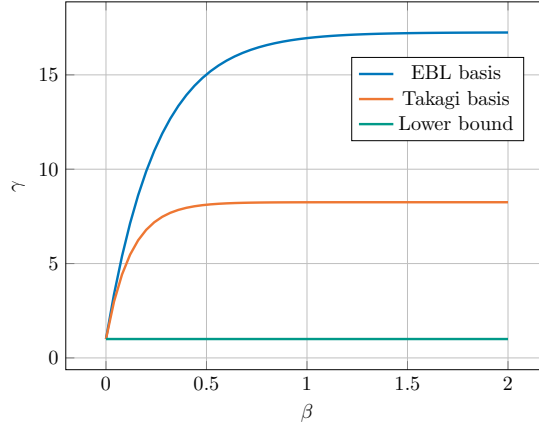


Figure 4. Scaling of the QPD cost γ with inverse temperature β for the decomposition of $e^{-\beta H}$ using different basis sets, where H is the 2-qubit Heisenberg Hamiltonian $H = -XX - YY - ZZ + II$. The Takagi basis (orange) outperforms the EBL basis (blue), illustrating the utility of including entangling gates in the QPD basis set. The lower bound, obtained using the diamond norm [62,63], is shown in teal. (See Appendix C for the analysis of optimal sampling cost and the effect of adding identity to the Hamiltonian for ITE.)

¹The trace of the state resulting from applying the ITE operation is dependent on the input state.

²The optimal cost γ_{opt} is precisely $\|\mathcal{T}\|_\diamond$ or $2\|\mathcal{T}\|_\diamond$ when allowing the basis to consist of arbitrary quantum instruments, see [62,63].

³See Appendix C for an explanation of the added identity term.

The problem of obtaining a QPD for a given map requires careful consideration of two important factors: the choice of basis set used for the decomposition and the value of γ obtained from the QPD. In general, the QPD corresponding to the optimal value of γ could be a linear combination of entangling operations or general quantum instruments that need to dilate to a bigger Hilbert space, which may be expensive (in gate depth) to implement directly on a quantum device. This tradeoff between the generality of the basis and the optimality of the decomposition [43, 64, 65] has been well studied for the specific case of QPDs for CPTP maps, with several results on optimizing [45, 61] the choice of basis for different use-cases. However, the solution to this problem for general HP maps, including NTP maps such as ITE which commonly occur in physics, is yet (to the best of our knowledge) unknown. We provide some discussions in Appendix A. Additionally, the problem of finding a sample-efficient yet hardware-friendly basis for decomposing maps acting on systems larger than 2-qubits quickly becomes intractable (since the size of the associated classical optimization problem grows polynomially with the dimension of the Hilbert space) and remains a challenge even for the case of decomposing CPTP maps.

3. QPD Implementation of Imaginary-Time Evolution

We now use the algorithms developed in Section 2 to implement imaginary time evolution $e^{-\beta H}$ for a local Hamiltonian H

$$H = \sum_{l=1}^L H_l, \quad (25)$$

presented as a sum of $L = O(\text{poly}(n))$ local terms H_l acting non-trivially on at most a constant k number of qubits. We may assume without loss of generality that $H \geq 0$ and $\|H\| \leq 1$, where $\|\cdot\|$ denotes the operator norm. Our goal is to implement the operator $e^{-\beta H}$ for real $\beta > 0$. More precisely, for a given initial state $|\psi\rangle$, we want to estimate expectation values of a given observable A with respect to the quantum state $|\psi_\beta\rangle \propto e^{-\beta H}|\psi\rangle$. Some basics of ITE and the optimal decomposition of ITE are presented in Appendix B and Appendix C, respectively.

3.1. Warm-Up Example: 2-Qubit Heisenberg Model

We first consider the simple case of implementing the ITE of a 2-qubit Heisenberg model

$$H = -XX - YY - ZZ \quad (26)$$

Since the basis $\{\mathcal{B}_i\}$ in Table A1 is known to be complete for all linear maps on single-qubits, we use the 256-element Cartesian product set $\{\mathcal{B}_i\} \times \{\mathcal{B}_j\}$ as the basis for decomposing the 2-qubit ITE operator $e^{-\beta H}$ into a QPD. In Figure 5, we simulate Algorithm 1 for estimating the energy of the 2-qubit Heisenberg model when it is initialized in the state $|00\rangle$ and undergoes ITE for increasing values of $\beta = 0.01t$, where t is the Trotter step. We note that the simulation is noiseless and there is no Trotter error because H commutes with itself. Therefore, the only error in the simulation is due to the QPD sampling.

Additionally, we implement the 2-qubit ITE operator $e^{-\beta H}$ using Algorithm 1 on a 5-qubit superconducting processor (`ibm_manila`). In this case, we use the noisy native gates on the device as the basis for QPD using noise characterization information. In particular, we take the noisy basis $\{\mathcal{B}'_i\}$ to be $\{\mathcal{N} \circ \mathcal{B}_i\}$ where $\mathcal{N}$ is the noise channel used to model the noise in the device, based on calibration data provided by IBM [66].

In the simulation and demonstration on hardware, we used 400, 800, 3200, and 25600 QPD samples for Trotter steps 1, 2, 3, and 4, respectively. For each sample, 512 shots of the corresponding circuit were performed (or simulated). For each Trotter step, this process was repeated 10 times; the data in Figure 5 show the mean estimate and the standard deviation. It is instructive to compare the number of samples we used with the cost estimate from 1 of 26000, 50000, 79000, and 114000 circuits for Trotter steps 1, 2, 3, and 4, respectively (for a target error of 1%). These numbers suggest that our analytical bound is not tight.

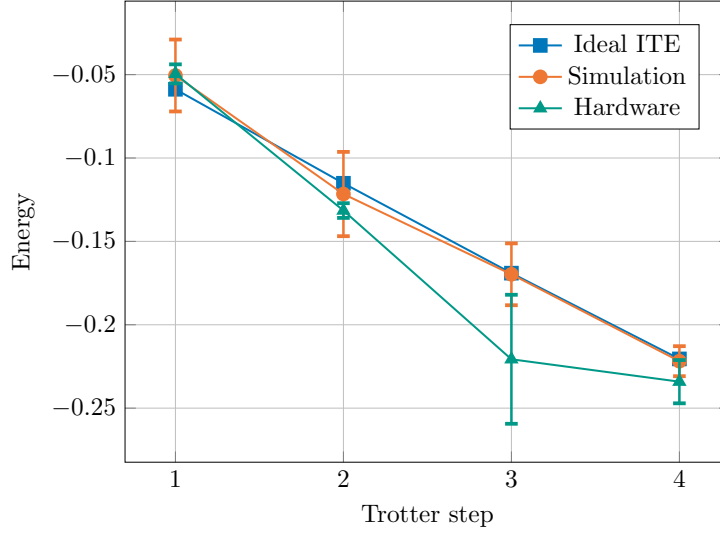


Figure 5. Energy estimation using ITE. We estimate the energy $\langle \psi_t | H | \psi_t \rangle / \langle \psi_t | \psi_t \rangle$, where H is the 2-qubit Heisenberg Hamiltonian and $|\psi_t\rangle = (e^{-0.01H})^t |00\rangle$ is the imaginary-time evolved state with t Trotter steps. The blue, orange, and teal points show the values for exact, simulated, and hardware ITE, respectively, where the simulation (resp., demonstration) implements Algorithm 2 on classical (resp., quantum) hardware.

The agreement between the demonstration on noisy hardware and the noiseless (classical) simulation indicates that our method can be made noise-resilient when information about the noise in a device is available. We note that the demonstration data for the third Trotter step is significantly worse than for the other Trotter steps. A possible explanation for this is that the device noise drifts over time, and so the noisy basis $\{\mathcal{B}'_i\}$ may become inaccurate, leading to bias in the results. This is a particular issue for demonstrations such as ours, where we ran many different circuits, each with a small number of shots, which is much slower than running a small number of circuits with many shots each. The results could also possibly be improved by using more QPD samples (compare the size of the (simulated data) error bars of the third and fourth Trotter steps in 5).

3.2. QPD Implementation of Trotterized Imaginary Time Evolution

We now define $\mathcal{T}(\cdot)$ to be the imaginary-time evolution for the n -qubit Hamiltonian in Eq. (28)

$$\mathcal{T}(\cdot) = e^{-\beta H}(\cdot)e^{-\beta H}. \quad (27)$$

We cannot apply Algorithm 1 directly to $\mathcal{T}(\cdot)$ since the task of obtaining the quasiprobability decomposition over an n -qubit basis set ($|\{\mathcal{B}_i\}| = \mathcal{O}(16^n)$) by solving a linear program can quickly become intractable for large n . Accordingly, we first decompose the n -qubit operation $e^{-\beta H}$ to a sequence of k -qubit operations by using the 1st-order Trotter approximation:

$$e^{-\beta H} \approx \left(e^{-\beta H_1/r} e^{-\beta H_2/r} \dots e^{-\beta H_L/r} \right)^r. \quad (28)$$

Since each H_l is k -local, we obtain the quasiprobability decomposition of each local ITE operation $e^{-\beta/r H_l}$ separately and implement a Trotterized ITE using Algorithm 2. Let $\mathcal{T}_l(\cdot)$ be the ITE operation for the k -local Hamiltonian H_l

$$\mathcal{T}_l(\cdot) = e^{-\beta H_l/r}(\cdot)e^{-\beta H_l/r} \quad (29)$$

We require the QPDs of each of the linear maps $\mathcal{T}_1, \dots, \mathcal{T}_L$ into a set of operations $\{\mathcal{B}_i\}$ as in Eq. (1). Suppose that we have computed these decompositions with the associated parameters $\gamma_1, \dots, \gamma_L$ respectively. Then, we define the n -qubit map $\mathcal{P}$ as the composition of the maps $(\mathcal{T}_1, \dots, \mathcal{T}_L)$ corresponding to the 1st-order Trotter approximation in Eq. (28). For a Hermitian observable A with $\|A\| \leq 1$, it follows from Lemma 2 that implementing $\mathcal{P}$ using

Algorithm 2 returns an estimate of the re-scaled expectation value $\langle A \rangle_{\mathcal{P}}$ within additive error $\epsilon \cdot (\text{tr}\{\mathcal{P}(\rho)\})^{-1}$ using $\mathcal{O}(\gamma_1^{2r} \dots \gamma_L^{2r} \epsilon^{-2})$ circuits.

