

PAPER • OPEN ACCESS

Near-term distributed quantum computation using mean-field corrections and auxiliary qubits

To cite this article: Abigail McClain Gomez *et al* 2024 *Quantum Sci. Technol.* **9** 035022

View the [article online](#) for updates and enhancements.

You may also like

- [Demonstration of a Bayesian quantum game on an ion-trap quantum computer](#)
Neal Solmeyer, Norbert M Linke, Caroline Figgatt et al.

- [A universal quantum circuit design for periodical functions](#)
Junxu Li and Sabre Kais

- [Towards a scalable discrete quantum generative adversarial neural network](#)
Smit Chaudhary, Patrick Huembeli, Ian MacCormack et al.



kiutra

Easy-to-use and Helium-3 free cryogenics solutions

[LEARN MORE](#)

Quantum Science and Technology

**OPEN ACCESS****PAPER**

Near-term distributed quantum computation using mean-field corrections and auxiliary qubits

RECEIVED
24 September 2023

REVISED
26 March 2024

ACCEPTED FOR PUBLICATION
16 April 2024

PUBLISHED
3 May 2024

Original Content from
this work may be used
under the terms of the
Creative Commons
Attribution 4.0 licence.

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Abigail McClain Gomez^{1,2,*} Taylor L Patti^{2,*} Anima Anandkumar^{2,3} and Susanne F Yelin¹

¹ Department of Physics, Harvard University, Cambridge, MA 02138, United States of America

² NVIDIA, Santa Clara, CA 95051, United States of America

³ Department of Computing + Mathematical Sciences (CMS), California Institute of Technology (Caltech), Pasadena, CA 91125, United States of America

* Authors to whom any correspondence should be addressed.

E-mail: amcclain@g.harvard.edu and tpatti@nvidia.com

Keywords: distributed quantum computing, near-term quantum computing, quantum simulation, variational quantum algorithms

Abstract

Distributed quantum computation is often proposed to increase the scalability of quantum hardware, as it reduces cooperative noise and requisite connectivity by sharing quantum information between distant quantum devices. However, such exchange of quantum information itself poses unique engineering challenges, requiring high gate fidelity and costly non-local operations. To mitigate this, we propose near-term distributed quantum computing, focusing on approximate approaches that involve limited information transfer and conservative entanglement production. We first devise an approximate distributed computing scheme for the time evolution of quantum systems split across any combination of classical and quantum devices. Our procedure harnesses mean-field corrections and auxiliary qubits to link two or more devices classically, optimally encoding the auxiliary qubits to both minimize short-time evolution error and extend the approximate scheme's performance to longer evolution times. We then expand the scheme to include limited quantum information transfer through selective qubit shuffling or teleportation, broadening our method's applicability and boosting its performance. Finally, we build upon these concepts to produce an approximate circuit-cutting technique for the fragmented pre-training of variational quantum algorithms. To characterize our technique, we introduce a non-linear perturbation theory that discerns the critical role of our mean-field corrections in optimization and may be suitable for analyzing other non-linear quantum techniques. This fragmented pre-training is remarkably successful, reducing algorithmic error by orders of magnitude while requiring fewer iterations.

1. Introduction

One prospective trajectory for quantum information hardware is distributed quantum computing [1–3], the quantum analog of the celebrated classical field [4–7]. Distributed quantum computing seeks to eliminate the need for large, monolithic quantum computers, which suffer from cooperative noise [8, 9]. Instead, large-scale problems will be split among many smaller quantum computers that are in communication with each other via a *quantum interconnect*, a standardized form of quantum communication between remote quantum computing platforms [10, 11].

While the benefits of distributed quantum computing are abundant, many obstacles complicate its realization. For instance, due to the no-cloning theorem [12], extensive quantum entanglement would be a required component of quantum interconnects in order to enable non-local operations such as quantum teleportation [1, 2, 9, 13]. Moreover, fault-tolerant quantum computing would be needed to compute and transmit quantum information between distributed simulators reliably [11, 14]. Finally, long coherence times or relatively local topology would be necessary to manage the time delays associated with communication between remote locations [9, 15].

Nevertheless, the promise of scalability continues to inspire research in various facets of distributed quantum computing. Researchers have characterized the compilation of quantum circuits into cohesive network instructions [16] and devised a language to communicate such instructions more efficiently than conventional circuit diagrams [17]. Likewise, much work has been done to develop the non-local operations integral to distributed quantum computing, which have been supported with experimental realizations [13, 18–20]. Other studies have developed algorithms tailored to quantum distributed architectures, including Shor’s algorithm, quantum sensing, and combinatorial optimization [21–24], while additional research has focused on the quantum advantage provided by quantum distributed computing [25–28]. Still other research has addressed how to approach distributed algorithm design [24, 29], the effect of noise in distributed quantum computing [30], architecture selection and scalability [14, 31, 32], and resource allocation [33–35], particularly to optimize teleportation cost [36–38].

Although the interest in its theoretical application continues to grow, a wide gap remains between much distributed quantum computing research and its physical implementation. Research along a different vein has instead concentrated on applications that are realizable using near-term hardware, stretching the limit of noisy quantum simulators’ utility. Although not distributed in the sense of the works discussed above (which assume that the distributed hardware forms a quantum network), these approaches involve small groups of qubits simulated in parallel or in sequence to address a larger problem. Entanglement forging is one such approach [39], which relies on shifting computation to classical post-processing in order to assemble information from two smaller circuits, thereby halving the maximum circuit size required for the calculation. Other more general circuit knitting techniques have been recently developed, such as the scheme presented in [40], which pieces together the simulation of weakly entangled subsystems through classical resources and explores the trade-off between sampling overhead and simulation accuracy. The quantum tensor network approach uses the framework of tensor networks to identify weakly entangled subgroups and parallelize quantum simulation [41]. Similarly, Quantum multi-programming (QMP) takes advantage of the increasing size of available quantum simulators to execute multiple shallow quantum circuits concurrently [42, 43].

In order to bridge the distributed quantum computing paradigm with the capabilities of near and moderate-term hardware, in this manuscript, we design two procedures that approximately link distributed simulators while remaining amenable to small-scale, noisy devices. Our schemes of fragmented quantum simulation explore what problems can be addressed without full information transfer between hardware. First, focusing on the task of time evolution, we partition a system of qubits into subgroups (referred to as *fragments*) that are treated separately. We harness mean-field measurements to inform mean-field corrections [44] that link the distinct fragments. These simulations could be executed in parallel on a single simulator (as in QMP [42, 43]), outsourced to different simulators (as in distributed computing [1, 2, 9, 13]), or even simulated using a mixture of classical and quantum resources (as in *heterogeneous* computing [45, 46]). We further make use of a limited number of auxiliary qubits to mimic the presence of the qubits located on distant simulators.

In our first approach to distributed time evolution, we rely on classical communication to transmit partial state information between distant simulators through measurements, omitting a quantum link between devices. Transmitting incomplete information reduces the generally exponential number of measurements required to relay complete information of a quantum state via a classical channel. For locally interacting systems, the classical fragmentation scheme closely approximates quantities local to each fragment—including the fidelity of the fragment—for timescales up to several $1/J$, where J weights the system’s interactions. We present a second scheme that is supplemented by limited quantum information transfer, consequently composing an interface of classical and partial quantum information transfer that approximately connects quantum simulators. We show numerically that the limited use of quantum communication significantly extends the scheme’s performance to longer evolution times, even for long-range interacting systems. As non-local operations become more available, this technique could be employed in moderate-term distributed applications before a fully connected quantum network is achievable.

Using the same fragmentation framework, we devise a fragmented pre-training approach for variational quantum algorithms, focusing on the variational quantum eigensolver algorithm (VQE) [47]. The pre-training can be performed classically or using resource-limited hardware, as only portions of the full circuit are considered. In this technique, gates between fragments are cut and ignored (that is, no gate reconstruction through classical post-processing is employed); however, through mean-field corrections and auxiliary qubits, the resultant error is kept small enough to still approach the space of optimal parameters through the fragmented optimization. For classical MaxCut problem graphs, the pre-training method reduces energy error by various orders of magnitude on average, and requires over an order of magnitude fewer circuit preparations. For transverse field Ising-like models [48, 49] outside of the classical domain, our pre-training scheme maintains a significant advantage in the regime of a small transverse field h .

The remainder of the paper is organized as follows. In section 2, we first present a fragmented approach to quantum simulation that only involves the classical transfer of partial state information. We further consider an alternate scheme for the case of linking quantum simulators with reduced quantum information transfer through selective qubit shuttling [50] or teleportation [51, 52], in addition to classical information transfer. In section 3, the performance of each scheme is evaluated for the time evolution of quantum Ising-like spin Hamiltonians [53], which are amenable to quantum simulation using trapped ions and Rydberg platforms [54, 55]. Finally, in section 4 we expand the scheme to apply to the optimization of quantum circuits. The use of our fragmentation scheme to assist VQE is evaluated in section 5 [47]. The role of mean-field corrections in the optimization through the lens of perturbation theory [56] is explored in sections 5.2.2 and 5.2.3. In section 5.2.3, we introduce a non-linear perturbation theory to study mean-field corrected Hamiltonians, analytically formalizing the success of our pre-training approach.

2. Fragmented quantum simulation

In our method of fragmented quantum simulation, we divide a system of N qubits into two or more sub-systems, here referred to as *fragments* (see figure 1(a)). Each fragment contains some number of qubits $N_f < N$, such that $\sum_f N_f = N$. The fragments are treated separately, but it is possible to approximate the presence of a fragment's *environment*, that is, the qubits outside of a given fragment, through corrective fields and interactions [57]. We devise mean-field corrections (described in detail in section 2.1) [44], which are informed by measurements of a fragment's environment, to actively adjust the state of a fragment. Corrective interactions are mediated by the inclusion of auxiliary qubits within each fragment's simulation, such that $\sum_f N_{f+a} > N$, where $N_{f+a} = N_f + N_a$ and N_a represents the number of auxiliary qubits included in fragment f . Each auxiliary qubit mimics the behavior of one environment qubit, which we refer to as the *target* qubit for that auxiliary. Each auxiliary qubit interacts with the fragment's qubits according to the same interaction terms as the corresponding target qubit, as prescribed by the original Hamiltonian, enabling entanglement to grow beyond the N_f fragment qubits.

Figure 1(b) provides an overview of our classically-linked fragmentation scheme, and a detailed diagram is provided in figure A1. We define a fragment's *interface* I to be the collection of interactions existing in the original Hamiltonian that act between fragment qubits and environment qubits. The combination of auxiliary qubits and mean-field corrections collectively mimics the action of the interface on the fragment. The growth and faithfulness of the entanglement within a fragment will be limited by the number of auxiliary qubits included—an unavoidable limitation of the scheme—but the effects of this limitation can be mitigated through judicious fragmentation of the system. Firstly, to mitigate fragmentation error (that is, the error produced by the omission of some system interactions and the resultant reduction of Hilbert space), one can choose to divide the system qubits such that the qubits interacting most influentially with each other are confined to a single fragment. Secondly, it is possible to make an informed choice of target qubit for each auxiliary. This is explored further in section 2.2.

Finally, we emphasize that this simulation technique is not exact. In fact, in the classically-linked case, the fully quantum state is not reconstructed from the evolution of individual fragments, and thus, observables that span more than one fragment are inaccessible. Nonetheless, observables local to one fragment can be estimated; this is demonstrated section 3.