In our implementation, we only consider a translation-invariant Hamiltonian, i.e., each local interaction term in H is the same. As a result, each of the exponentials $e^{-\beta H_i/r}$ has the same decomposition in a given basis with an associated parameter γ . Now, the Trotter approximation causes the expectation value $\langle A \rangle_{\mathcal{P}}$ to differ from the target imaginary-time expectation $\langle A \rangle_{\mathcal{T}}$, but this difference can be bounded using known results [67]. We show in Appendix E that for any observable A ,

$$|\text{tr}[A\mathcal{T}(\rho)] - \text{tr}[A\mathcal{P}(\rho)]| = \mathcal{O}\left(\frac{\beta^2 L^2}{r} e^{\beta L/r}\right) \quad (30)$$

$$|\text{tr}[\mathcal{T}(\rho)] - \text{tr}[\mathcal{P}(\rho)]| = \mathcal{O}\left(\frac{\beta^2 L^2}{r} e^{\beta L/r}\right) \quad (31)$$

again assuming $\|A\| \leq 1$. Note that we can choose

$$r = \mathcal{O}\left(\frac{\beta^2 L^2}{\epsilon}\right) \quad (32)$$

to make the error on the RHS at most $\mathcal{O}(\epsilon)$. We know from Lemma 2 that Algorithm 2 can estimate the expectation value $\langle A \rangle_{\mathcal{P}} = \frac{\text{tr}(A\mathcal{P}(\rho))}{\text{tr}(\mathcal{P}(\rho))}$ to within error

$$\frac{2\epsilon}{|\text{tr}[\mathcal{P}(\rho)]|} \left(1 + \frac{\epsilon}{|\text{tr}[\mathcal{P}(\rho)]|}\right) \quad (33)$$

using

$$\mathcal{O}\left(\frac{\gamma^{2Lr}}{\epsilon^2}\right) \quad (34)$$

circuits where each circuit consists of at most $\mathcal{O}(Lr)$ gates. Now, the algorithm is suitable only when the normalization $\text{tr}(\mathcal{P}(\rho))$ is much larger than the target precision ϵ . In that case, for the choice of r in Eq. (32), it follows that

$$|\langle A \rangle_{\mathcal{T}} - \langle A \rangle_{\mathcal{P}}| \leq \mathcal{O}(\epsilon), \quad (35)$$

and moreover, the error in estimating $\langle A \rangle_{\mathcal{P}}$ via Eq. (33) can be upper-bounded by

$$\frac{4\epsilon}{|\text{tr}[\mathcal{T}(\rho)]|} \left(1 + \frac{4\epsilon}{|\text{tr}[\mathcal{T}(\rho)]|}\right). \quad (36)$$

Even though the cost as stated scales as γ^{2Lr} , we note that the factor γ depends on the local exponential $e^{-\beta H_i/r}$ and is therefore likely to decrease as we increase the Trotter number r . Moreover, our error bounds above can likely be further tightened, in particular because we appeal to standard additive-error bounds for Trotter formula, which may not be well-suited for non-trace-preserving imaginary-time evolution. It may also be possible to improve the bounds by using multiplicative-error estimates for Trotter-Suzuki formula [67,68]. We leave a detailed analysis of the complexity of our algorithm to future work.

3.2.1. Estimating Thermal Expectation Values

We now show how our QPD implementation of imaginary-time evolution can be used to compute expectation values of observables in the Gibbs state. Recall that the Gibbs state is defined to be

$$\rho_{\beta} = \frac{e^{-\beta H}}{\mathcal{Z}_{\beta}}, \quad (37)$$

where H is the Hamiltonian of the quantum system, β is the inverse temperature, and $\mathcal{Z}_{\beta} = \text{tr}(e^{-\beta H})$ is the partition function. Preparing the Gibbs state and estimating thermal expectation values for general local Hamiltonians are known to be difficult tasks, and therefore unlikely to admit efficient algorithms in general [14].

Instead of directly preparing the (mixed) Gibbs state, we seek to prepare a *thermal pure quantum state* (TPQ) [69]. The latter is a suitably constructed n -qubit pure state $|\psi\rangle$ which, with high probability, can be used to estimate the Gibbs state expectation values for a set of observables, up to a tunable precision ϵ . Formally, the TPQ state for a given set of observables $\{\hat{O}\}$, is defined as a pure state $|\psi\rangle$ that satisfies

$$\Pr\left[\langle\psi|\hat{O}|\psi\rangle - \text{tr}(\rho_\beta\hat{O}) \geq \epsilon\right] \leq f(N, \epsilon), \quad (38)$$

where $\lim_{N \rightarrow \infty} f(N, \epsilon) = 0$. [70,71] It is known that applying imaginary-time evolution to a state chosen uniformly at random from the Haar measure gives a TPQ state [71–73]. Since Haar-random states cannot be efficiently prepared on a quantum computer, we follow the approach of Ref. [74] where the initial Haar-random state is replaced by a random 3-design. Specifically, the authors in Ref. [74] apply a random Clifford operator to a computational basis state followed by imaginary-time evolution, preparing the state,

$$|\psi_\beta\rangle = \frac{e^{-\beta H/2} U |0\rangle}{\sqrt{\langle 0|U^\dagger e^{-\beta H} U|0\rangle}}, \quad (39)$$

where U is the random Clifford operator, and the state $|\psi_\beta\rangle$ is in normalized form. Note that a random n -qubit Clifford operator can be implemented using $\mathcal{O}(n^2)$ one- and two-qubit gates [75], a significant gain over the exponential number of such gates required to generate a Haar-random unitary. Leveraging the 3-design property of the Clifford group, it can be shown [74,76] that the above state (Eq (39)) is a valid TPQ state satisfying the definition in Eq. (38).

Ref. [74] used algorithms based on quantum signal processing to implement the ITE operation to prepare the Clifford-random TPQ. Here, we implement ITE using our quasi-probabilistic approach based on applying Algorithm 2 to a Trotterized approximation of ITE. We note that preparing TPQ states is a highly suitable application of Algorithm 2 because the TPQ state error rapidly approaches zero as n becomes large [74]. In addition, for a k -local n -qubit Hamiltonian, all that is needed is the QPDs of the local terms, which are much more tractable to compute than the QPD of the entire n -qubit Hamiltonian. The method we outline here can be used to estimate Gibbs-state expectation values for any k -local Hamiltonian at a given temperature.

Numerical Simulations

We test the performance of our technique in this case by simulating the preparation of TPQ states using Algorithm 2 and comparing their performance with that of TPQ states prepared using ideal ITE. Specifically, we consider TPQ states of the form $|\psi_i\rangle = e^{-0.02H} U_i |0\rangle^{\otimes n}$, where H is the 1D Heisenberg Hamiltonian on $n \in \{4, \dots, 8\}$ qubits, and U_i is a randomly chosen Clifford unitary. We generate 10 such random states and average the error in the thermal expectation value, $|\langle\psi_i|O_j|\psi_i\rangle - \text{tr}(|\psi_\beta\rangle\langle\psi_\beta|O_j)|$, over 30 randomly chosen Pauli observables, O_j . For each simulated TPQ state, we use $N \in \{1024, 9400, 51200, 409600, 1638400\}$ QPD samples for $n \in \{4, 5, 6, 7, 8\}$ qubits, respectively. The results are shown in Figure 1 and show good agreement between our simulation and states prepared using ideal ITE.

4. Comparison with Prior Work

As highlighted in the Introduction, there has been extensive work on quantum algorithms for imaginary-time evolution and thermal state preparation. Here we summarize relevant prior work and highlight the salient features of our algorithm that distinguish it from those.

4.1. Imaginary-Time Evolution

Algorithms based on techniques such as block-encoding and quantum signal processing [28,30,32,60,77] have rigorous performance guarantees but usually require complicated controlled-unitaries and a large number of ancilla qubits. The need for ancilla qubits in these algorithms arises because the non-unitary ITE operation has to be embedded into a unitary quantum circuit. The well-known QITE algorithm of Ref. [34] takes a different path to dealing with this nonunitarity. QITE approximates the ITE as a sequence of local non-unitary operations, $\exp(-\beta H) \approx \prod_{j=1}^J T_j$, and then embeds each local operation T_j into a unitary U_j that acts on a slightly larger subsystem than the T_j it embeds. As a result, the QITE algorithm does not require any ancilla qubits, but determining the unitary U_j requires performing intermediate measurements on the state prepared at the $(j-1)$ th step. Like QITE, our algorithm also requires no ancilla qubits and proceeds by approximating $\exp(-\beta H)$ as a product of local operators. The key difference

is that we implement the local operations T_j by decomposing them into linear combination of operations with the *same* locality. In addition, our algorithm is nonadaptive and can be implemented without intermediate measurements. Unlike the block-encoding algorithms mentioned above, we do not require complicated controlled-unitary operations. We note that a number of alternative ways of effecting imaginary-time evolution with quantum circuits have been proposed in the literature [38,39,78], which also share similar benefits of requiring few to no ancilla qubits. We leave a detailed comparison with these algorithms to future work.