2.1. Mean-field corrections

Consider the class of spin models:

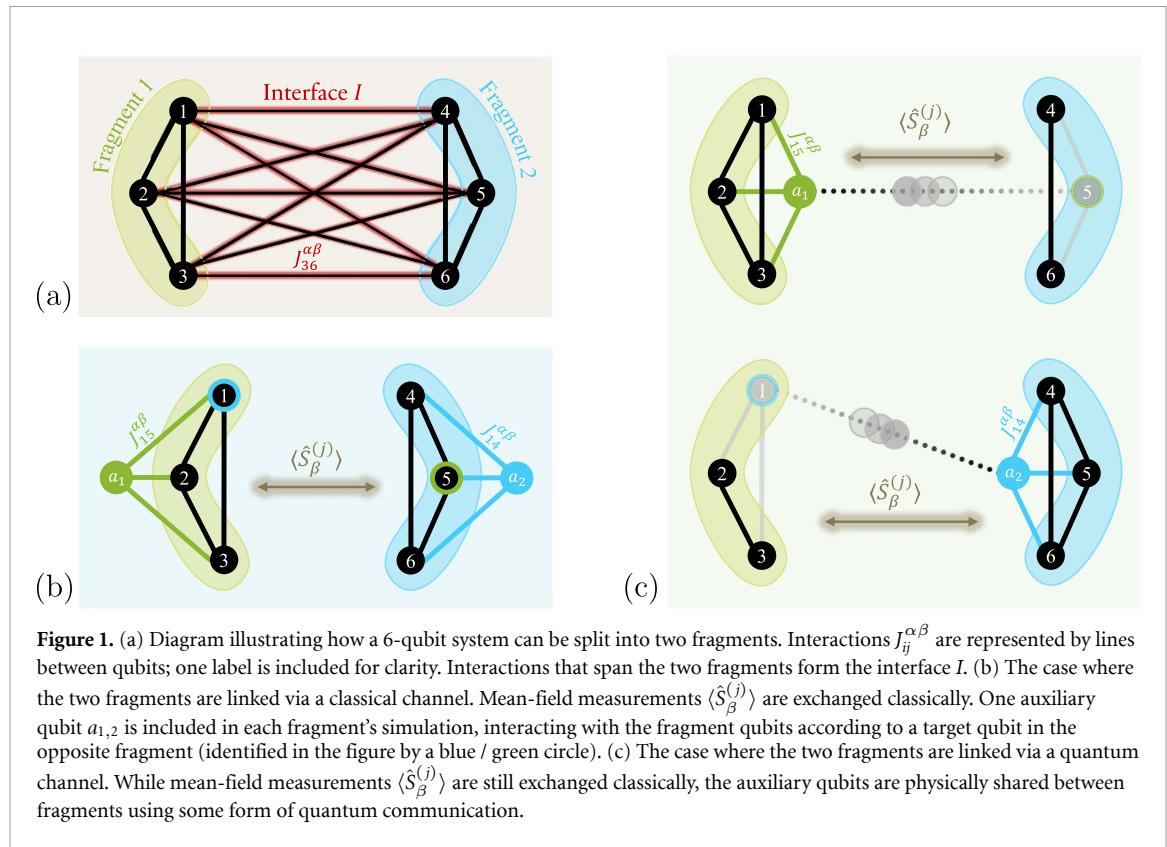
$$H = - \sum_{\langle i,j \rangle} \sum_{\alpha, \beta} J_{ij}^{\alpha, \beta} \hat{S}_\alpha^{(i)} \hat{S}_\beta^{(j)} - \sum_{i=1}^N h_i \hat{S}_x^{(i)}. \quad (1)$$

Here, $\hat{S}_\alpha^{(i)}$ and $\hat{S}_\beta^{(j)}$ are spin-1/2 spin operators acting on sites i and j , where $\alpha, \beta \in \{x, y, z\}$. The coefficient $J_{ij}^{\alpha, \beta}$ gives the strength and sign of the interaction. For concreteness and without loss of generality, we have selected transverse fields h_i to point along the x -axis. The Hamiltonian acting strictly within some sub-system f will neglect any operators acting outside of f , yielding

$$H^{(f)} = - \sum_{\langle i,j \rangle \in f} \sum_{\alpha, \beta} J_{ij}^{\alpha, \beta} \hat{S}_\alpha^{(i)} \hat{S}_\beta^{(j)} - \sum_{i \in f} h_i \hat{S}_x^{(i)}, \quad (2)$$

the bare Hamiltonian that acts within a fragment f when no corrections are included.

Clearly, the simple exclusion of interactions that span the interface between f and its environment (i.e. the fragmented evolution of f under $H^{(f)}$) will, in general, poorly approximate the evolution of the



sub-system under the full Hamiltonian. The fragment qubits will behave as a closed system without external interactions. Although generally these interactions cannot be exactly simulated without modeling all of the system's spins on a single fragment, we introduce a mean-field to partially capture the action of each missing interaction. Mean-field methods have frequently been used to simplify the simulation and study of quantum systems, and statistical physics [44, 58, 59]. Here, the strength and sign of the introduced mean-field correction is informed by the measurement of the corresponding environment spin, while the correction's axis is determined by that of the corresponding interaction's spin operator that would act within fragment f . The resulting mean-field corrected Hamiltonian is given by:

$$H_{\text{MF}}^{(f)} = - \sum_{\substack{\langle i,j \rangle \in I, \alpha, \beta \\ i \in f}} \sum J_{ij}^{\alpha\beta} \hat{S}_\alpha^{(i)} \langle \hat{S}_\beta^{(j)} \rangle - \sum_{\langle i,j \rangle \in f} \sum_{\alpha, \beta} J_{ij}^{\alpha\beta} \hat{S}_\alpha^{(i)} \hat{S}_\beta^{(j)} - \sum_{i \in f} h_i \hat{S}_x^{(i)}. \quad (3)$$

The strength and direction of the mean-fields appearing in $H_{\text{MF}}^{(f)}$ should be updated regularly to reflect the current state of the environment spins. Physically, this requires regular mean-field measurements of the fragments. Evolution must therefore be reset to the initial state in order to proceed by one time step dt , with each new mean-field measurement being stored to progress the evolution. The process of incrementing the time evolution by one time step per simulation is commonly implemented in order to track the time dynamics of an observable [60], resulting in a complexity that scales as $\mathcal{O}(N_t^2)$ in the number of time steps N_t .

2.2. Auxiliary target spin selection

For nearest-neighbor spin models (e.g. the transverse field Ising model [49]), the selection of a target spin for each auxiliary is somewhat trivial, as at most two qubits interact with a fragmented section of the chain. The choice of auxiliary qubit encoding may be unclear for more general systems. Here, we present a method for auxiliary target qubit selection that yields, on average, the optimal auxiliary qubit encoding. Specifically, we consider how auxiliary target selection affects the simulation error to the first non-vanishing order in dt . This simulation error arises from the omission of interactions forming the interface of some particular fragment f and the remaining environment spins E . The full derivation of the leading error is provided in appendix A.2; here, we sketch the derivation and build on the result.

To derive the leading error, we define the fidelity between the evolved state of a fragmented system and that of the full system to be:

$$F(t) = |\langle \Psi | U_I^{(f)}(t) | \Psi \rangle|^2. \quad (4)$$

The unitary operator $U(t) = \exp(-iHt)$ evolves the system exactly under the full Hamiltonian H , while $U_I^{(f)}(t) = \exp(-iH_I^{(f)}t)$ evolves the system under a fragmented Hamiltonian. Notably, the Hamiltonian $H_I^{(f)}$ around which this discussion revolves is not the Hamiltonian that acts only within fragment f ; rather, $H_I^{(f)}$ additionally includes all interactions between qubits that are external fragment f . Thus, $H_I^{(f)}$ lives in the Hilbert space of the full system, neglecting only the interactions crossing the interface of fragment f in order to isolate the fragmentation error associated with f . In appendix A.2, equation (4) is expanded for short times t to understand how the evolution error $\epsilon(t) = 1 - F(t)$ depends on the strength of the neglected interactions. Through the use of Taylor expansion and the Baker–Campbell–Hausdorff (BCH) formula [61], we arrive at the first non-vanishing correction to the fidelity:

$$F(t) \approx 1 - \text{var} \left(H - H_I^{(f)} \right) t^2, \quad (5)$$

where $\text{var}(\mathcal{O})$ is the quantum variance of operator \mathcal{O} . The error $\epsilon(t) = 1 - F(t)$ is thus given by $\text{var}(H - H_I^{(f)})t^2$ for short times t .

The form of the short-time error provides a simple rule for choosing the target auxiliary qubits for fragment f to minimize error; namely, select the environment qubit(s) whose interactions contribute most significantly to the variance $\text{var}(H - H_I^{(f)})$. This choice will minimize the short-time error of evolving the state by the fragmented Hamiltonian, which will lead to higher fidelity performance, on average (see section 3.3). Moreover, if the auxiliary selection is updated sufficiently often, the selection becomes exact as the short-time error dominates from the time of one auxiliary encoding to the next.

2.3. Practical implementation of the optimal auxiliary encoding

Although the final form of the short-time evolution error provides insight into optimal auxiliary selection, the procedure for estimating a particular qubit's contribution to the error within the distributed framework is less straightforward. For a general spin Hamiltonian, this variance is given by:

$$\begin{aligned} \text{var} \left(H - H_I^{(f)} \right) &= \text{var} \left(- \sum_{\langle i,j \rangle \in I} \sum_{\alpha,\beta} J_{ij}^{\alpha,\beta} \hat{S}_{\alpha}^{(i)} \hat{S}_{\beta}^{(j)} \right) \\ &= \sum_{\langle i,j \rangle \in I} \sum_{\langle i',j' \rangle \in I} \sum_{\alpha,\beta} J_{ij}^{\alpha,\beta} J_{i'j'}^{\alpha',\beta'} \left(\left\langle \hat{S}_{\alpha}^{(i)} \hat{S}_{\beta}^{(j)} \hat{S}_{\alpha'}^{(i')} \hat{S}_{\beta'}^{(j')} \right\rangle - \left\langle \hat{S}_{\alpha}^{(i)} \hat{S}_{\beta}^{(j)} \right\rangle \left\langle \hat{S}_{\alpha'}^{(i')} \hat{S}_{\beta'}^{(j')} \right\rangle \right). \end{aligned} \quad (6)$$

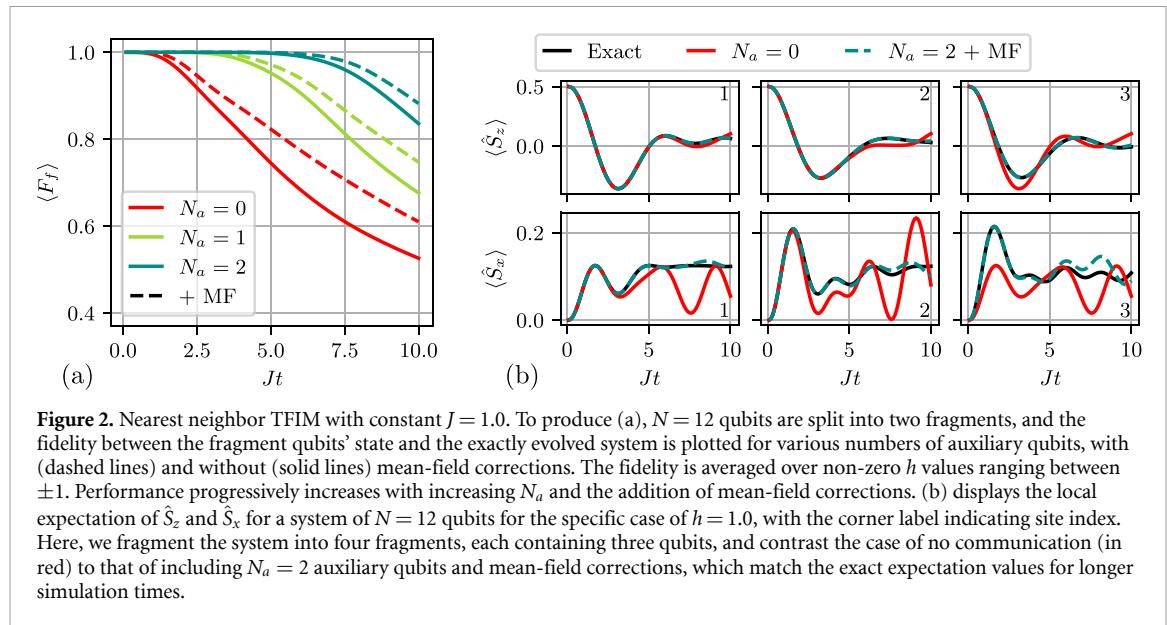
Estimating the full variance of equation (6) requires 4-point correlation measurements. If the distributed simulators are linked solely via classical channels, correlation measurements are only accessible when all relevant qubits are local to a single fragment. This implies that two auxiliary qubits—one targeting j and one targeting j' —must already be placed within the fragment in order to access the required 4-point correlator measurements. For N_E environment qubits, there are $\mathcal{O}(N_E^2)$ combinations, requiring $\mathcal{O}(N_E^2)$ copies of the system in order to estimate all required 4-point correlators, undermining (although not necessarily precluding) the motivations for fragmented quantum simulation with such a technique.

The correlator calculation simplifies significantly when the variance is calculated with respect to a known product state, but a new issue arises: for many spin model Hamiltonians, the variance will vanish for certain initial product states. In fact, for the case of the transverse field Ising model [49], this quantity vanishes for all computational basis states, providing no insight into the proper auxiliary choice.

We propose a two-part solution that addresses these issues. First, we propose a proxy $v(a)$ that estimates the contribution of one potential auxiliary a to the variance:

$$v(a) = \sum_{\langle i,j \rangle \in I} \sum_{\langle i',j' \rangle \in I} \sum_{\alpha,\beta} \sum_{\alpha',\beta'} J_{ij}^{\alpha,\beta} J_{i'j'}^{\alpha',\beta'} \delta_{j,a} \delta_{j',a} \left(\left\langle \hat{S}_{\alpha}^{(i)} \hat{S}_{\beta}^{(j)} \hat{S}_{\alpha'}^{(i')} \hat{S}_{\beta'}^{(j')} \right\rangle - \left\langle \hat{S}_{\alpha}^{(i)} \hat{S}_{\beta}^{(j)} \right\rangle \left\langle \hat{S}_{\alpha'}^{(i')} \hat{S}_{\beta'}^{(j')} \right\rangle \right). \quad (7)$$

Inserting the two Dirac delta functions $\delta_{j,a} \delta_{j',a}$ eliminates the cross-terms in equation (6) that depend on multiple environment qubits. Thus, a single auxiliary is required to estimate $v(a)$, and in total $\mathcal{O}(N_E)$ partitions are required to acquire $v(a)$ for all potential auxiliary targets. In addition to requiring fewer measurements, this proxy focuses on a 's contribution to the variance while neglecting the cross-terms that involve contributions from other potential auxiliary qubits. Secondly, to avoid scenarios where the variance vanishes for initial product states, we suggest first evolving the system for one time step dt for a particular choice of a before estimating $v(a)$. Although this procedure is more involved than calculating $v(a)$ for the initial product state directly, the overhead remains linear in the number of potential auxiliary qubits.



3. Application 1: fragmented time evolution

3.1. Simulators linked via classical information

We first focus on the scheme that involves no quantum information transfer. In this scheme, the auxiliary qubits are selected at the beginning of the simulation and fixed to target a single environment qubit throughout the evolution. We refer the reader to figure 1 and appendix A.1 for an in-depth look at how a system is fragmented for time evolution. As a representative example, consider the transverse field Ising model (TFIM) [48, 49] with a uniform transverse field:

$$H_{\text{TFIM}} = -J \sum_{\langle i,j \rangle} \hat{S}_z^{(i)} \hat{S}_z^{(j)} - h \sum_i^N \hat{S}_x^{(i)}. \quad (8)$$

This system has been studied in depth to better understand the physics of quantum phase transitions [62–64]. We consider the evolution of a quantum system initialized in the computational basis state $|0\rangle$ under H_{TFIM} , implementing exact unitary evolution numerically using PennyLane [65] with mean-field measurements updated every $dt = 0.1/J$. Figure 2 displays the scheme's performance for a 12-qubit model with nearest neighbor interactions of $J = 1.0$. The results presented in figure 2(a) are averaged over non-zero transverse fields h ranging from ± 1 , while figure 2(b) features the specific case of $h = 1.0$. In figure 2(a), the system is split into two fragments, each simulating six of the system qubits and some number of auxiliary qubits. The average of the quantity F_f is plotted, which we define as the fidelity between the reduced density operator of the system qubits within the fragment (tracing out any auxiliary qubits a) and the reduced density of the same system qubits for the exact evolution of the full system (tracing out all environment qubits forming E). We use the generalization of fidelity for density matrices [66] to enable the focused evaluation of the fragment sub-system:

$$F_f = \left(\text{Tr} \sqrt{\sqrt{\rho_f} \rho_f^{\text{ex}} \sqrt{\rho_f}} \right)^2, \quad (9)$$

$$\rho_f = \text{Tr}_a \rho_{f+a}, \quad (10)$$

$$\rho_f^{\text{ex}} = \text{Tr}_E \rho^{\text{ex}}. \quad (11)$$

For a short evolution time, the scheme captures the correct state of the system qubits within the fragment. This time can be extended by the inclusion of additional auxiliary qubits.

In figure 2(b), we consider a specific instance of the TFIM with $J = 1.0$ and $h = 1.0$. To test the scheme, we split the system into smaller partitions with $N_f = 3$. When we make use of two auxiliary qubits and mean-field corrections, the local expectation values exhibit little error for several units of Jt , as expected from the fidelity results.

In figure 3, we examine the scaling performance for the nearest neighbor model, increasing the number of fragments simulated with increasing N (keeping N_f constant) with $N_a = 2$. In the left panel, the first

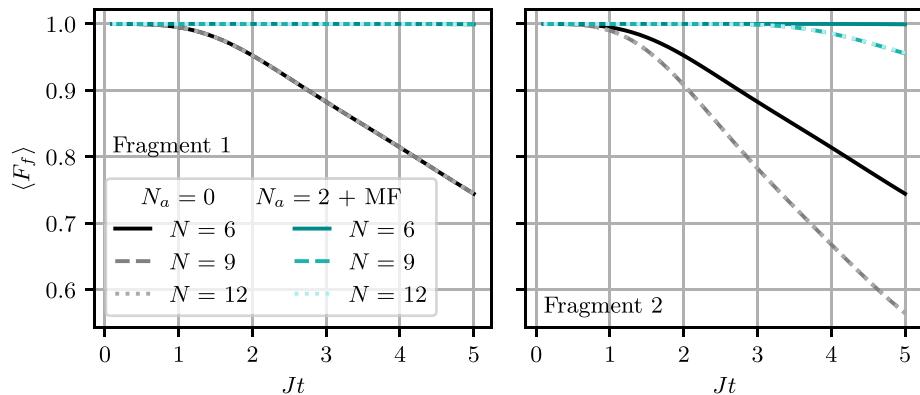


Figure 3. Scaling performance of the classical scheme for the nearest neighbor TFIM with $J = 1.0$, averaged over h values ranging from ± 1 . For each simulation, the number of system qubits within a fragment is fixed to be $N_f = 3$ as N is increased, with N_a fixed to be zero (no communication in black / gray) or two (in blue). The left panel plots $\langle F_f \rangle$ for the first fragment (which includes the boundary qubit and thus only involves one interaction crossing the interface), while the right panel plots $\langle F_f \rangle$ for the second fragment (which, for $N = 9$ and $N = 12$, is an *interior* fragment with two interactions crossing the interface).

fragment is considered. This fragment contains the boundary of the chain, and consequently, the fragment's interface consists of only one missing interaction. The right panel considers the second fragment, which is on the interior of the chain for $N > 6$ and consequently neglects two interactions, leading to reduced performance. This is manifested in the reduced fragment fidelity going from $N = 6$ to $N = 9$ for fragment 2. However, there is no such visible drop going from $N = 9$ to $N = 12$ due to the monogamy of entanglement [67]—that is, although the number of qubits in the system grows, the qubits that are most strongly entangled with each other remain local to one fragment, and thus the amount of lost information shrinks as N is further increased. We therefore expect our classical scheme to scale well with N for systems that are locally interacting, and to serve as a strong approximation for moderate evolution times.

3.2. The addition of quantum information transfer

Next, we examine the case of selective quantum information transfer between quantum simulators, applicable when non-local operations are available, even if only in a limited capacity. In this case, the fragmentation scheme can be modified to include limited quantum information transfer (a *quantum channel* [9]). The role of the auxiliary qubits shifts from being bystanders confined to a single fragment to qubits that are physically shared between simulators through selective non-local interactions, accomplished through qubit shuttling [50] or teleportation [51, 52] (see figure 1(c)). If the simulations are being executed in parallel on a single quantum simulator [42, 43], this would only require a few additional SWAP gates to include a limited number of cross-simulation interactions. In addition to providing more complete information transfer, a quantum channel further enables the active correction of auxiliary encoding as the system evolves. The selected number of auxiliary qubits places a limit on the number of qubits that are physically teleported / shuttled to a fragment; however, which environment qubits play this role can be changed from one time step to the next depending on which potential auxiliary qubit(s) have the largest contribution to the most recent estimate of the short-time error. When quantum channels and synchronized measurements are available, all correlation measurements are accessible. The quantity $v(a)$ can thus be estimated for any a at any time. As the potential auxiliary qubits' contribution to the variance shift, new auxiliary qubits can be selected—that is, we can make a new selection for which qubit(s) physically interact with a fragment native to a different simulator. If the time steps are sufficiently small such that the first non-vanishing order in the error dominates, then this becomes optimal even for long simulation times.

To evaluate this scheme numerically, we abstract away the details of information transport; this topic has been investigated by other research in the context of distributed time evolution [68]. In our actively updated simulation, the quantity $v(a)$ is calculated for each potential auxiliary qubit at each time step to determine its contribution to the short-time error. This requires the estimation of correlators between each potential auxiliary and each fragment system qubit (requiring $\mathcal{O}(N_f N_E N_{\alpha\beta}^2)$ per fragment, where $N_{\alpha\beta}$ is the number of $\alpha\beta$ interaction types), but if there is only one kind of interaction (as is the case for the TFIM and other Ising-like models, with $\alpha = \beta = z$), all relevant correlators can be estimated from a set of full system snapshot measurements. The largest contributors are selected to be auxiliaries—numerically, this amounts to keeping the interactions between these qubits and the fragment qubits, while zeroing the $J_{ij}^{\alpha\beta}$ coefficients of all other environment-fragment interactions (see appendix A.3). Any zeroed interactions can be approximately

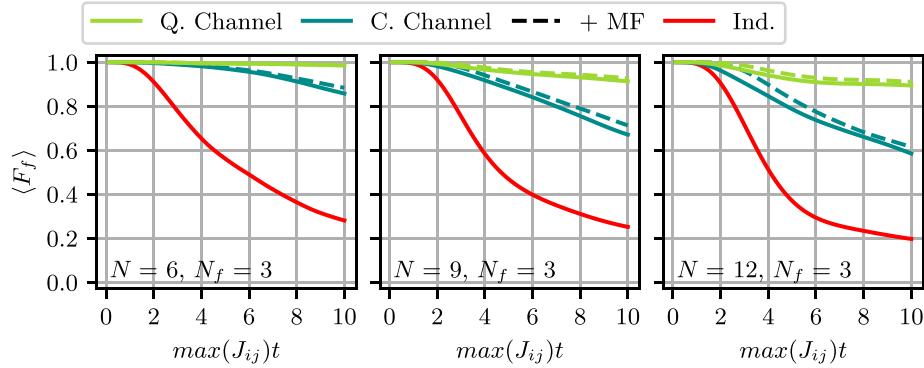


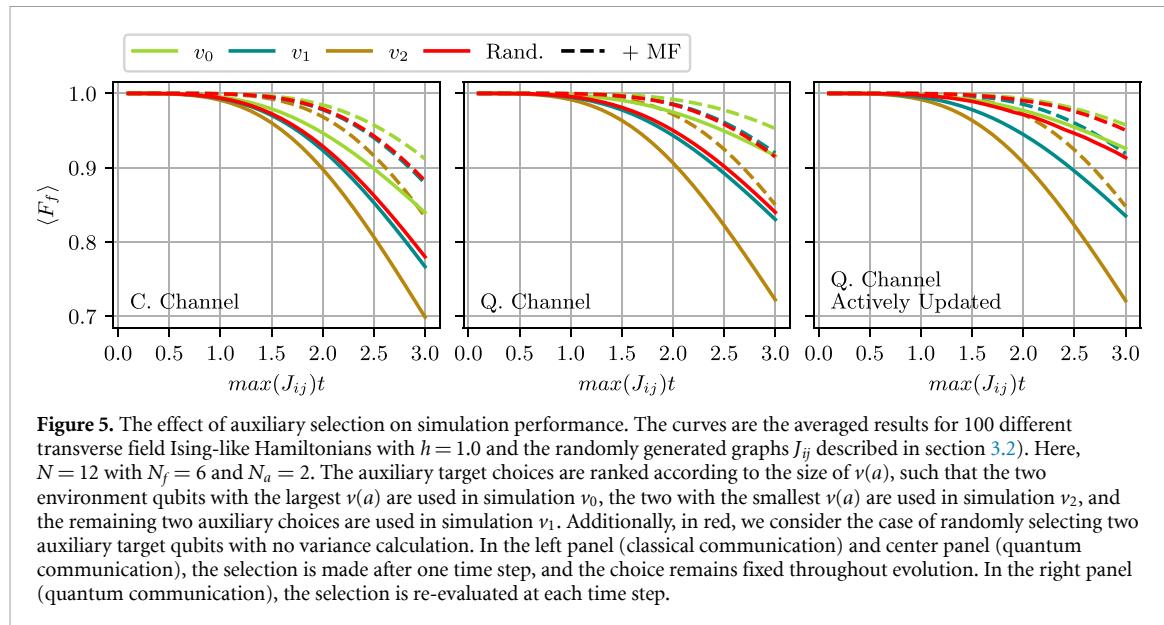
Figure 4. Comparison between scheme involving only classical information transfer (dark teal) to that involving limited quantum and classical information transfer (light green). For reference, the independent case (no information transfer) is included in red. The results are averaged over 100 Ising-like Hamiltonians with constant $h = 1.0$ and randomly generated graphs J_{ij} (see section 3.2). The N qubits are split into groups such that $N_f = 3$, with an additional $N_a = 2$ auxiliary qubits employed in the simulation.

included via mean-field corrections. At the next time step, the selection of zeroed interactions might change due to a change in the selected auxiliaries for each fragment, as dictated by the short-time error.