4.2. Thermal State Expectation Values

Our proposed method for evaluating thermal expectation values enjoys the same feature as our ITE algorithm of requiring no ancillary qubits. In contrast, quantum algorithms for Gibbs-state preparation based on block-encoding techniques require at least n ancilla qubits to prepare the Gibbs state of an n -qubit Hamiltonian [25,28,30,77]. Ref. [74] used the TPQ idea with block-encoding algorithms to reduce the number of ancilla qubits from n to a constant for the purpose of estimating thermal expectation values. Using our quasiprobabilistic algorithm for performing ITE instead of block-encoding eliminates the need for any ancillary qubits. We note that a number of recent works have investigated reducing quantum resource requirements for Gibbs-state preparation and partition function evaluation. These include the Gibbs sampling algorithm of Ref. [79], which also requires no ancilla qubits. Other algorithms have combined randomized ideas with block-encoding techniques to reduce ancilla count [37,80].

There has been a large amount of work in recent years on quantum algorithms that prepare Gibbs states by simulating Lindbladian evolution [81–85]. These methods appear promising as they have been shown to prepare certain Gibbs states in polynomial time [86–88]. However, with few exceptions such as Refs. [83,84], these generally require complex subroutines and are thus more suitable for fault-tolerant quantum computers.

Among feasible near-term implementations, there are a number of variational algorithms for Gibbs-state preparation [89–92], which minimize a cost function over parametrized quantum circuits to prepare the Gibbs state. Even though the relatively simple circuits make them attractive for experimentation, the evaluation of cost functions often require a large number of measurements. Moreover, convergence can be hampered by phenomenon such as barren plateaus [93].

We also emphasize that our method only gives access to expectation values of observables in the Gibbs state, and not the Gibbs state itself. Most of the algorithms we mentioned above are perhaps more powerful in that they give access to the state itself. However, our method can prove adequate for most physics applications where the goal would be to extract observables of interest.

5. Conclusions

In this work, we propose using quasiprobability decompositions to implement imaginary time evolution, a useful primitive in many quantum algorithms. This “algorithmic” use of quasiprobability decompositions inherits the error mitigation properties of related techniques such as probabilistic error cancellation. Our method also extends to implementing linear Hermiticity-preserving maps, of which imaginary time evolution is one example. We provide algorithms for estimating the expectation values of observables with respect to states evolved under one or more applications of a given map. Using our algorithms, we can estimate the expectation value of an observable for imaginary-time evolved states. This has many applications including preparing thermal pure quantum states, which can be used to estimate the Gibbs state expectation value of a set of observables. We simulate the performance of our method for this task, and find good agreement with the ideal imaginary time evolution operation. Our method is well-suited to near-term hardware, as it requires only 1- and 2-qubit gates and measurement, and the quasiprobability decomposition can be performed over a device’s native (noisy) gateset. As a proof-of-principle demonstration, we implement imaginary time evolution using a superconducting qubit processor (provided by IBM), and use this to estimate the ground state energy of the 2-qubit Heisenberg Hamiltonian.

A natural next step will be to extend our proof-of-principle demonstration on hardware to Hamiltonians on more than two qubits. Although this will entail increased sampling costs, when one considers a more specific use-case, namely, of thermal pure quantum states to estimate (thermal) expectation values, the error in the expectation value decays rapidly with the number of qubits, a promising sign for application of our method on near-term devices. There are two hardware-level optimizations that would make the methods presented in this work more efficient. First, in our demonstration using IBM hardware, if a mid-circuit measurement gives the incorrect result, i.e., post-selection is unsuccessful, then we still execute the remainder of the circuit because of the limitations of the measurement-based feedforward. If instead one has the ability to apply quantum circuits based on the results of mid-circuit measurements,

then this waste of resources can easily be avoided. Second, sampling all of the required circuits in software and then loading each circuit individually into hardware is highly inefficient. If instead the sampling is implemented at the level of the field-programmable gate-array control system (as was recently demonstrated for randomized compiling [94]), then we expect that the overall runtime of the circuits will be significantly reduced.

Author Contributions

Project conception by A.C. Simulations by A.R., E.S., N.C., and M.V. Hardware implementation by A.R. Analysis and proofs by A.R., N.C., and A.C. Project management and coordination by N.C., M.V., and A.C. Manuscript prepared by all the authors. All authors have read and agreed to the published version of the manuscript.

Funding

NC and AC were partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant RGPIN-2019-04198. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities.

Conflicts of Interest Statement

The authors declare no conflicts of interest.

Data Availability Statement

The data pertaining to this study will be shared upon reasonable request.

Acknowledgments

The authors thank David Gosset, Joel Klassen, Raymond Laflamme, and Christophe Piveteau for enlightening discussions. This research was enabled in part by support provided by Compute Ontario and the Digital Research Alliance of Canada. This work was completed while AR was a doctoral candidate at the University of Waterloo. Part of this research was conducted while AC was affiliated with the University of Waterloo.

Appendix A. Choi Representation and Decompositions for Linear Maps

The Choi representation [95,96] (a.k.a Choi–Jamiołkowski isomorphism) captures many important properties of linear maps, hence plays a vital role in quantum information science [97]. The Choi representation $C(\mathcal{T})$ of a linear map $\mathcal{T} : \mathbb{B}_{\mathcal{H}_1} \rightarrow \mathbb{B}_{\mathcal{H}_2}$ is defined as

$$C(\mathcal{T}) := \mathcal{T} \otimes I(|\psi_+\rangle \langle \psi_+|), \quad (\text{A1})$$

where $\mathbb{B}_{\mathcal{H}_1}$ and $\mathbb{B}_{\mathcal{H}_2}$ are bounded operators on the Hilbert space $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, $|\psi_+\rangle = \sum_{i=1}^d |ii\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_1$ is an unnormalized maximally entangled state. The map $\mathcal{T}$ is Hermiticity-preserving (HP) iff $C(\mathcal{T})$ is a Hermitian matrix. The map $\mathcal{T}$ is completely positive (CP) iff $C(\mathcal{T}) \geq 0$ (i.e., $C(\mathcal{T})$ is positive semi-definite) [97].

Choi representation $C(\mathcal{T})$ preserves the linear combination of maps, meaning that

$$C(\mathcal{T}) = \sum_i q_i C(\mathcal{B}_i)$$

if $\mathcal{T} = \sum_i q_i \mathcal{B}_i$. Therefore, the QPD decomposition can be done within the Choi representation. For HP linear maps, and the coefficients q_i can be chosen to be real numbers with a Hermitian basis. Note that CP maps are a subset of all HP maps. Since ITE is not only HP, it is also CP. Therefore, the basis elements, $\{\mathcal{B}_i\}$, that we use for demonstrations (see Table A1) are all CP and the coefficients q_i are all real.

The basic idea of QPD is decomposing the desired operation $\mathcal{T}$ as a linear combination of other physically realizable operations $\{\mathcal{B}_j\}$, then sampling each operation $\mathcal{B}_j$ according to the probability $p_j = q_j / \sum_i |q_i|$ calculated from coefficients $\{q_i\}$. The “free operations,” i.e., $\gamma = \sum_i |q_i| = 1$, are the convex combination $\text{Cov}(\mathcal{B}_j)$ of $\{\mathcal{B}_j\}$. For example, the free operations for the 2 fixed operations $\mathcal{B}_1$ and $\mathcal{B}_2$ would be the line segment between them, and the operation they can express is any operation on the line. The choice of $\{\mathcal{B}_j\}$ is critical regarding the implementation of QPD.

One can choose the physical operations $\{\mathcal{B}_j\}$ to be a fixed finite set, then the QPD method can achieve all linear combinations of the basis. For an n -qubit linear operation, the degree of freedom is 4^{2n} . Hence, a linearly independent set of bases with cardinality 4^{2n} can express all linear operations for n qubits. For example, Table A1, proposed in [65], is complete for single-qubit operations, and the tensor product of these bases is also complete for more qubits. The basis in Table A1 comprises 10 unitaries and 6 measurements. The unitary operators are generated by S gate and Hadamard gate H , and are thus Clifford. The measurements are generated by S , H and projection P_0 . They are all relatively easy and directly implementable on hardware and relatively easy concerning classical simulation.

Table A1. EBL basis [43]. A complete basis for single-qubit linear maps that contains 10 trace-preserving elements ($[1]$ to $[R_{XY}]$) and 6 non-trace-preserving elements ($[\pi_X]$ to $[\pi_{XY}]$), where we use the notation $[U](\cdot) := U(\cdot)U^\dagger$. The map $[P_0]$ refers to projection onto the state $|0\rangle$.