Figure 4 compares this scheme (labeled ‘Q. Channel’ to indicate the addition of quantum information transfer) to the previous scheme in section 3.1, which involves only classical information transfer (‘C. Channel’). The graph plots the fragment fidelity F_f averaged over 100 transverse field Ising-like models with $h = 1.0$ and randomly generated graphs J_{ij} . Each edge ij exists with probability 0.5, and edge weights J_{ij} are sampled from a Gaussian distribution with mean $\mu = 0.0$ and width $\sigma = 1.0$. Furthermore, we randomly select a computational basis state to initialize the fragmented system. Although both schemes outperform the case of no information transfer (in red), the complicated long-range nature of the Hamiltonians considered challenges the previous scheme, which only employs classical information transfer. In contrast, the quantum scheme preserves a large fragment fidelity, even at late simulation times.

3.3. Short-time error auxiliary selection

The benefit of using short-time error to inform auxiliary selection can be isolated by evaluating the performance of each auxiliary choice independently. Consider a system of $N = 12$ qubits, fragmented into two groups of $N_f = 6$. This leaves six environment qubits from the perspective of each fragment that could be targeted by an auxiliary qubit. Selecting two auxiliary qubits ($N_a = 2$), we rank the six potential choices for target auxiliary encoding according to the size of $v(a)$. In figure 5, the six target encoding choices are divided into three groups of two based on $v(a)$, and each option is explored for randomly generated transverse field Ising-like Hamiltonians with $h = 1.0$, as considered in the previous section. A total of 100 such Hamiltonians are generated and simulated; the averaged results are presented in figure 5, where v_0 corresponds to encoding the two environment qubits with the largest value for $v(a)$. On the left, the results are plotted for the case of classical information transfer. Any separation between the fidelity curves corresponding to different auxiliary choices indicates that the $v(a)$ metric meaningfully separates the potential auxiliary choices according to fidelity performance. The fact that the ordering corresponds to the ranked choice is evidence that using short-time error to select auxiliary encoding propagates to better performance at later times. In red, we consider random auxiliary encoding. The random performance roughly converges to the middle-ranked choice v_1 and can be thought of as the performance averaged over auxiliary encoding. In the center, the results are plotted for the case of additional quantum information transfer, without actively updating the auxiliary encoding. The results qualitatively match those of the classical case, with slightly better performance overall, consistent with figure 4. In the right panel, we consider the quantum channel with actively updated auxiliary encoding. In this case, v_0 (v_2) corresponds to selecting the two auxiliary targets with the largest (smallest) values for $v(a)$ at each decision. The performance of v_0 marginally increases with the introduction of active updates, while the performance of v_2 marginally decreases. However, the random performance increases most markedly. Here, the rapid shuffling of auxiliary qubit encoding allows the fragments to quickly share information, leading to performance comparable to the optimal variance choice, v_0 . The random, actively updated case has the added advantage of being measurement-efficient as it forgoes any variance estimation, but the highly frequent change of auxiliary encoding may lead to an overhead in qubit routing / swapping in order to be realized.



Finally, we note that in the averaged results presented in figure 5, the mean-field corrected simulation (plotted with a dashed line) outperforms the corresponding simulation that fully neglects these interface interactions for every case considered. Appendix B investigates the use of mean-field corrections to reduce simulation error through a numerical study.

4. Fragmented quantum circuits

We now investigate the use of fragmentation in quantum circuit evolution. Specifically, we focus on the fragmentation of a parameterized quantum circuit (PQC), the basic model architecture of variational quantum algorithms with applications ranging from ground state preparation to classification and recognition tasks [69–71]. Consider the fragmentation of a PQC of size N into multiple smaller PQCs. To fragment a circuit, multi-qubit unitaries that act on qubits outside the N_{f+a} qubits devoted to a single sub-system's PQC are neglected. Although this resembles the first step of circuit-cutting techniques [40, 72], no data processing is required to reconstruct the cut gates; they are simply ignored. Crucially, some auxiliary qubits are included in each sub-system PQC, such that the full set of sub-system PQCs overlap with one another and $\sum_f N_{f+a} > N$ (see figure 6). Including extra registers each fragmented circuit helps reduce the error of the fragmentation scheme, just as the auxiliary qubits were shown to extend the accurate time evolution of fragmented systems in the previous sections. We stress that the ensemble of fragmented circuits is not equivalent to the full circuit; nonetheless, for finding the ground state of certain classes of models, the linked optimization of the fragmented circuit ensemble can approach the optimal parameters of the full circuit. In particular, we optimize the collection of fragmented circuits prior to optimizing the full circuit as a new approach to *pre-training*, commonly employed to boost variational quantum algorithms [73–79]. Pre-training generally uses classical resources and can greatly increase the accuracy of a variational algorithm's solution, which is crucial for many applications such as reaching chemical accuracy for quantum chemistry problems [80–82]. Our pre-training approach is motivated by the fact that the parameter solutions of the smaller circuits are expected to be smoothly connected to the parameter solutions of the full quantum circuit, as explored by [83]. We constrain the pre-training to use small circuits that are cheap to simulate classically. Furthermore, employing smaller circuits limits entanglement growth, which has been shown to improve training and avoid barren plateaus [83–88].

5. Application 2: fragment-initialized VQE

Our method of fragmenting a quantum circuit can be applied to classically pre-train quantum circuit parameters for the VQE [47]. For this application, a PQC of size N is divided into smaller PQCs, each having size $N_{f+a} < N$. To optimize each sub-system PQC, the mean-field-corrected Hamiltonian given in equation (3) is minimized. In addition to facilitating the study of quantum systems and statistical physics, mean-field methods have been introduced for data analysis and loss function modification [89–92]. In our pre-training technique, employing mean-field terms serves to link the optimization of the separate circuits

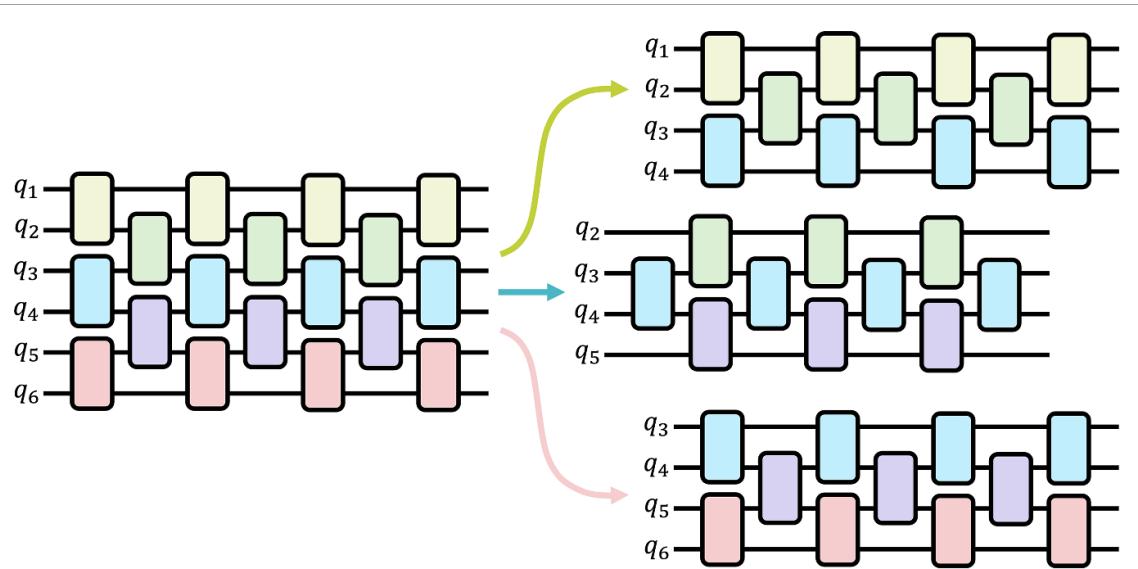


Figure 6. Diagram depicting how a circuit can be fragmented into a number of smaller circuits with overlapping registers, analogous to the inclusion of auxiliary qubits. The $N = 6$ qubits are partitioned into three groups of two—that is, $N_f = 2$ for each circuit, with q_1, q_2 addressed by the top PQC, q_3, q_4 addressed by the middle PQC, and q_5, q_6 addressed by the bottom PQC. Two additional auxiliary registers are included in each small PQC, such that some of the parameterized two-qubit gates appear in multiple PQCs. Gates that address qubits beyond the scope of one PQC are neglected by that particular circuit.

by their current mean-field measurements. Overlapping parameters (that is, parameters shared by two fragmented PQCs) are initialized for one PQC using the most recent values from the other, further uniting the separate circuit optimizations. The mean-field measurements are updated regularly, and optimization halts when the steady state (up to some set precision) is reached for all parameters—those shared and those unique to one PQC—or the maximum number of iterations is reached. The algorithm is outlined in algorithm 1.

Algorithm 1. Fragment pre-training with mean-field corrections.

```

(Randomly) initialize  $\{\theta_i\}$  for the brickwork section of the full PQC.
Divide  $\{\theta_i\}$  into a set  $\{\theta_{f,i}\}$  for each fragment  $f$ .
Initialize  $\langle \hat{S}_\beta^{(j)} \rangle(0) = 0$ .
repeat
  for  $f$  in system do
     $\theta_{f,i=a}(k) \leftarrow \theta_{i=a}(k)$  for auxiliary spins  $a$  in  $f$ .
     $\theta_{f,i}(k+1) \leftarrow \theta_{f,i}(k) - \eta \nabla_{\theta_{f,i}} \langle H_{\text{MF}}^{(f)} \rangle_f$ .
     $\langle \hat{S}_\beta^{(j)} \rangle(k+1) \leftarrow \langle \hat{S}_\beta^{(j)} \rangle_f(k+1)$  for system spins  $j \in f$ .
     $\theta_j(k+1) \leftarrow \theta_{f,j}(k+1)$  for system spins  $j \in f$ .
  end for
until Parameters  $\{\theta_i\}$  converge.