$[1]$				
$[\sigma^X]$			$=$	$[H][S]^2[H]$
$[\sigma^Y]$			$=$	$[H][S]^2[H][S]^2$
$[\sigma^Z]$			$=$	$[S]^2$
$[R_X]$	$=$	$[\frac{1+i\sigma^X}{\sqrt{2}}]$	$=$	$[H][S]^3[H]$
$[R_Y]$	$=$	$[\frac{1+i\sigma^Y}{\sqrt{2}}]$	$=$	$[S][H][S]^3[H][S]^3$
$[R_Z]$	$=$	$[\frac{1+i\sigma^Z}{\sqrt{2}}]$	$=$	$[S]^3$
$[R_{YZ}]$	$=$	$[\frac{\sigma^Y+\sigma^Z}{\sqrt{2}}]$	$=$	$[H][S]^3[H][S]^2$
$[R_{ZX}]$	$=$	$[\frac{\sigma^Z+\sigma^X}{\sqrt{2}}]$	$=$	$[S]^3[H][S]^3[H][S]^3$
$[R_{XY}]$	$=$	$[\frac{\sigma^X+\sigma^Y}{\sqrt{2}}]$	$=$	$[H][S]^2[H][S]^3$
$[\pi_X]$	$=$	$[\frac{1+\sigma^X}{2}]$	$=$	$[S][H][S][H][P_0][H][S]^3[H][S]^3$
$[\pi_Y]$	$=$	$[\frac{1+\sigma^Y}{2}]$	$=$	$[H][S]^3[H][P_0][H][S][H]$
$[\pi_Z]$	$=$	$[\frac{1+\sigma^Z}{2}]$	$=$	$[P_0]$
$[\pi_{YZ}]$	$=$	$[\frac{\sigma^Y+i\sigma^Z}{2}]$	$=$	$[S][H][S][H][P_0][H][S][H][S]^3$
$[\pi_{ZX}]$	$=$	$[\frac{\sigma^Z+i\sigma^X}{2}]$	$=$	$[H][S]^3[H][P_0][H][S][H][S]^2$
$[\pi_{XY}]$	$=$	$[\frac{\sigma^X+i\sigma^Y}{2}]$	$=$	$[P_0][H][S]^2[H]$

The number of degrees of freedom in the problem can be reduced with constraints on the map $\mathcal{T}$. For instance, knowing $\mathcal{T}$ is TP reduces the degrees of freedom to $d^4 - d^2 + 1$. In [64], the author proposed a new basis set (see Table A2) for 2-qubit CPTP maps, which reduced the number of basis from 256 to 241. Note that, as shown in Figure 4, decomposing the imaginary time evolution according to the EBL basis has a higher sampling cost compared to the Takagi basis. It is not only because the Takagi basis contains more entangling operations. As mentioned above, the geometry of the basis sets also affects the cost.

Table A2. Takagi Basis [64]. A complete basis for 2-qubit CPTP maps. $\{\mathcal{B}_i\}_{i=1}^{13}$ are the first 13 elements of the EBL basis (A1). $\mathcal{CX}, \mathcal{CS}, \mathcal{CH}, \mathcal{C}_H\mathcal{X}$ are channel versions of CNOT, controlled-phase, controlled-Hadamard, and NOT controlled with ± 1 eigenstates of the Hadamard gate, respectively. $\mathcal{K}_i$ is the channel version of $K := SH$ acting on qubit i . $\mathcal{S}_W$ and $i\mathcal{S}_W$ are channel versions of the SWAP and iSWAP gates. $\mathcal{U}$ conjugated by $\mathcal{V}$ denotes $\mathcal{V}^\dagger \circ \mathcal{U} \circ \mathcal{V}$ and “ $\mathcal{U}$ + conjugation with $\mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$ ” collects the nine conjugations of $\mathcal{U}$ by $I_{12}, \mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_1^\dagger, \mathcal{K}_2^\dagger, \mathcal{K}_1 \otimes \mathcal{K}_2, \mathcal{K}_1 \otimes \mathcal{K}_2^\dagger, \mathcal{K}_1^\dagger \otimes \mathcal{K}_2,$ and $\mathcal{K}_1^\dagger \otimes \mathcal{K}_2^\dagger$.

$\mathcal{B}_1\text{--}\mathcal{B}_{169}$	$\{\mathcal{B}_i\}_{i=1}^{13} \otimes \{\mathcal{B}_i\}_{i=1}^{13}$
$\mathcal{B}_{170}\text{--}\mathcal{B}_{178}$	$\mathcal{CX} + \text{conjugation with } \mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$
$\mathcal{B}_{179}\text{--}\mathcal{B}_{187}$	$\mathcal{X}_1 \circ \mathcal{CX} \circ \mathcal{X}_1 + \text{conjugation with } \mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$
$\mathcal{B}_{188}\text{--}\mathcal{B}_{196}$	$\mathcal{CS} + \text{conjugation with } \mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$
$\mathcal{B}_{197}\text{--}\mathcal{B}_{205}$	$\mathcal{CH} + \text{conjugation with } \mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$
$\mathcal{B}_{206}\text{--}\mathcal{B}_{214}$	$\mathcal{C}_H\mathcal{X} + \text{conjugation with } \mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$
$\mathcal{B}_{215}\text{--}\mathcal{B}_{223}$	$\mathcal{CX} \circ \mathcal{H}_1 + \text{conjugation with } \mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$
$\mathcal{B}_{224}\text{--}\mathcal{B}_{226}$	$\mathcal{S}_W + \text{conjugation with } \mathcal{K}_2, \mathcal{K}_2^\dagger$
$\mathcal{B}_{227}\text{--}\mathcal{B}_{232}$	$i\mathcal{S}_W + \text{conjugation with } \mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$
$\mathcal{B}_{233}\text{--}\mathcal{B}_{241}$	$\mathcal{S}_W \circ \mathcal{H}_1 + \text{conjugation with } \mathcal{K}_{1,2}, \mathcal{K}_{1,2}^\dagger$

On the other hand, one can also try decomposing $\mathcal{T}$ to more general sets of physical operations. It is known that any HPTP map $\mathcal{T}$ can be decomposed to a linear combination of two CPTP maps $\mathcal{T} = q_1 \mathcal{T}_1 - q_2 \mathcal{T}_2$ [98]. Similarly, a HP map can be decomposed by the subtraction of two CP maps. This also allows us to implement QPD without fixing

a set of basis operations [98]. Similarly, allowing general quantum instruments can also empower us to implement QPD for broader classes of maps. In [62], the authors proved the optimal cost of QPD will be achieved by allowing even one quantum instrument. The optimal sampling cost γ for implementing with measurement-controlled post-processing (a generalized QPD method) equals the diamond norm $\|\mathcal{T}\|_\diamond$ of the desired map $\mathcal{T}$ [62]. We analyze the optimal decomposition of ITE in Appendix C.

Appendix B. Basics of Imaginary Time Evolution

The imaginary time evolution without renormalization is a CP non-TP map,

$$\mathcal{T}(\rho) = e^{-\beta H} \rho e^{-\beta H},$$

where $\beta > 0$ and $\rho \in \mathcal{D}(\mathcal{H}^d)$.

Diagonalizing the ITE operator

$$e^{-\beta H} = U \text{diag}[e^{-\beta \lambda_i}] U^\dagger, \quad (\text{A2})$$

where $\text{diag}[e^{-\beta \lambda_i}]$ is a diagonal matrix and $\{\lambda_i\}$ are eigenvalues of H with the increasing order (λ_0 is the ground state energy). It is a convention to assume that H is positive semi-definite (i.e., the $\lambda_i \geq 0$). However, we need to be slightly careful when working with non-unitary evolution.

From Eq. (A2), when $\beta \rightarrow \infty$, if $\lambda_0 = 0$, the ITE operator is the projection onto the space of the ground states; if $\lambda_0 > 0$, the limit $\lim_{\beta \rightarrow \infty} e^{-\beta H}$ will become a zero matrix; if H has negative eigenvalues, then $\lim_{\beta \rightarrow \infty} e^{-\beta H}$ is an unbounded operator that projects onto the ground space and multiplies the unbounded scalar $e^{-\beta \lambda_0}$.

Appendix C. Optimal Sampling Cost for ITE

If we allow ourselves the ability to optimize over the choice of the basis elements $\{\mathcal{B}_i\}$, we can upper bound the cost of implementing $e^{-\beta H}$ in terms of β and the operator norm of H . The diamond norm $\|\mathcal{T}\|_\diamond$ of the HP map $\mathcal{T}$ gives a lower bound for the QPD sampling cost, which is also the optimal cost of a generalized QPD implementation [62]. In the following, we analyze the diamond norm for the imaginary time evolution.

The imaginary time evolution without renormalization is a CP non-TP map,

$$\mathcal{T}(\rho) = e^{-\beta H} \rho e^{-\beta H},$$

where β is fixed and $\beta > 0$ and $\rho \in \mathcal{D}(\mathcal{H}^d)$. The diamond norm of $\mathcal{T}$ is

$$\|\mathcal{T}\|_\diamond := \sup_{n \geq 1} \|\mathcal{T} \otimes \mathbb{I}_n\|_{tr},$$

where $\|\mathcal{T} \otimes \mathbb{I}_n\|_{tr} = \max_\rho \|\mathcal{T} \otimes \mathbb{I}_n(\rho)\|_{tr}$, and $\rho \in \mathcal{D}(\mathcal{H}^d \otimes \mathcal{H}^n)$. Therefore,

$$\|\mathcal{T}\|_\diamond = \sup_{n \geq 1} \max_\rho \|\mathcal{T} \otimes \mathbb{I}_n(\rho)\|_{tr} = \max_{|\psi\rangle} \|\mathcal{T} \otimes \mathbb{I}_d(|\psi\rangle \langle \psi|)\|_{tr},$$

where d is the dimension of the system, and now $\rho \in \mathcal{D}(\mathcal{H}^d \otimes \mathcal{H}^d)$. The second equality is due to the convexity of trace norm and Schmidt decomposition.

Perform the spectral decomposition of the Hamiltonian H , $H = U \Sigma U^\dagger$. For imaginary time evolution $\mathcal{T}(\rho) = e^{-\beta H} \rho e^{-\beta H}$, its diamond norm is

$$\begin{aligned} \|\mathcal{T}\|_\diamond &= \max_{|\psi\rangle} \left\| [(U e^{-\beta \Sigma} U^\dagger) \otimes \mathbb{I}_d](|\psi\rangle \langle \psi|) [(U e^{-\beta \Sigma} U^\dagger) \otimes \mathbb{I}_d] \right\|_{tr} \\ &= \max_{|\psi\rangle} \text{tr}(\sqrt{|\phi\rangle \langle \phi| |\phi\rangle \langle \phi|}) \\ &= \max_{|\psi\rangle} a_\phi \end{aligned} \quad (\text{A3})$$

where $|\phi\rangle := [(Ue^{-\beta\Sigma}U^\dagger) \otimes \mathbb{I}_d] |\psi\rangle$ (note that $|\phi\rangle$ is unnormalized since $\mathcal{T}$ is not trace-preserving) and $a_\phi := \langle\phi|\phi\rangle$. Therefore, the diamond norm of ITE is

$$\|\mathcal{T}\|_\diamond = \max_{|\phi\rangle} a_\phi = e^{-2\beta\lambda_0}, \quad (\text{A4})$$

where λ_0 is the smallest eigenvalue of H .

If H is positive definite (PD), the eigenvalues $\lambda_i > 0$, $\|\mathcal{T}\|_\diamond = e^{-2\beta\lambda_0}$. As β increases, the optimal sampling cost γ_{opt} gets smaller; if H is positive semi-definite (PSD), $\lambda_0 = 0$, $\|\mathcal{T}\|_\diamond = 1$, i.e. the optimal cost γ_{opt} will not change with β ; if H is indefinite, the smallest eigenvalue λ_0 is negative, then the cost will increase as β increases.

Based on our analysis, introducing a scaled identity operator into the Hamiltonian appears to improve the sampling cost. However, since imaginary-time evolution is not trace-preserving, the signal strength $\text{tr}(\mathcal{T}(\rho))$ may not be maintained. In the following, we study the balance between sampling cost and signal strength.

Proposition A1. *Given a fixed Hamiltonian H and any given inverse temperature β , the trace $\text{tr}(\mathcal{T}(\rho))$ of the output state from ITE is upper bounded by the diamond norm $\|\mathcal{T}\|_\diamond$, and lower bounded by the rescaled overlap $\text{tr}(\rho_{gs}\rho)$ between the input state ρ and ground state ρ_{gs} of H (with ground state energy λ_0), i.e.*

$$\|\mathcal{T}\|_\diamond \geq \text{tr}(\mathcal{T}(\rho)) \geq e^{-2\beta\lambda_0} \text{tr}(\rho_{gs}\rho). \quad (\text{A5})$$

Proof. The first inequality is straightforward from the definition of the diamond norm. The second inequality is also straightforward:

$$\begin{aligned} \text{tr}(\mathcal{T}(\rho)) &= \text{tr}(e^{-\beta H} \rho e^{-\beta H}) \\ &= \sum_i e^{-2\beta\lambda_i} \langle i | U^\dagger \rho U | i \rangle \end{aligned} \quad (\text{A6})$$

Every term in Eq. (A6) is non-negative, and $e^{-2\beta\lambda_0} \text{tr}(\rho_{gs}\rho)$ is only one term in the summation. $\square$

Since we know that $\|\mathcal{T}\|_\diamond = e^{-2\beta\lambda_0}$, Eq. (A5) can be rewritten as

$$1 \geq e^{2\beta\lambda_0} \text{tr}(\mathcal{T}(\rho)) \geq \text{tr}(\rho_{gs}\rho). \quad (\text{A7})$$

Similar to the diamond norm $\|\mathcal{T}\|_\diamond$, the signal strength $\text{tr}(\mathcal{T}(\rho))$ is also affected by the scalar $e^{-2\beta\lambda_0}$. When $\lambda_0 > 0$, as β increases, the signal also decreases. Any (unnormalized) expectation $\text{tr}(\mathcal{T}(\rho)A)$ for a bounded observable A will vanish as β approaches infinity. Therefore, the sampling cost for such values with a fixed precision ϵ also approaches zero. Conversely, when $\lambda_0 < 0$, we get a signal amplification, causing an increase in sampling cost. We can avoid this issue when $\lambda_0 = 0$, i.e., H is PSD. Therefore, the optimal adjustment for adding a scaled identity term αI , where $\alpha \in \mathbb{R}$, to the Hamiltonian H is thus the one that shifts the ground-state energy to zero.

Determining the appropriate scaling factor α can be challenging in general, given that λ_0 is unknown. However, for a frustration-free Hamiltonian, a global rescaling can be obtained by adjusting each local term accordingly. When implementing the global ITE via Trotterization into local components, each local evolution step drives the state towards the ground state of that local term. Because the global Hamiltonian is frustration-free, this process ensures that the signal strength is maintained throughout the entire procedure. This does not extend to frustrated Hamiltonians, thereby constraining the general applicability of our method. Nonetheless, it is important to note that frustration-free Hamiltonians can still encode computationally hard problems. A notable example is the Feynman–Kitaev mapping [12], which can map complex computational problems into such Hamiltonians.

Appendix D. Estimating Expectation Values: Proofs

Proof of 1. Recall Eq. (1)

$$\text{tr}[A\mathcal{T}(\rho)] = \sum_{i=1}^D \text{sgn}(q_i) \cdot |q_i| \cdot \text{tr}[A\mathcal{B}_i(\rho)]. \quad (\text{A8})$$

Let I be a random variable that takes values in $\{1, \dots, D\}$ with probability

$$\Pr[I = i] = \frac{q_i}{\gamma}. \quad (\text{A9})$$

Let $L_I \in \{0, 1\}$ be the indicator variable that is 1 when the operation $\mathcal{B}_I$ is successfully implemented, and let Y_I denote the outcome of measuring the observable A . Note that

$$\Pr[L_i = 1] = 1, \text{ if } \mathcal{B}_i \text{ is trace preserving,} \quad (\text{A10})$$

$$\Pr[L_i = 1] = \text{tr}[\mathcal{B}_i(\rho)] \leq 1, \text{ if } \mathcal{B}_i \text{ is non-trace preserving.} \quad (\text{A11})$$

Now, we define the random variables

$$F = \gamma \text{sgn}(q_I) L_I Y_I \quad G = \gamma \text{sgn}(q_I) L_I. \quad (\text{A12})$$

Computing their expectations below as

$$\mathbb{E}[F] = \gamma \sum_{i=1}^D \sum_{L_i} \mathbb{E} \text{sgn}(q_i) \Pr[I = i] \Pr[L_i = 1] Y_i \quad (\text{A13})$$

$$= \gamma \sum_{i=1}^D \frac{|q_i|}{\gamma} \text{sgn}(q_i) \text{tr}[A \mathcal{B}_i(\rho)] \quad (\text{A14})$$

$$= \text{tr}[A \mathcal{T}(\rho)] \quad (\text{A15})$$

and

$$\mathbb{E}[G] = \gamma \sum_{i=1}^D \sum_{L_i} \text{sgn}(q_i) \Pr[I = i] \Pr[L_i = 1] \quad (\text{A16})$$

$$= \gamma \sum_{i=1}^D \frac{|q_i|}{\gamma} \text{sgn}(q_i) \text{tr}[\mathcal{B}_i(\rho)] \quad (\text{A17})$$

$$= \text{tr}[\mathcal{T}(\rho)], \quad (\text{A18})$$

we see that the rescaled expectation value is given by

$$\langle A \rangle_{\mathcal{T}} = \frac{\mathbb{E}[F]}{\mathbb{E}[G]} \quad (\text{A19})$$

proving the correctness of Algorithm 1.

We use standard concentration inequalities to bound the number of measurements needed to estimate $\langle A \rangle_{\mathcal{T}}$ within a given precision. We note that both F and G lie between $\pm \gamma$. For N independent trials (each trial consists of sampling an index i , running the corresponding quantum circuit and accepting/rejecting, and measuring A), define $\bar{F} = \frac{1}{N}(F_1 + F_2 + \dots + F_N)$ and $\bar{G} = \frac{1}{N}(G_1 + G_2 + \dots + G_N)$. Then, Hoeffding's inequality implies

$$\Pr[|\bar{F} - \mathbb{E}[F]| \geq \epsilon] \leq \exp\left(-\frac{N\epsilon^2}{2\gamma^2}\right) \quad (\text{A20})$$

$$\Pr[|\bar{G} - \mathbb{E}[G]| \geq \epsilon] \leq \exp\left(-\frac{N\epsilon^2}{2\gamma^2}\right). \quad (\text{A21})$$