```

5.1. Details of ansatz

We focus on pre-training brickwork circuits with a limited number of layers to constrain entanglement growth between fragments. We note that this is analogous to restricting ourselves to short evolution times to ensure more accurate results from a fragmented scheme. By utilizing a shallow circuit for pre-training, the error in the output of the ensemble of fragmented circuits is kept manageable. Although a circuit ansatz with high complexity is often necessary for interesting VQE applications in order to provide enough expressivity to reach the ground state [93, 94], fragmentation-based pre-training is still beneficial through the use of a layer-wise approach [95]. If a shallow brickwork circuit is placed ahead of a more expressive PQC ansatz, the brickwork layers can first be optimized using the fragmented approach. These layers serve to bring the state of the system to have some ground state overlap. The full circuit VQE can then be performed, initializing the leading brickwork layers of the circuit with the pre-trained parameter values and initializing the remaining gates of the ansatz to approximately act as identity—specifically, we choose to randomly initialize these parameters to be small values bounded by $\pm \varepsilon$ (with $\varepsilon = 10^{-5}$ for our results), to balance maintaining the optimized action of the initial layers after pre-training while avoiding training issues associated with a true identity initialization [73, 96]. The overall circuit layout is outlined in figure 7. Although we employ a

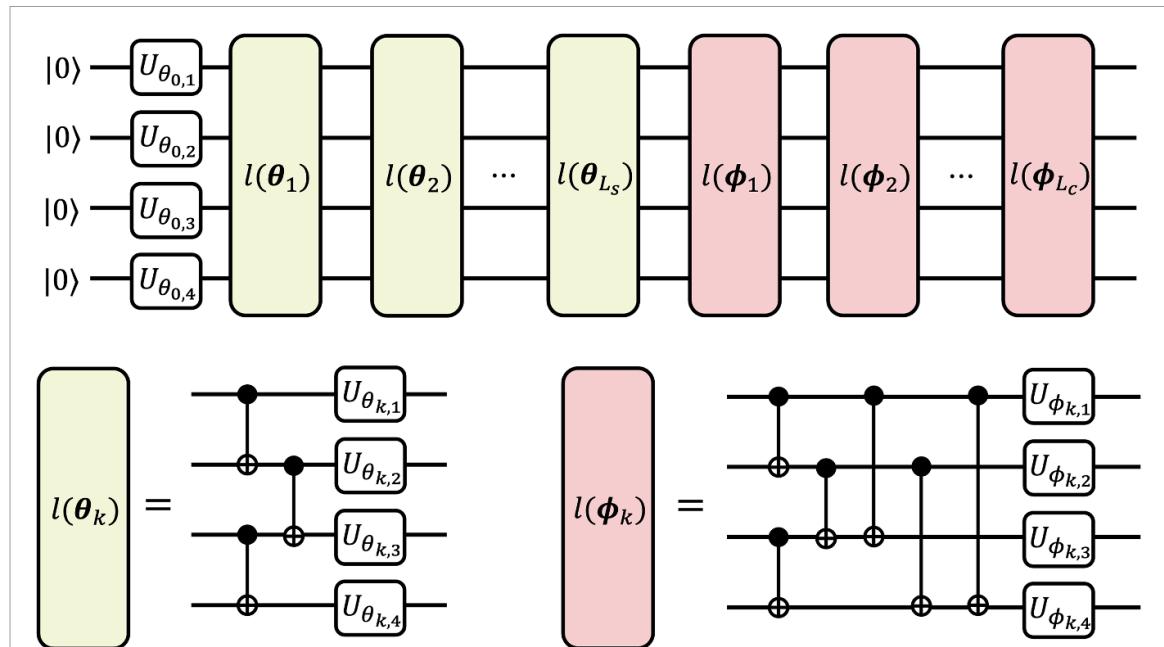


Figure 7. The circuit ansatz is built from L_s layers $l(\theta)$ with linear entangling gates, which are amenable to fragmentation. These are followed by a set of L_c layers $l(\phi)$ with an all-to-all entangling architecture. Only the brickwork layers parameterized by θ_i are pre-trained using the fragmented scheme, while the layers parameterized by ϕ_j are employed only in the final training process.

fully-entangling architecture in the latter portion of the circuit to increase expressibility, we note that one might limit the number of gates connecting fragments, as is explored in [97], to efficiently create a fully distributed architecture.

5.2. Performance and analysis of fragmented pre-training

To evaluate our pre-training method, we focus on random Ising-like models. In section 5.2.1, we present the numerical performance of the scheme for the classical case of zero transverse field ($h = 0$). Having established the advantage of the approach, in section 5.2.2 we derive its success as stemming from the mean-field corrective terms included in the loss function, which shift the global minimum of the collective fragmented circuit to coincide with that of the full optimization problem. Finally, in section 5.2.3, we use perturbation theory and numerical simulation to demonstrate that our approach remains beneficial for $|h| > 0$, in the regime of a weak transverse field.

5.2.1. VQE results for MaxCut

We first benchmark the scheme using randomly generated classical Ising Hamiltonians, where the all-to-all J_{ij} interactions are sampled from a Gaussian distribution (mean $\mu = 0.0$, width $\sigma = 1.0$) and the transverse field h is fixed to be zero. Finding the ground state of such models can be mapped to weighted MaxCut problems with positive and negative weights that are randomly generated [87]. The MaxCut problem is a graph partitioning problem that is known to be NP-hard, with applications ranging from network optimization to circuit design [98]. For the circuit ansatz, a fixed number of brickwork layers is used ($L_s = 4$) to keep this portion of the circuit shallow, while the all-to-all entangling portion of the circuit is made up of $L_c = N$ layers. The parameterized single qubit rotations within each layer are selected to be one rotation about x followed by one rotation about y , and the entangled gates are selected to be controlled z (CZ) rotations. All simulations are performed numerically using PennyLane [65], using the Adam optimizer built into PyTorch for all optimization [99, 100]. Lastly, note that parameter convergence (evaluated every 100 iterations) is used as the stopping criterion for both the fragmented circuit and full circuit optimization, with a maximum of 5000 iterations permitted.

To assess the performance of pre-training using circuit fragmentation, the same circuit is optimized using random initial values (referred to as ‘vanilla VQE’). In this scenario, the parameters that correspond to the shallow, brickwork portion of the circuit are drawn from a uniform distribution ranging between $[0, 1]$. The all-to-all entangling portion of the gate is initialized near identity, as is done for the pre-training scheme. Figure 8 provides a case-by-case comparison between fragment-initialized VQE and vanilla VQE for 500 such models, for circuits of up to 15 qubits. In the top panels, the final percent error $\epsilon = (E - E_0)/|E_0|$ (where E_0 is the true ground state energy) is plotted for both approaches, along with the geometric mean of the results.

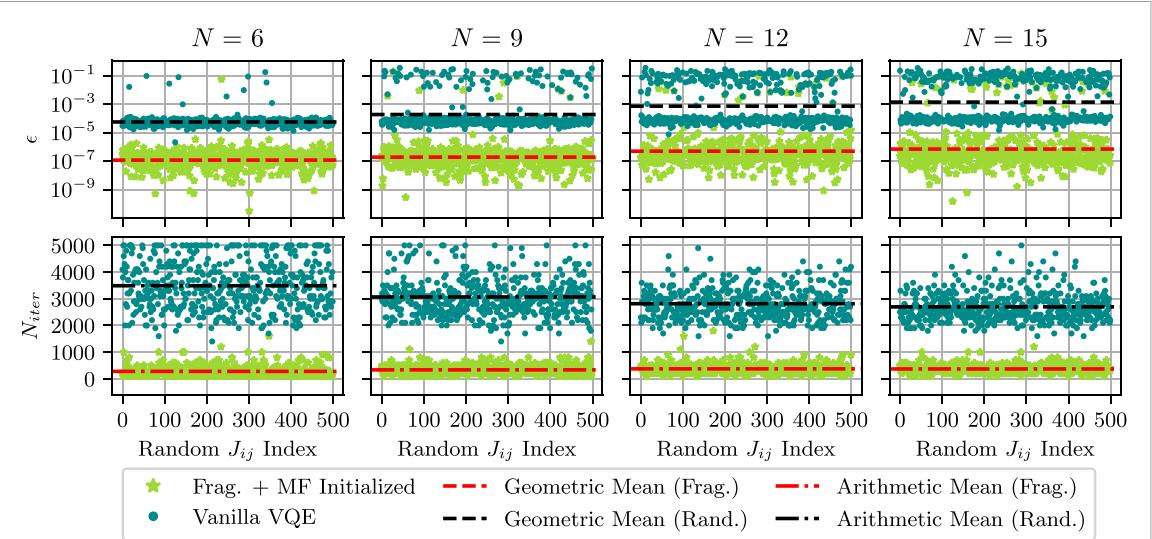


Figure 8. Comparison between fragment pre-trained VQE and vanilla VQE for 500 different J_{ij} matrices (graphs). The full PQC is split into fragments with $N_a = 2$ and at most $N_f = 3$ during pre-training. A total of $T = 10$ different partitionings are considered, and the best pre-trained solution is used to initialize the final optimization. The final percent error ϵ is provided in the top panel, while the required number of iterations N_{iter} to reach convergence is provided in the bottom panel. For the fragment initialized case, these metrics refer to the full circuit training that occurs after pre-training. Fragment pre-training reduces the geometric mean of ϵ by orders of magnitude, even as the system size increases. Likewise, the mean number of required iterations is reduced by nearly an order of magnitude.

The geometric mean of the fragment-initialized final error lies roughly three orders of magnitude below that of the vanilla VQE, with this gap growing even larger with increasing system size. For the larger system sizes, the vanilla VQE struggles to find a solution having $\epsilon < 10^{-2}$, while the fragment-initialized approach reaches $\epsilon \sim 10^{-7}$ for the same problem Hamiltonian. Moreover, using the same stopping criterion, the fragment-initialized VQE reaches this solution in fewer iterations (N_{iter}), decreasing the average number by nearly an order of magnitude, as illustrated by the bottom panels of figure 8. After successful pre-training, the parameters of the stitched-together circuit produce a loss that is already in the neighborhood of the minimum, so fewer iterations are required to reach convergence. We note that this corresponds to a significant reduction in the required execution time on quantum hardware, which can be prohibitively long to fully train a VQA [101]. For this simulation, we employ a batched optimization of T different fragmented circuits performed in parallel. In this approach, T different circuit partitions are randomly generated, and the resulting ensembles of fragmented circuits are optimized in parallel. The trained ensemble with lowest error is then used to initialize the full circuit. See appendix D.1 for an in-depth description of this approach and discussion of performance with batch size T .

5.2.2. Solving maxcut with mean-field terms

Our modification of fragmented loss functions to replace missing (that is, inaccessible) interactions with mean-field terms is critical to the success of pre-training. We here demonstrate that when there is no transverse field (as is the case for Ising-like Hamiltonians that map to classical graph problems), mean-field replacement of interactions results in a ground state and ground state energy that coincide with that of the exact Hamiltonian. This can be shown using a simple logical argument. First, it is well-established that the ground state of a classical Ising Hamiltonian will be a computational basis state—indeed, this is why the ground state can be mapped to the solution of a classical problem. We denote the ground state by $|x^*\rangle$. The ground state energy is simply a sum of the expected values of weighted ZZ interactions, taken with respect to the computational basis state $|x^*\rangle$: $E_g = -\sum_{\langle i,j \rangle} J_{ij} \langle x^* | \hat{S}_z^{(i)} \hat{S}_z^{(j)} | x^* \rangle$. Notice that for any computational basis state $|x\rangle$, the value of the expectation of a ZZ interaction exactly equals the value of the product of the expectation of the individual Z operators; that is, $\langle x | \hat{S}_z^{(i)} \hat{S}_z^{(j)} | x \rangle = \langle x | \hat{S}_z^{(i)} | x \rangle \langle x | \hat{S}_z^{(j)} | x \rangle$. Thus, if any weighted interaction $J_{ij} \langle \hat{S}_z^{(i)} \hat{S}_z^{(j)} \rangle$ is replaced by its mean-field counterpart $J_{ij} \langle \hat{S}_z^{(i)} \rangle \langle \hat{S}_z^{(j)} \rangle$, the resultant energy is unchanged: $E_g = \langle x^* | H | x^* \rangle = \langle x^* | H_{\text{MF}}(|x^*\rangle) | x^* \rangle$, where H_{MF} is the union of the fragmented, mean-field corrected Hamiltonians $\{H_{\text{MF}}^{(f)}\}$ and we have explicitly included the state dependence due to the presence of mean-field terms. Having established this fact, we must now show that $|x^*\rangle$ is the ground state of H_{MF} , such that $\langle x^* | H_{\text{MF}}(|x^*\rangle) | x^* \rangle \leq \langle \psi | H_{\text{MF}}(|\psi\rangle) | \psi \rangle \forall |\psi\rangle$. Observe that the quantity $\langle \hat{S}_z^{(i)} \hat{S}_z^{(j)} \rangle$ is bounded by $\pm 1/4$ and equals one of these extremum values for any computational basis state. The mean-field counterpart

$\langle \hat{S}_z^{(i)} \rangle \langle \hat{S}_z^{(j)} \rangle$ shares the same bounds; therefore, we cannot expect any state $|\psi\rangle$ to produce a smaller energy $\langle \psi | H_{\text{MF}}(|\psi\rangle) | \psi \rangle$ than $|x^*\rangle$, the ground state of the full Hamiltonian.