Choosing

$$N = \frac{2\gamma^2}{\epsilon^2} \log(1/\delta) \quad (\text{A22})$$

ensures that the RHS of Eqs (A20) and (A21) are at most δ . Then, we have with probability at least $(1 - \delta)$

$$\left| \frac{\bar{F}}{\bar{G}} - \frac{\mathbb{E}[F]}{\mathbb{E}[G]} \right| = \left| \frac{\mathbb{E}[G](\bar{F} - \mathbb{E}[F]) + \mathbb{E}[F](\mathbb{E}[G] - \bar{G})}{\bar{G}\mathbb{E}[G]} \right| \quad (\text{A23})$$

$$\leq \left| \frac{\mathbb{E}[G](\bar{F} - \mathbb{E}[F])}{\bar{G}\mathbb{E}[G]} \right| + \left| \frac{\mathbb{E}[F](\mathbb{E}[G] - \bar{G})}{\bar{G}\mathbb{E}[G]} \right| \quad (\text{A24})$$

$$\leq \frac{1}{|\bar{G}|} \left(\epsilon + \epsilon \left| \frac{\mathbb{E}[F]}{\mathbb{E}[G]} \right| \right) \quad (\text{A25})$$

$$\leq \frac{2\epsilon}{|\bar{G}|} \quad (\text{A26})$$

where we have used Eqs (A15) and (A18), and that $\frac{\mathbb{E}[F]}{\mathbb{E}[G]} \leq 1$. Moreover

$$\frac{1}{|\bar{G}|} \leq \frac{1}{|\mathbb{E}[G] - \epsilon|} \leq \frac{1}{|\mathbb{E}[G]|} \left(1 + \frac{\epsilon}{|\mathbb{E}[G]|} \right). \quad (\text{A27})$$

Using Eqs (A26) and (A27) and recalling Eq. (A18) we see that

$$\left| \frac{\bar{F}}{\bar{G}} - \langle A \rangle_{\mathcal{T}} \right| \leq \frac{2\epsilon}{|\text{tr}[\mathcal{T}(\rho)]|} \left(1 + \frac{\epsilon}{|\text{tr}[\mathcal{T}(\rho)]|} \right) \quad (\text{A28})$$

with probability at least $(1 - \delta)$ for choice of N as in Eq. (A22).

□

Proof of Lemma 2. The proof essentially follows that of Lemma 1 with more involved notation. First note that for a $\mathcal{E}$ given by R repeated applications of $\mathcal{T}$, we have

$$\text{tr}[A\mathcal{E}(\rho)] = \sum_{i_1=1}^D \sum_{i_2=1}^D \cdots \sum_{i_R=1}^D \text{sgn}(q_{i_1}q_{i_2} \cdots q_{i_R}) |q_{i_1}q_{i_2} \cdots q_{i_R}| \cdot \text{tr}[A\mathcal{B}_{i_R}(\cdots \mathcal{B}_{i_2}(\mathcal{B}_{i_1}(\rho)))]. \quad (\text{A29})$$

We define I_{i_r} be a random variable that takes values in $\{1, \dots, D\}$ with probability

$$\Pr[I_{i_r} = i] = \frac{q_i}{\gamma}. \quad (\text{A30})$$

Similar to the proof of Lemma 1, for each i_r we define $L_{i_r} \in \{0, 1\}$ to be the indicator variable that is 1 when the operation $\mathcal{B}_{i_r}$ is successfully implemented. We have $L_{i_r} = 1$ if $\mathcal{B}_{i_r}$ is trace-preserving and

$$\Pr[L_{i_r} = 1] = \frac{\text{tr}[\mathcal{B}_{i_r}(\mathcal{B}_{i_{r-1}}(\cdots \mathcal{B}_{i_1}(\rho)))]}{\text{tr}[\mathcal{B}_{i_{r-1}}(\cdots \mathcal{B}_{i_1}(\rho))]}, \quad (\text{A31})$$

if $\mathcal{B}_{i_r}$ is non-trace-preserving. We let Y be the random variable denoting the measurement outcome of the observable A and define

$$F := \gamma^R \sum_{i_1=1}^D \sum_{i_2=1}^D \cdots \sum_{i_R=1}^D \text{sgn}(q_{i_1}q_{i_2} \cdots q_{i_R}) \cdot L_{i_R} \cdots L_{i_2} L_{i_1} Y \quad (\text{A32})$$

$$G := \gamma^R \sum_{i_1=1}^D \sum_{i_2=1}^D \cdots \sum_{i_R=1}^D \text{sgn}(q_{i_1}q_{i_2} \cdots q_{i_R}) \cdot L_{i_R} \cdots L_{i_2} L_{i_1}. \quad (\text{A33})$$

The expectation values of F and G are given by

$$\begin{aligned}\mathbb{E}[F] &= \sum_{i_1=1}^D \sum_{i_2=1}^D \cdots \sum_{i_R=1}^D \text{sgn}(q_{i_1} q_{i_2} \cdots q_{i_R}) |q_{i_1} q_{i_2} \cdots q_{i_R}| \cdot \text{tr} [\mathcal{B}_{i_R} (\cdots \mathcal{B}_{i_2} (\mathcal{B}_{i_1} (\rho)))] \\ &= \text{tr} [A\mathcal{E}(\rho)],\end{aligned}\tag{A34}$$

$$\begin{aligned}\mathbb{E}[G] &= \sum_{i_1=1}^D \sum_{i_2=1}^D \cdots \sum_{i_R=1}^D \text{sgn}(q_{i_1} q_{i_2} \cdots q_{i_R}) |q_{i_1} q_{i_2} \cdots q_{i_R}| \cdot \text{tr} [\mathcal{B}_{i_R} (\cdots \mathcal{B}_{i_2} (\mathcal{B}_{i_1} (\rho)))] \\ &= \text{tr} [\mathcal{E}(\rho)]\end{aligned}\tag{A35}$$

and thus

$$\langle A \rangle_{\mathcal{E}} = \frac{\mathbb{E}[F]}{\mathbb{E}[G]}\tag{A36}$$

proving correctness of Algorithm 2. Now, both F and G lie between $\pm\gamma^R$ (recall that $\|A\| \leq 1$). For N independent trials (each trial consists of sampling a multiset of indices $\{i_1, i_2, \dots, i_R\}$, running the corresponding quantum circuit and accepting/rejecting, and finally measuring A), we define $\bar{F} = \frac{1}{N}(F_1 + F_2 + \cdots + F_N)$ and $\bar{G} = \frac{1}{N}(G_1 + G_2 + \cdots + G_N)$. Then Hoeffding's inequality implies

$$\Pr[|\bar{F} - \mathbb{E}[F]| \geq \epsilon] \leq \exp\left(-\frac{N\epsilon^2}{2\gamma^{2R}}\right)\tag{A37}$$

$$\Pr[|\bar{G} - \mathbb{E}[G]| \geq \epsilon] \leq \exp\left(-\frac{N\epsilon^2}{2\gamma^{2R}}\right).\tag{A38}$$

Following the same analysis as in the proof of Lemma 1 (c.f. Eq. (A22) to Eq. (A28)), we see that choosing

$$N = \frac{2\gamma^{2R}}{\epsilon^2} \log(1/\delta)\tag{A39}$$

ensures that

$$\left| \frac{\bar{F}}{\bar{G}} - \langle A \rangle_{\mathcal{E}} \right| \leq \frac{2\epsilon}{|\text{tr}[\mathcal{E}(\rho)]|} \left(1 + \frac{\epsilon}{|\text{tr}[\mathcal{E}(\rho)]|} \right)\tag{A40}$$

with probability at least $(1 - \delta)$. $\square$

Appendix E. Trotterized ITE: Error Bounds

Here, we determine the cost of estimating imaginary-time evolved expectation values to within a given precision. We define $P = e^{-\beta H}$ and denoted its 1st-order Trotter approximation (see Eq. (28)) as $T = \left(e^{-\frac{\beta}{r} H_1} e^{-\frac{\beta}{r} H_2} \cdots e^{-\frac{\beta}{r} H_L} \right)^r$. The corresponding maps are $\mathcal{T}(\cdot)$ and $\mathcal{P}(\cdot)$ i.e., $\mathcal{P}(\cdot) = P(\cdot)P$ and $\mathcal{T}(\cdot) = T(\cdot)T$. From Theorem 6 in [67], it can be shown that

$$\|T - P\| = \mathcal{O}\left(\left(\sum_{l=1}^L \|H_l\|\beta\right)^2 e^{\sum_{l=1}^L \|H_l\|\beta}\right)\tag{A41}$$

$$= \mathcal{O}\left(\frac{\beta^2 L^2}{r} e^{\beta L/r}\right).\tag{A42}$$

For an input state ρ , we can bound the error in the expectation value $\text{tr} \left\{ \left[A e^{-\beta H} \rho e^{-\beta H} \right] \right\}$ for an observable A as below

$$\begin{aligned} |\text{tr}[AT(\rho)] - \text{tr}[AP(\rho)]| &= |\text{tr}[AT\rho T] - \text{tr}[AP\rho P]| \\ &= |\text{tr}[AT\rho T] - \text{tr}[AT\rho P] + \text{tr}[AT\rho P] - \text{tr}[AP\rho P]| \\ &= |\text{tr}[AT\rho(T-P)] + \text{tr}[A(T-P)\rho P]| \end{aligned} \quad (\text{A43})$$

$$\leq |\text{tr}[AT\rho(T-P)]| + |\text{tr}[A(T-P)\rho P]| \quad (\text{A44})$$

$$\leq \|A\| \cdot \|T\| \cdot \|(T-P)\| + \|A\| \cdot \|(T-P)\| \cdot \|P\| \quad (\text{A45})$$

$$\leq 2\|e^{-\beta H}\| \cdot \mathcal{O}\left(\frac{\beta^2 L^2}{r} e^{\beta L/r}\right) \quad (\text{A46})$$

$$= \mathcal{O}\left(\frac{\beta^2 L^2}{r} e^{\beta L/r}\right) \quad (\text{A47})$$

where we have used Eq. (A42) in the second-last step, and that H is positive semidefinite in the last step. Similarly,

$$|\text{tr}[\mathcal{T}(\rho)] - \text{tr}[\mathcal{P}(\rho)]| = \mathcal{O}\left(\frac{\beta^2 L^2}{r} e^{\beta L/r}\right) \quad (\text{A48})$$

We choose $r = \mathcal{O}\left(\frac{\beta^2 L^2}{\epsilon}\right)$ such that the error on the RHS above and thus also in Eq. (A47) are both at most $\mathcal{O}(\epsilon)$. In our case, each term H_i in the Hamiltonian is of the form $h \otimes I$, where h is a k -local interaction, meaning that we have the conditions as in Lemma 2, where we repeatedly apply the same local map to different subsets of qubits.

Appendix F. Further Implementation Details

For the demonstration on hardware shown in Figure 5, we were also able to run circuits for the fifth Trotter step. However, due to retirement of the device (5-qubit IBM systems) in November 2023, only 20000 QPD samples were collected, where the circuit corresponding to each sample was run for 512 shots. The energy estimate we computed from this data is -0.275 , which should be compared with the ideal value of -0.269 and the simulated value of -0.275 ± 0.010 .

References

1. A.J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer and P. Zoller (2022). “Practical quantum advantage in quantum simulation”. *Nature*, 607, 667–676.
2. B. Bauer, S. Bravyi, M. Motta and G. Kin-Lic Chan (2020). “Quantum algorithms for quantum chemistry and quantum materials science”. *Chemical Reviews*, 120, 12685–12717.
3. R. Orús, S. Muel and E. Lizaso (2019). “Quantum computing for finance: Overview and prospects”. *Reviews Physics*, 4, 100028.
4. A. M. Dalzell, S. McArdle, M. Berta, P. Bienias, C.-F. Chen, A. Gilyén, C. T. Hann, M. J. Kastoryano, E. T. Khabiboulline, A. Kubica, G. Salton, S. Wang and F. G. S. L. Brandão (2025). “Quantum algorithms: A survey of applications and end-to-end complexities”. *arXiv:2310.03011*.
5. Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis and A. Aspuru-Guzik (2019). “Quantum chemistry in the age of quantum computing”. *Chemical Reviews*, 119, 10856–10915.
6. S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin and X. Yuan (2020). “Quantum computational chemistry”. *Reviews of Modern Physics*, 92, 015003.
7. R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Pater, A. Fowler and H. Neven (2018). “Encoding electronic spectra in quantum circuits with linear t complexity”. *Physical Reviews X*, 8, 041015.
8. J. Lemieux, G. Duclos-Cianci, D. Sénéchal and D. Poulin (2021). “Resource estimate for quantum many-body ground-state preparation on a quantum computer”. *Physical Reviews A*, 103, 052408.
9. P. Rall (2020). “Quantum algorithms for estimating physical quantities using block encodings”. *Physical Reviews A*, 102, 022408.
10. N. M. Tubman, C. Mejuto-Zaera, J. M. Epstein, D. Hait, D. S. Levine, W. Huggins, Z. Jiang, J. R. McClean, R. Babbush, M. Head-Gordon and K. Birgitta Whaley (2018). “Postponing the orthogonality catastrophe: efficient state preparation for electronic structure simulations on quantum devices”. *arXiv:1809.05523*.

11. I. D. Kivlichan, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, W. Sun, Z. Jiang, N. Rubin, A. Fowler, A. Aspuru-Guzik, H. Neven and R. Babbush (2020). “Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via trotterization”. *Quantum*, 4, 296.
12. A. Y. Kitaev, A. Shen and M. N. Vyalyi (2002). *Classical and Quantum Computation*. American Mathematical Society.
13. S. Gharibian, Y. Huang, Z. Landau and S. Woo Shin (2015). “Quantum Hamiltonian complexity”. *Foundations and Trends in Theoretical Computer Science*, 10, 159–282.
14. S. Bravyi, A. Chowdhury, D. Gosset and P. Wocjan (2022). “Quantum Hamiltonian complexity in thermal equilibrium”. *Nature Physics*, 18, 1367–1370.
15. S. Gharibian and F. Le Gall (2023). “Dequantizing the quantum singular value transformation: hardness and applications to quantum chemistry and the quantum PCP conjecture”. *SIAM Journal on Computing*, 52, 1009–1038.
16. F. Verstraete, J. J. García-Ripoll and J. I. Cirac (2004). “Matrix product density operators: Simulation of finite-temperature and dissipative systems”. *Physical Reviews Letters*, 93, 207204.
17. M. Zwolak and G. Vidal (2004). “Mixed-state dynamics in one-dimensional quantum lattice systems: A time-dependent superoperator renormalization algorithm”. *Physical Reviews Letters*, 93, 207205.
18. F. Alexander Wolf, A. Go, I. P. McCulloch, A. J. Millis and U. Schollwöck (2015). “Imaginary-time matrix product state impurity solver for dynamical mean-field theory”. *Physical Reviews X*, 5, 041032.
19. L. Lehtovaara, J. Toivanen and J. Eloranta (2007). “Solution of time-independent Schrödinger equation by the imaginary time propagation method”. *Journal of Computational Physics*, 221, 148–157.
20. C. V. Kraus and J. I. Cirac (2010). “Generalized Hartree–Fock theory for interacting fermions in lattices: numerical methods”. *New Journal of Physics*, 12, 113004.
21. J. R. McClean and A. Aspuru-Guzik (2015). “Compact wavefunctions from compressed imaginary time evolution”. *RSC Advances*, 5, 102277–102283.
22. T. Shi, E. Demler and J. I. Cirac (2018). “Variational study of fermionic and bosonic systems with non-gaussian states: Theory and applications”. *Annals of Physics*, 390, 245–302.
23. J. R. McClean, J. A. Parkhill and A. Aspuru-Guzik (2013). “Feynman’s clock, a new variational principle, and parallel-in-time quantum dynamics”. *Proceedings of the National Academy of Sciences*, 110, E3901–E3909.
24. J. R. McClean and A. Aspuru-Guzik (2015). “Clock quantum monte carlo technique: An imaginary-time method for real-time quantum dynamics”. *Physical Reviews A*, 91, 012311.
25. D. Poulin and P. Wocjan (2009). “Sampling from the thermal quantum Gibbs state and evaluating partition functions with a quantum computer”. *Physical Reviews Letters*, 103, 220502.
26. E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino and R. L. Sugar (1990). “Sign problem in the numerical simulation of many-electron systems”. *Physical Reviews B*, 41, 9301–9307.
27. M. Troyer and U.-J. Wiese (2005). “Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations”. *Physical Reviews Letters*, 94, 170201.
28. A. N. Chowdhury and R. D. Somma (2017). “Quantum algorithms for Gibbs sampling and hittingtime estimation”. *Quantum Information & Computation*, 17, 41–64.
29. S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin and X. Yuan (2019). “Variational ansatz-based quantum simulation of imaginary time evolution”. *npj Quantum Information*, 5, 75.
30. J. van Apeldoorn, A. Gilyén, S. Gribling and R. de Wolf (2020). “Quantum SDP-solvers: Better upper and lower bounds”. *Quantum*, 4, 230.
31. Z. Holmes, G. Muraleedharan, R. D. Somma, Y. Subasi and B. Şahinoğlu (2022). “Quantum algorithms from fluctuation theorems: Thermal-state preparation”. *Quantum*, 6, 825.
32. T. L. Silva, M. M. Taddei, S. Carrazza and L. Aolita (2023). “Fragmented imaginary-time evolution for early-stage quantum signal processors”. *Scientific Reports*, 13, 18258.
33. E. T. Campbell, B. M. Terhal and C. Vuillot (2017). “Roads towards fault-tolerant universal quantum computation”. *Nature*, 549, 172–179.
34. M. Motta, C. Sun, A. T. K. Tan, M. J. O’Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão and G. Kin-Lic Chan (2019). “Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution”. *Nature Physics*, 16, 205–210.
35. T.J. Sewell, C.D. White and B. Swingle (2022). “Thermal multi-scale entanglement renormalization ansatz for variational Gibbs state preparation”. *arXiv:2210.16419*.