The above is a central reason for the success of our fragmented training: for Ising-like models with zero transverse field, optimizing a Hamiltonian with mean-field corrections will solve the original problem mapped to the full Hamiltonian. Two potential error sources can arise: (1) the state produced by stitching the optimized circuits together can differ from the output of the individual circuits, and (2) the fragmented optimization may have limited success, e.g. by landing in a local minimum or stalling in a barren plateau. A balance should be struck between these complications: the first error source can be mitigated by considering larger fragments with a larger number of auxiliary qubits or possibly by limiting the number of inter-fragment unitaries, as done in [102], while the second can be mitigated by considering smaller fragments with fewer circuit parameters.

5.2.3. Mean-field terms as first-order perturbation corrections

We now use perturbation theory to elucidate our technique of replacing multi-qubit interactions with mean fields when $h \neq 0$. In the previous section, it is established that the ground state and ground state energy of an Ising-like Hamiltonian with zero transverse fields remain unchanged when one or more of the interactions are replaced by the corresponding mean-field approximation term. Following a similar argument, one can further establish that the computational basis states are stationary states of the mean-field corrected Hamiltonian $H_{\text{MF}}(|\psi\rangle)$, and therefore $H_{\text{MF}}(|\psi\rangle)$ and the unaltered Hamiltonian H share the same spectrum and set of eigenstates (although this term is used loosely for $H_{\text{MF}}(|\psi\rangle)$, as the dependence on $|\psi\rangle$ causes the stationary Schrödinger equation to deviate from a linear eigenvalue problem).

In this section, we consider adding a small transverse field to the classical Ising-like model, propelling the problem into the quantum domain. The first-order corrections to the ground state $|x^*\rangle$ and ground state energy E_g are computed using perturbation theory. The case of the mean-field corrected Hamiltonian $H_{\text{MF}}(|\psi\rangle)$ is treated with a version of perturbation theory modified to accommodate mean-field terms, and notably, the same first-order corrections to $|x^*\rangle$ and E_g are recovered. For a full derivation, please refer to appendix C.

Adding a transverse field, the unaltered Hamiltonian containing all interactions is given by:

$$H = H_0 + H_I + \lambda V, \quad (12)$$

where H_0 contains the intra-fragment interactions:

$$H_0 = - \sum_{\langle i,j \rangle \notin I} J_{ij} \hat{S}_z^{(i)} \hat{S}_z^{(j)}, \quad (13)$$

H_I contains the inter-fragment interactions:

$$H_I = - \sum_{\langle i,j \rangle \in I} J_{ij} \hat{S}_z^{(i)} \hat{S}_z^{(j)}, \quad (14)$$

V contains the perturbing transverse field:

$$V = -h \sum_i \hat{S}_x^{(i)}, \quad (15)$$

and λ is a perturbation parameter. We remind the reader that the inter-fragment interactions H_I are those that will be replaced by mean-field corrections.

In contrast, the mean-field corrected Hamiltonian denoted H_{MF} is given by:

$$H_{\text{MF}}(|\psi\rangle) = H_0 + H_{I, \text{MF}}(|\psi\rangle) + \lambda V, \quad (16)$$

where the form of the Hamiltonian now depends on the state of the system due to the mean-field corrections:

$$H_{I, \text{MF}}(|\psi\rangle) = - \sum_{\langle i,j \rangle \in I} J_{ij} \hat{S}_z^{(i)} \langle \psi | \hat{S}_z^{(j)} | \psi \rangle. \quad (17)$$

Before any corrections can be computed, it is imperative to establish the correct zeroth order energies and eigenstates for each Hamiltonian. Following perturbation theory, the zeroth order eigenstates of H and H_{MF} generally equal those of the unperturbed counterparts (that is, taking $h = 0$); these coincide with the set of computational basis states $\{|x\rangle\}$ —including the unperturbed ground state, $|x^*\rangle$. However, there are degeneracies in the unperturbed Hamiltonians, and thus, degenerate perturbation theory is required.

When the unperturbed spectrum contains degeneracies, the proper linear combinations of the unperturbed eigenstates forming the degenerate subspace must be determined; these are the states that the perturbed eigenstates approach as $h \rightarrow 0$. The unperturbed Ising-like model possesses \mathbb{Z}_2 symmetry. Practically, this means that for each eigenstate $|x\rangle$, the ‘flipped’ eigenstate $|\bar{x}\rangle := \bigotimes_i X_i |x_i\rangle$ is degenerate. For the unaltered Ising-like model H , the proper zeroth order eigenstates for the degenerate subspace containing the ground state are given by $|\pm_{x^*}\rangle = \frac{1}{\sqrt{2}}(|x^*\rangle \pm |\bar{x}^*\rangle)$. The transverse field will break the ground state degeneracy of H , and the positive superposition $|+_{x^*}\rangle$ is preferred by the ground state.

Shifting attention to the mean-field corrected Hamiltonian H_{MF} , the stationary Schrödinger equation is no longer linear in $|\psi\rangle$, and the linearity that characterizes quantum mechanics no longer applies. The notion of finding proper linear combinations is not an appropriate procedure due to the problem’s nonlinearity. In particular, superpositions of degenerate eigenstates can yield different energies for H_{MF} and thus effectively exist outside the degenerate subspace.

To illustrate this, consider a single mean-field factor, $\langle \hat{S}_z^{(i)} \rangle$, such as those within H_{MF} . While the expectation value of this quantity with respect to a computational basis state $|x^*\rangle$ yields

$$\langle x^* | \hat{S}_z^{(i)} | x^* \rangle = \frac{1}{2} (-1)^{x_i^*}, \quad (18)$$

evaluating the same term with respect to $|+_{x^*}\rangle$ leads to the term vanishing as

$$\begin{aligned} \langle +_{x^*} | \hat{S}_z^{(i)} | +_{x^*} \rangle &= \frac{1}{2} \left(\langle x^* | \hat{S}_z^{(i)} | x^* \rangle + \langle x^* | \hat{S}_z^{(i)} | \bar{x}^* \rangle + \langle \bar{x}^* | \hat{S}_z^{(i)} | x^* \rangle + \langle \bar{x}^* | \hat{S}_z^{(i)} | \bar{x}^* \rangle \right) \\ &= \frac{1}{4} \left((-1)^{x_i^*} + (-1)^{\bar{x}_i^*} \right) \\ &= 0. \end{aligned} \quad (19)$$

Notably for the ground state of the unperturbed Hamiltonian $|x^*\rangle$, this means that the pure computational basis states $|x^*\rangle, |\bar{x}^*\rangle$ are energetically preferred over any linear combination of them. Thus, for H_{MF} , the computational basis states remain the proper zeroth order eigenstates with a perturbative transverse field.

After establishing the zeroth order eigenstates and eigenenergies ($|k^{(0)}\rangle$ and $E_k^{(0)}$, respectively) of conventional Hamiltonians such as equation (12), perturbation theory proceeds by expanding $|k\rangle$ and E in λ in the stationary Schrödinger equation and equating orders of λ :

$$\begin{aligned} (H_0 + H_1 + \lambda V) \left(|k^{(0)}\rangle + \lambda |k^{(1)}\rangle + \lambda^2 |k^{(2)}\rangle + \dots \right) \\ = \left(E_k^{(0)} + \lambda E_k^{(1)} + \lambda^2 E_k^{(2)} + \dots \right) \left(|k^{(0)}\rangle + \lambda |k^{(1)}\rangle + \lambda^2 |k^{(2)}\rangle + \dots \right). \end{aligned} \quad (20)$$

Following this procedure for H and carefully treating the degeneracy, the first-order energy correction $E_k^{(1)}$ vanishes and the first-order eigenstate correction takes the form:

$$|k^{(1)}\rangle = \sum_{m \notin D_k} \frac{\langle m^{(0)} | V | k^{(0)} \rangle}{E_k^{(0)} - E_m^{(0)}} |m^{(0)}\rangle, \quad (21)$$

where D_k represents the degenerate subspace that $|k^{(0)}\rangle$ occupies.

To derive the analogous correction to H_{MF} , we employ a modified approach to perturbation theory that can accommodate the nonlinearity of the stationary Schrödinger equation. In particular, the expanded form of $|k_{\text{MF}}\rangle$ is explicitly inserted into the state-dependant terms of H_{MF} prior to equating orders of λ to compute the corrections. Following this procedure, the first order energy correction $E_{k,\text{MF}}^{(1)}$ again vanishes, and the first order eigenstate correction takes on an identical form to that of H :

$$|k_{\text{MF}}^{(1)}\rangle = \sum_{m \notin D_k} \frac{\langle m_{\text{MF}}^{(0)} | V | k_{\text{MF}}^{(0)} \rangle}{E_{k,\text{MF}}^{(0)} - E_{m,\text{MF}}^{(0)}} |m_{\text{MF}}^{(0)}\rangle. \quad (22)$$

There is one crucial difference between equations (21) and (22): the zeroth order eigenstates $|m^{(0)}\rangle$ and $|m_{\text{MF}}^{(0)}\rangle$. For the full Hamiltonian, each $|m^{(0)}\rangle$ has the form $|\pm_y\rangle \propto (|y\rangle \pm |\bar{y}\rangle)$, while each $|m_{\text{MF}}^{(0)}\rangle$ is a single computational basis state, $|y\rangle$. This leads to the fidelity between the first order ground state $|\psi_{g,\text{MF}}\rangle \propto |x^*\rangle + |x_{\text{MF}}^{*(1)}\rangle$ and that of the full Hamiltonian $|\psi_g\rangle \propto |+_{x^*}\rangle + |+_{x^*}^{(1)}\rangle$ to be $F = 0.5$ rather than perfect unity (see appendix C.3). Nonetheless, half overlap provides significant information about the true ground state for pre-training.

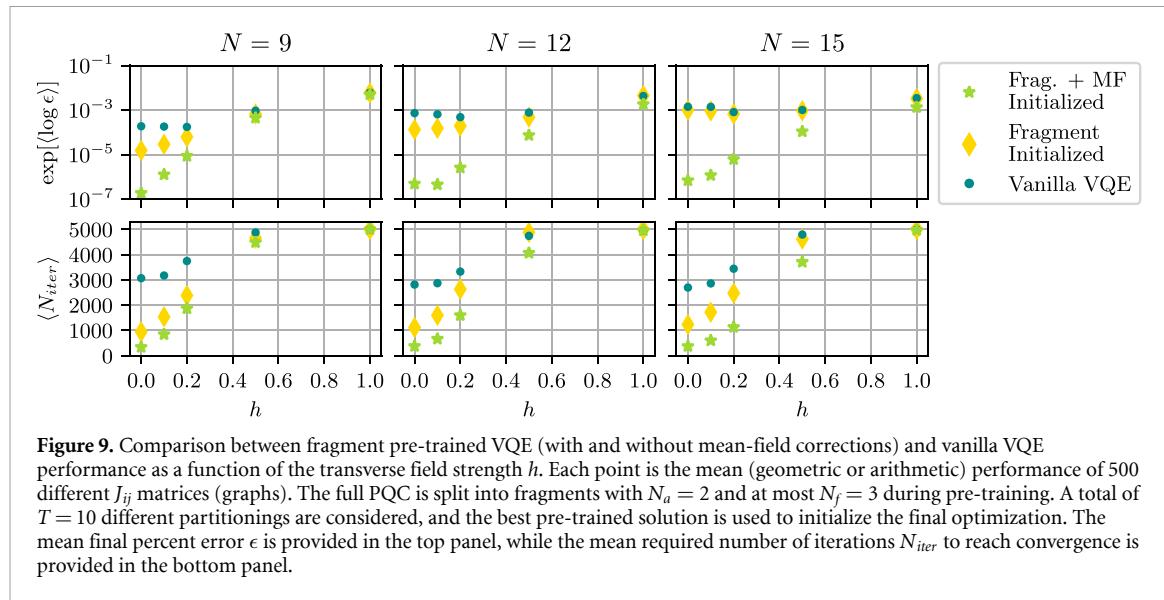


Figure 9. Comparison between fragment pre-trained VQE (with and without mean-field corrections) and vanilla VQE performance as a function of the transverse field strength h . Each point is the mean (geometric or arithmetic) performance of 500 different J_{ij} matrices (graphs). The full PQC is split into fragments with $N_a = 2$ and at most $N_f = 3$ during pre-training. A total of $T = 10$ different partitionings are considered, and the best pre-trained solution is used to initialize the final optimization. The mean final percent error ϵ is provided in the top panel, while the mean required number of iterations N_{iter} to reach convergence is provided in the bottom panel.