36. M. Benedetti, M. Fiorentini and M. Lubasch (2021). “Hardware-efficient variational quantum algorithms for time evolution”. *Physical Reviews Research*, 3, 033083.
37. T. de Lima Silva, L. Borges and L. Aolita (2024). “Partition function estimation with a quantum coin toss”. *arXiv:2411.17816*.
38. K. Hejazi, M. Motta and G. K.-L. Chan (2023). “Adiabatic quantum imaginary time evolution”. *arXiv:2308.03292*.
39. N. Gomes, A. Mukherjee, F. Zhang, T. Iadecola, C.-Z. Wang, K.-M. Ho, P. P. Orth and Y.-X. Yao (2021). “Adaptive variational quantum imaginary time evolution approach for ground state preparation”. *Advanced Quantum Technologies*, 4, 2100114.
40. J. C. Getelina, N. Gomes, T. Iadecola, P. P. Orth and Y.-X. Yao (2023). “Adaptive variational quantum minimally entangled typical thermal states for finite temperature simulations”. *SciPost Physics*, 15, 102.
41. K. Temme, S. Bravyi and J. M. Gambetta (2017). “Error mitigation for short-depth quantum circuits”. *Physical Reviews Letters*, 119, 180509.
42. Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean and T. E. O’Brien (2023). “Quantum error mitigation”. *Reviews of Modern Physics*, 95, 045005.
43. S. Endo, S. C. Benjamin and Y. Li (2018). “Practical quantum error mitigation for near-future applications”. *Physical Reviews X*, 8, 031027.
44. A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow and J. M. Gambetta (2019). “Error mitigation extends the computational reach of a noisy quantum processor”. *Nature*, 567, 491–495.
45. C. Piveteau, D. Sutter and S. Woerner (2022). “Quasiprobability decompositions with reduced sampling overhead”. *npj Quantum Information*, 8, 12.
46. K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin and F. Verstraete (2011). “Quantum metropolis sampling”. *Nature*, 471, 87–90.
47. J. J. Wallman and J. Emerson (2016). “Noise tailoring for scalable quantum computation via randomized compiling”. *Physical Reviews A*, 94, 052325.
48. E. Van Den Berg, Z. K. Mineev, A. Kandala and K. Temme (2023). “Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors”. *Nature Physics*, 19, 1116–1121.
49. H. Pashayan, J. J. Wallman and S. D. Bartlett (2015). “Estimating outcome probabilities of quantum circuits using quasiprobabilities”. *Physical Reviews Letters*, 115, 070501.
50. M. Howard and E. Campbell (2017). “Application of a resource theory for magic states to fault-tolerant quantum computing”. *Physical Reviews Letters*, 118, 090501.
51. J. R. Seddon and E. T. Campbell (2019). “Quantifying magic for multi-qubit operations”. *Proceedings of the Royal Society A*, 475, 20190251.
52. M. Heinrich and D. Gross (2019). “Robustness of magic and symmetries of the stabiliser polytope”. *Quantum*, 3, 132.
53. J. R. Seddon, B. Regula, H. Pashayan, Y. Ouyang and E. T. Campbell (2021). “Quantifying quantum speedups: Improved classical simulation from tighter magic monotones”. *PRX Quantum*, 2, 010345.
54. K. Mitarai and K. Fujii (2021). “Constructing a virtual two-qubit gate by sampling single-qubit operations”. *New Journal of Physics*, 23, 023021.
55. R. Wiersema, L. Guerini, J. F. Carrasquilla and L. Aolita (2022). “Circuit connectivity boosts by quantum-classical-quantum interfaces”. *Physical Reviews Research*, 4, 043221.
56. A. Lowe, M. Medvidović, A. Hayes, L. J. O’Riordan, T. R. Bromley, J. M. Arrazola and N. Killoran (2023). “Fast quantum circuit cutting with randomized measurements”. *Quantum*, 7, 934.
57. L. Brenner, C. Piveteau and D. Sutter (2025). “Optimal wire cutting with classical communication”. *IEEE Transactions on Information Theory*, 71, 7742–7752.
58. C. Piveteau and D. Sutter (2024). “Circuit knitting with classical communication”. *IEEE Trans. Inf. Theory*, 70, 2734–2745.
59. C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta and K. Temme (2021). “Error mitigation for universal gates on encoded qubits”. *Physical Reviews Letters*, 127, 200505.
60. D. Poulin and P. Wocjan (2009). “Sampling from the thermal quantum Gibbs state and evaluating partition functions with a quantum computer”. *Physical Reviews Letters*, 103, 220502.
61. C. Piveteau (2020). “Advanced methods for quasiprobabilistic quantum error mitigation”. Master’s thesis. ETH Zurich. Zurich.

62. X. Zhao, L. Zhang, B. Zhao and X. Wang (2025). “Power of quantum measurement in simulating unphysical operations”. *Physical Reviews Research*, 7, 013334.
63. B. Regula, R. Takagi and M. Gu (2021). “Operational applications of the diamond norm and related measures in quantifying the non-physicality of quantum maps”. *Quantum*, 5, 522.
64. R. Takagi (2021). “Optimal resource cost for error mitigation”. *Physical Reviews Research*, 3, 033178.
65. S. Endo, Q. Zhao, Y. Li, S. Benjamin and X. Yuan (2019). “Mitigating algorithmic errors in a Hamiltonian simulation”. *Physical Reviews A*, 99, 012334.
66. “About calibration jobs, IBM Q webpage”. <https://docs.quantum.ibm.com/admin/calibration-jobs>. [Online; accessed 28-Jan-2025].
67. A. M. Childs, Y. Su, M. C. Tran, N. Wiebe and S. Zhu (2021). “Theory of trotter error with commutator scaling”. *Physical Reviews X*, 11, 011020.
68. S. Bravyi and D. Gosset (2017). “Polynomial-time classical simulation of quantum ferromagnets”. *Physical Reviews Letters*, 119, 100503.
69. S. Sugiura and A. Shimizu (2013). “Canonical thermal pure quantum state”. *Physical Reviews Letters*, 111, 010401.
70. S. Sugiura and A. Shimizu (2012). “Thermal pure quantum states at finite temperature”. *Physical Reviews Letters*, 108, 240401.
71. S. Sugiura and A. Shimizu (2013). “Canonical thermal pure quantum state”. *Physical Reviews Letters*, 111, 010401.
72. J. Richter and A. Pal (2021). “Simulating hydrodynamics on noisy intermediate-scale quantum devices with random circuits”. *Physical Reviews Letters*, 126, 230501.
73. C. Powers, L. Bassman Otfelie, D. Camps and W. A. de Jong (2023). “Exploring finite temperature properties of materials with quantum computers”. *Scientific Reports*, 13, 1986.
74. L. Coopmans, Y. Kikuchi and M. Benedetti (2023). “Predicting Gibbs-state expectation values with pure thermal shadows”. *PRX Quantum*, 4, 010305.
75. S. Bravyi and D. Maslov (2021). “Hadamard-free circuits expose the structure of the Clifford group”. *IEEE Transactions on Information Theory*, 67, 4546–4563.
76. O. Watts, Y. Kikuchi and L. Coopmans (2023). “Quantum semidefinite programming with thermal pure quantum states”. *arXiv:2310.07774*.
77. A. Gilyén, Y. Su, G. H. Low and N. Wiebe (2019). “Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics”. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing*, pp. 193–204.
78. M. Gluza, J. Son, B. H. Tiang, R. Zander, R. Seidel, Y. Suzuki, Z. Holmes and N. H. Y. Ng (2024). “Double-bracket quantum algorithms for quantum imaginary-time evolution”. *arXiv:2412.04554*.
79. D. Zhang, J. L. Bosse and T. Cubitt (2023). “Dissipative quantum Gibbs sampling”. *arXiv:2304.04526*.
80. A. D. C. Tosta, T. de Lima Silva, G. Camilo and L. Aolita (2023). “Randomized semi-quantum matrix processing”. *npj Quantum Information*. <https://api.semanticscholar.org/CorpusID:260125591>.
81. C.-F. Chen, M. J. Kastoryano, F. G. S. L. Brandão and A. Gilyén (2023). “Quantum thermal state preparation”. *arXiv:2303.18224*.
82. P. Rall, C. Wang and P. Wocjan (2023). “Thermal state preparation via rounding promises”. *Quantum*, 7, 1132.
83. O. Shtanko and R. Movassagh (2023). “Preparing thermal states on noiseless and noisy programmable quantum processors”. *arXiv:2112.14688*.
84. J. Lloyd and D. A. Abanin (2025). “Quantum thermal state preparation for near-term quantum processors”. *arXiv:2506.21318*.
85. Z. Ding, B. Li and L. Lin (2025). “Efficient quantum Gibbs samplers with Kubo–Martin–Schwinger detailed balance condition”. *Communications in Mathematical Physics*, 406, 67.
86. C. Rouzé, D. S. França and Á. M. Alhambra (2025). “Efficient thermalization and universal quantum computing with quantum Gibbs samplers”. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pp. 1488–1495.
87. Y. Tong and Y. Zhan (2025). “Fast mixing of weakly interacting fermionic systems at any temperature”. *PRX Quantum*, 6, 030301.
88. Š. Šmíd, R. Meister, M. Berta and R. Bondesan (2025). “Polynomial time quantum Gibbs sampling for fermi-Hubbard model at any temperature”. *arXiv:2501.01412*.

89. A. N. Chowdhury, G. Hao Low and N. Wiebe (2020). “A variational quantum algorithm for preparing quantum Gibbs states”. *arXiv:2002.00055*.
90. T. J. Sewell, C. D. White and B. Swingle(2022). “Thermal multi-scale entanglement renormalization ansatz for variational Gibbs state preparation”. *arXiv:2210.16419*.
91. J. Wu and T. H. Hsieh (2019). “Variational thermal quantum simulation via thermofield double states”. *Physical Reviews Letters*, 123, 220502.
92. Y. Wang, G. Li and X. Wang (2021). “Variational quantum Gibbs state preparation with a truncated Taylor series”. *Physical Reviews Applications*, 16, 054035.
93. M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes and M. Cerezo (2025). “Barren plateaus in variational quantum computing”. *Nature Reviews Physics*, 7, 1–16.
94. N. Fruitwala, A. Hashim, A. D. Rajagopala, Y. Xu, J. Hines, R. K. Naik, I. Siddiqi, K. Klymko, G. Huang and K. Nowrouzi (2024). “Hardware-efficient randomized compiling”. *arXiv:2406.13967*.
95. M.-D. Choi (1975). “Completely positive linear maps on complex matrices”. *Linear Algebra and its Applications*, 10, 285–290.
96. A. Jamiołkowski (1972). “Linear transformations which preserve trace and positive semidefiniteness of operators”. *Reports on Mathematical Physics*, 3, 275–278.
97. J. Watrous (2018). *The Theory of Quantum Information*. Cambridge University Press.
98. J. Jiang, K. Wang and X. Wang (2021). “Physical implementability of linear maps and its application in error mitigation”. *Quantum*, 5, 600.