In figure 9, we examine the mean performance of the pre-training scheme as a function of the transverse field strength h . The case of neglecting mean-field terms during pre-training is also considered to highlight the vital role these corrections play at small values of h . When $h = 0$, the model is classical, and the global fragmented Hamiltonian with mean-field corrections shares the same ground state and ground state energy of the full model, as discussed in section 5.2.2. This leads to remarkable pre-training performance, even as N is increased. The fragment initialization that neglects mean-field terms is not guaranteed to share the ground state of the full model—in fact, the two are likely to be orthogonal. The pre-training will still feature low entanglement, which likely explains why the mean-field free initialization scheme outperforms random initialization, but overall, the average error exceeds that of the mean-field corrected case by orders of magnitude, particularly as N is increased. When a small transverse field is added to the model, the half overlap between the fragmented and full Hamiltonian ground states provided by including mean-field corrections leads to error orders of magnitude smaller than the other approaches. Only as h is further increased—entering the regime where first-order perturbation theory is inadequate to describe the ground state—do the approaches begin to perform comparably, with increasing error and required iterations.

6. Conclusion

We have presented two near-term approaches to the distributed Hamiltonian evolution of a quantum system and a pre-training technique for variational quantum circuits. Our time evolution schemes are built upon the idea that the relative importance of interactions spanning a sub-system and its environment can be ascertained using the principle of minimizing the short-time evolution error, which is derived to be proportional to the quantum variance of the difference between the full and fragmented Hamiltonians. The first scheme employs only classical information transfer in the form of mean-field measurements to update mean-field corrections, as well as a limited number of auxiliary qubits anchored to each fragment, enabling limited entanglement growth. Although our method is lossy, metrics local to the system qubits addressed by a single fragment can closely mimic the true values from exact evolution, including the gold standard comparison of state fidelity. Moreover, this scheme is flexible, as it is amenable to any mixture of classical and quantum hardware and can process the fragments in series or parallel. In our second scheme, the information stored by qubits designated to be auxiliaries is physically shared between fragments, either through qubit shuttling or quantum teleportation. This approach is appropriate when quantum hardware is available and limited quantum communication is feasible. If desired, the choice of which qubits act as auxiliaries can be updated from one time step to the next, as dictated by the minimum error rule, to extend the performance of the approximate scheme.

Finally, we examine how our fragmented simulation scheme can be modified to apply to quantum circuits. Here, a single circuit is fractured into several smaller overlapping circuits, which are more manageable (requiring lower connectivity and less prone to suffer from noise and barren plateaus) and, if sufficiently small, even classically treatable. We devise a scheme that employs fragmented circuits to pre-train the parameters of the full PQC. Crucially, the use of overlapping registers coupled with the mean-field corrective terms in the loss function links the optimization of the individual circuits. The inclusion of

mean-field corrections shifts the solution of the collective circuit optimization to have a large overlap with the solution of the full problem. We demonstrate that the pre-training scheme reduces the final percent error by orders of magnitude as well as the number of iterations required when compared to randomly initialized full circuit optimizations of VQE. Although the scheme's performance is particularly strong for classical Ising Hamiltonians, we develop a non-linear perturbation theory to analytically show that the mean-field terms included in optimization act as first-order perturbation corrections when a small transverse field is added, extending the success of the scheme into the quantum realm.

This manuscript motivates and facilitates numerous future research directions. Although we emphasized limited quantum information transfer, subsequent studies might explore how the number of auxiliary qubits and the frequency of re-encoding affect distributed simulation, or devise the details of physically implementing a limited quantum channel. Likewise, higher-order moments beyond mean-field terms may be explored as higher-order corrections. Moreover, rather than using auxiliary qubits to target specific environment qubits, a method of mapping salient environment *states* to auxiliary qubits (as employed by some classical fragmentation methods such as DMET [57]) may further improve the method, although the measurement-efficiency of such a technique may prove challenging in a quantum setting. Regarding fragment pre-training for variational algorithms, future works might develop efficient circuit fragmentations that are tailored to specific problem Hamiltonians and/or symmetries, rather than our more general, batched approach. Finally, alternative partitioning schemes might be considered to enable pre-training of non-brickwork circuits.

This work represents a pivotal stepping stone on the path to large-scale distributed quantum computing. In the near term, our distributed computing method with classical channels can be implemented by a single small simulator in sequence, or by a collection of small simulators that are either quantum or classical in nature. This permits the simulation of large system quantum dynamics without the noise and connectivity concerns of a large-scale quantum device [8], allowing experimentalists to address challenging problems in quantum chemistry and condensed matter physics [103, 104]. As non-local operations on quantum hardware improve, our proposal for limited quantum information transfer can be implemented, enabling cross-simulator measurements and higher accuracy. Lastly, our fragmented pre-training method can reduce the error of large-scale variational quantum algorithms by orders of magnitude while reducing the number of training epochs. Such improvements are vital to this field, which seeks to address problems ranging from drug discovery to NP-hard optimization on quantum hardware despite persistent training difficulties [47, 69, 80, 105, 106].

Data availability statement

The data that support the findings of this study will be openly available following an embargo at the following URL/DOI:<https://github.com/amcclaingomez3/MF-DQC>.

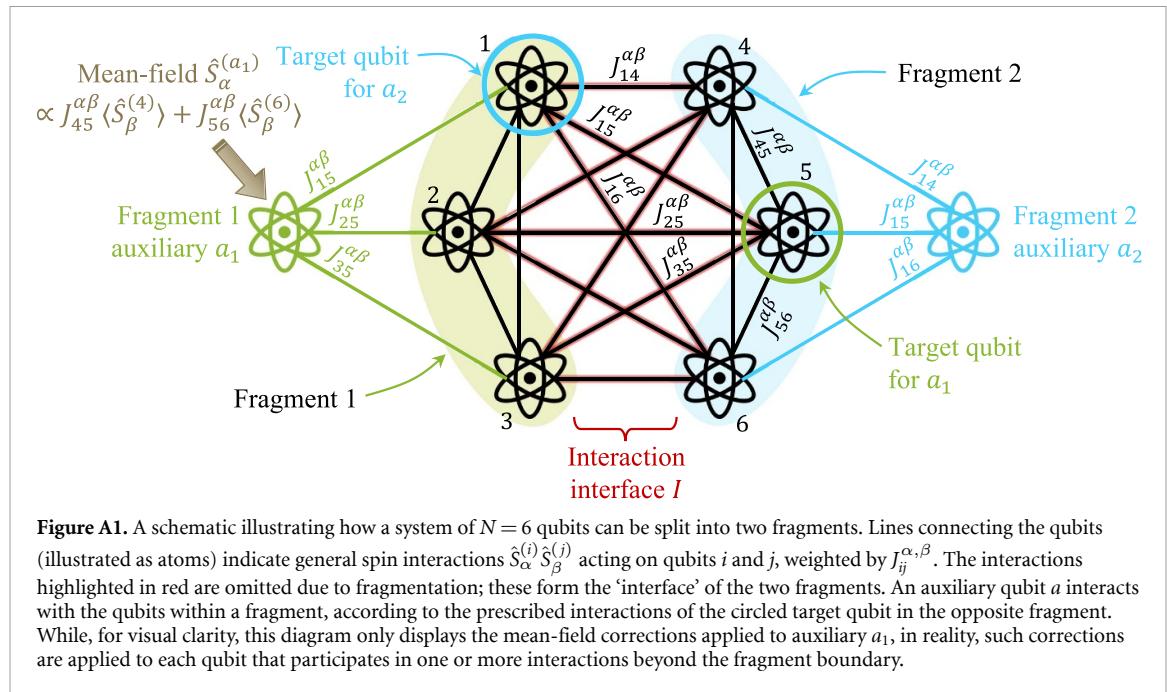
Acknowledgments

This work was done during A M G's internship at NVIDIA. A M G acknowledges support from the NSF through the Graduate Research Fellowships Program (DGE 2140743) and from the Theodore H Ashford Fellowships in the Sciences. At CalTech, A A is supported in part by the Bren-endowed chair. S F Y thanks the AFOSR (FA9550-19-1-0233) and the NSF through the Qu-IDEAS HDR Institute (OAC-2118310), the CUA PFC (PHY-2317134), and Q-SEnSE QLCI (OMA-2016244) for funding.

Appendix A. Fragmentation for time evolution

A.1. A toy model

For added clarification, here we consider a toy model of $N = 6$ qubits fragmented into two groups such that $N_f = 3$ (see figure A1, a more explicitly labeled version of figure 1(b)). One auxiliary qubit is included for each fragment ($N_a = 1$). The auxiliary qubit of Fragment 1, a_1 , targets qubit 5 in Fragment 2. Similarly, the auxiliary qubit of Fragment 2, a_2 , targets qubit 1 in Fragment 1. In the figure, the interactions mediated by the auxiliary qubits are explicitly labeled to illustrate how each auxiliary qubit plays the role of one qubit from the environment. Which environment qubit is selected for this role will depend on the quantum variance $\text{var}(H - H_1^{(f)})$ (see section 2.2). The mean-field corrections applied to a_1 are included in the figure to account for the missing interactions with qubits 4 and 6 that the target qubit (qubit 5) would participate in if all qubits were present. Although only one such arrow appears in the figure, these corrections exist for each



qubit participating in fewer interactions than it would in the full system, regardless of whether the qubit is categorized as a fragment qubit or an auxiliary targeting an environment qubit.

A.2. Fragmentation error

Consider the error caused by the omission of interactions that form the interface of some particular fragment f and the remaining environment spins E . This Hamiltonian which we denote $H_I^{(f)}$ is subtly different than the previously introduced $H^{(f)}$, as it includes operators acting in the space of E in order to isolate the error caused by the section of the interaction interface produced by a particular fragment f :

$$H_I^{(f)} = H^{(f)} + H^{(E)} \\ = - \sum_{\langle i, j \rangle / \in I} \sum_{\alpha, \beta} J_{ij}^{\alpha, \beta} \hat{S}_\alpha^{(i)} \hat{S}_\beta^{(j)} - \sum_i^N h_i \hat{S}_x^{(i)}. \quad (\text{A.1})$$

The fidelity of the state evolved under $H_I^{(f)}$ with that evolved under H is given by:

$$F(t) = |\langle \Psi | U^\dagger(t) U_I^{(f)}(t) | \Psi \rangle|^2, \quad (\text{A.2})$$

where $U(t) = \exp(-iHt)$, $U_I^{(f)}(t) = \exp(-iH_I^{(f)}t)$, and $|\Psi_0\rangle$ is the current state of the system [66].

To analyze the fidelity metric, we can combine the product $U^\dagger(t) U_I^{(f)}(t)$ into a single exponential $\exp(Z)$. This final exponential argument is obtained using the BCH formula [61], which provides an expression for Z in terms of the nested commutators of the individual exponential arguments of $U^\dagger(t)$ and $U_I^{(f)}(t)$:

$$Z = it \left(H - H_I^{(f)} \right) + \frac{1}{2} t^2 \left[H, H_I^{(f)} \right] + \frac{1}{12} t^3 \left(i \left[H, \left[H, H_I^{(f)} \right] \right] + (-i) \left[H_I^{(f)}, \left[H_I^{(f)}, H \right] \right] \right) \\ - \frac{1}{24} t^4 \left[H_I^{(f)}, \left[H, \left[H, H_I^{(f)} \right] \right] \right] + \dots \quad (\text{A.3})$$

Notice that each subsequent term in the BCH formula has a higher order in t . To approximate the error, we keep the terms up to small orders in t —specifically, we can write $Z = \sum_{n=1}^{\infty} z_n t^n$, where the coefficients z_n do not depend on t and generally do not commute with each other, and truncate the sum after some n . We note that the first and second order z_n are given by:

$$z_1 = i \left(H - H_I^{(f)} \right), \quad (\text{A.4})$$

$$z_2 = \frac{1}{2} \left[H, H_I^{(f)} \right]. \quad (\text{A.5})$$

Taylor expanding the exponential $\exp(Z)$ to second order in t :

$$\begin{aligned} U^\dagger(t) U_I^{(f)}(t) &= \exp\left(\sum_{n=1}^{\infty} z_n t^n\right) \\ &= 1 + \sum_{n=1}^{\infty} z_n t^n + \frac{1}{2} \left(\sum_{n=1}^{\infty} z_n t^n\right)^2 + \dots \\ &= 1 + z_1 t + t^2 \left(z_2 + \frac{1}{2} z_1^2\right) + \mathcal{O}(t^3). \end{aligned} \quad (\text{A.6})$$

The fidelity is then the modulus square of the expectation of the above expression, taken with respect to the initial state of the full system:

$$\begin{aligned} F(t) &\approx \left|1 + \langle z_1 \rangle t + t^2 \left(\langle z_2 \rangle + \frac{1}{2} \langle z_1^2 \rangle\right)\right|^2 \\ &= 1 + t(\langle z_1 \rangle + \langle z_1 \rangle^*) + t^2 \left(|\langle z_1 \rangle|^2 + \langle z_2 \rangle + \langle z_2 \rangle^* + \frac{1}{2} \langle z_1^2 \rangle + \frac{1}{2} \langle z_1^2 \rangle^* + \mathcal{O}(t^3)\right). \end{aligned} \quad (\text{A.7})$$

Plugging in the expressions for z_1 and z_2 :

$$\begin{aligned} F(t) &\approx 1 + t \left(i \left\langle H - H_I^{(f)}\right\rangle - i \left\langle H - H_I^{(f)}\right\rangle^*\right) + t^2 \left(|\left\langle H - H_I^{(f)}\right\rangle|^2 + \frac{1}{2} \left\langle \left[H, H_I^{(f)}\right]\right\rangle + \frac{1}{2} \left\langle \left[H, H_I^{(f)}\right]\right\rangle^* \right. \\ &\quad \left. + (i)^2 \frac{1}{2} \left\langle \left(H - H_I^{(f)}\right)^2\right\rangle + (-i)^2 \frac{1}{2} \left\langle \left(H - H_I^{(f)}\right)^2\right\rangle^*\right). \end{aligned} \quad (\text{A.8})$$

Notice that, as both H and $H_I^{(f)}$ are Hermitian, $\langle H - H_I^{(f)}\rangle^* = \langle H \rangle^* - \langle H_I^{(f)}\rangle^* = \langle H \rangle - \langle H_I^{(f)}\rangle$. Thus, the terms first order in t cancel with one another. The second-order terms can be simplified by expanding them:

$$\begin{aligned} F(t) &\approx 1 + t^2 \left(|\left\langle H - H_I^{(f)}\right\rangle|^2 + \frac{1}{2} \left(\left\langle HH_I^{(f)}\right\rangle - \left\langle H_I^{(f)}H\right\rangle\right) + \frac{1}{2} \left(\left\langle HH_I^{(f)}\right\rangle^* - \left\langle H_I^{(f)}H\right\rangle^*\right) \right. \\ &\quad \left. - \frac{1}{2} \left(\langle H^2 \rangle - \left\langle HH_I^{(f)}\right\rangle - \langle H_I^{(f)}H \rangle + \left\langle \left(H_I^{(f)}\right)^2\right\rangle\right) \right. \\ &\quad \left. - \frac{1}{2} \left(\langle H^2 \rangle^* - \left\langle HH_I^{(f)}\right\rangle^* - \left\langle H_I^{(f)}H\right\rangle^* + \left\langle \left(H_I^{(f)}\right)^2\right\rangle^*\right)\right). \end{aligned} \quad (\text{A.9})$$

Using the fact that $\langle AB \rangle^* = \langle B^\dagger A^\dagger \rangle$ to simplify, we arrive at a compact expression for the fidelity to second order in t :

$$\begin{aligned} F(t) &= 1 + t^2 \left(|\left\langle H - H_I^{(f)}\right\rangle|^2 - \left\langle \left(H - H_I^{(f)}\right)^2\right\rangle\right) + \mathcal{O}(t^3) \\ &\approx 1 - \text{var}(H - H_I^{(f)}) t^2. \end{aligned} \quad (\text{A.10})$$

The error $\epsilon(t) = 1 - F(t)$ is thus given by $\text{var}(H - H_I^{(f)}) t^2$ for short times t .

A.3. Numerical simulation of a quantum channel

When simulators are linked via a quantum channel, the states of the fragments no longer live in separate Hilbert spaces; they comprise the state of the collective system, which lives in the larger Hilbert space of size 2^N . To simulate this numerically, we evolve a modified Hamiltonian that lives in the full Hilbert space. This Hamiltonian involves adjusted connectivity matrices $J_{ij}^{\alpha\beta}$ to match the connectivity provided by the current auxiliary encoding. That is, defining a set S_f for each fragment containing the list of fragment qubits and target qubits for auxiliaries of fragment f :

$$S_f = \{i \in f\} \cup \{a_f\}, \quad (\text{A.11})$$

we implement the union of connectivity matrices $\cup J_{ij}^{\alpha\beta, (f)}$ for each $\alpha\beta$ interaction type, defined by:

$$J_{ij}^{\alpha\beta, (f)} = \begin{cases} J_{ij}^{\alpha\beta} & \text{if } i, j \in S_f \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.12})$$

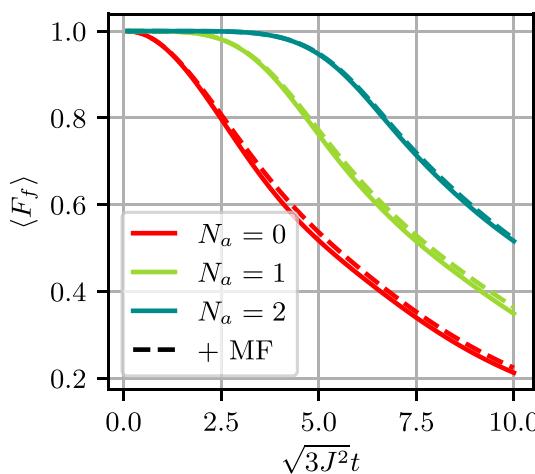


Figure A2. Nearest neighbor Heisenberg model with constant $J = 1.0$. $N = 12$ qubits are split into two fragments, and the fidelity between the fragment qubits' state and the exactly evolved system is plotted for various numbers of auxiliary qubits, with (dashed lines) and without (solid lines) mean-field corrections. The fidelity is averaged over 100 random computational basis initial states. Performance progressively increases with increasing N_a and the addition of mean-field corrections.

Any interactions still zeroed in the union $\bigcup J_{ij}^{\alpha\beta,(f)}$ are approximately included via mean-field corrections. If the auxiliary encoding is updated, the union $\bigcup J_{ij}^{\alpha\beta,(f)}$ will change due to the change in selected auxiliaries $\{a_j\}$ of each fragment.

A.4. Fragmented simulation of a Heisenberg model

In this section, we report results for the time simulation of the Heisenberg model. We consider a 1D Heisenberg model with uniform coupling J , governed by the following Hamiltonian:

$$H_{\text{Heis}} = -J \sum_{\langle i,j \rangle \in n.n.} \left(\hat{S}_x^{(i)} \hat{S}_x^{(j)} + \hat{S}_y^{(i)} \hat{S}_y^{(j)} + \hat{S}_z^{(i)} \hat{S}_z^{(j)} \right). \quad (\text{A.13})$$

Note that this model features significantly larger interaction as compared to the Ising model previously considered. To adjust for this increased interaction, we modify the effective simulation time to be $\sqrt{3J^2}t$. Figure A2, presents the fragment fidelity for the fragmented evolution of the Heisenberg model. Repeating the setup of figure 2(a), we split a system of 12 qubits into 2 fragments and evolve the system with different numbers of auxiliary qubits, both with and without mean-field corrections. The results provided are averaged over 100 different initial states, sampled from computational basis states. The general trend again matches that of figure 2(a), with progressively better performance for larger N_a . However, we note that the mean-field corrections provide less benefit due to the increased interaction likely leading to a more entangled structure that cannot be mimicked by mean-field corrections. This concept is explored further in appendix B.

Appendix B. When mean-field corrections are beneficial to time evolution

In this appendix, we numerically investigate the role of mean-field corrections in time evolution, showing that for the majority of distributed states, mean-field corrections reduce the fragmentation error. The second order fragmentation error is derived in the main text to be the variance of the difference between H and $H_1^{(f)}$ (see equation (6)). In this appendix, we will denote this quantity by V . If we include mean-field corrections in the fragmented Hamiltonian, the second order fragmentation error can be amended to include additional terms:

$$V_{\text{MF}} := \text{var} \left(H - H_{1,\text{MF}}^{(f)} \right) = \text{var} \left(- \sum_{\langle i,j \rangle \in I} \sum_{\alpha,\beta} J_{ij}^{\alpha,\beta} \hat{S}_\alpha^{(i)} \left(\hat{S}_\beta^{(j)} - \langle \hat{S}_\beta^{(j)} \rangle \right) \right). \quad (\text{B.1})$$

The improved performance due to mean-field corrections present in the results of the main text can be attributed to the fact that in those scenarios, $V_{\text{MF}} < V$ on average, such that the error is decreased. However, situations can arise where $V_{\text{MF}} > V$. In what follows, we show that these situations generally occur when the qubits spanning the interface are significantly entangled with one another.

To gain insight into when mean-field corrections are beneficial, consider two fragments each containing a single qubit. The qubits interact according to a quantum Ising-like Hamiltonian, with $J_{12} = 4$ such that $J_{12}\hat{S}_z^{(1)}\hat{S}_z^{(2)}$ reduces to Pauli operators $\hat{\sigma}_z^{(1)}\hat{\sigma}_z^{(2)}$. The interface of the fragments is comprised of the single J_{12} linking the two qubits of the system. To explore the possible values of $V - V_{\text{MF}}$, we randomly 10 000 two-qubit states $|\psi\rangle$ from the Haar measure. A general two-qubit state can be written as:

$$|\psi\rangle = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle, \quad (\text{B.2})$$

where the complex coefficients c_{ij} are properly normalized. For each $|\psi\rangle$, the variance difference $V - V_{\text{MF}}$ is computed (recall that a positive $V - V_{\text{MF}}$ implies that mean-field corrections have reduced the simulation error). We additionally compute the concurrence to measure the entanglement between qubit 1 and qubit 2. The concurrence is defined to be [107]:

$$C(|\psi\rangle) = 2|c_{00}c_{11} - c_{01}c_{10}|, \quad (\text{B.3})$$

where $C(|\psi\rangle) = 0$ for any pure state $|\psi\rangle$, and $C(|\psi\rangle)$ monotonically increases with entanglement to a maximum value of $C(|\psi\rangle) = 1$ (e.g. for a maximally entangled Bell pair). The results are plotted in light green in figure B1. The scattered data falls within a closed area, with $V - V_{\text{MF}}$ dropping below zero for concurrence $0 < C(|\psi\rangle) < 1$. The difference $V - V_{\text{MF}}$ is also bounded between -0.25 and 1 , reaching its maximum positive value for product states (where $C(|\psi\rangle) = 1$).

To understand the features of a state that extremizes $V - V_{\text{MF}}$, we probe the boundaries of the envelope of possible states by examining parameterized two-qubit states. The specific parameterized states discussed in the following paragraphs are meant to be representative of the form of the states that lie on the edge of the closed area in figure B1, but generally, many possible two-qubit states will produce identical values of $(C(|\psi\rangle), V - V_{\text{MF}})$.

First, we consider a parameterized state $|\psi(\alpha)\rangle$ that smoothly approaches the Bell state $|\Phi_+\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle)$:

$$|\psi(\alpha)\rangle := \sqrt{\alpha}|00\rangle + \sqrt{1-\alpha}|11\rangle. \quad (\text{B.4})$$

When $\alpha = 0$, the resulting state $|11\rangle$ is a product state with zero concurrence. Furthermore, $|11\rangle$ is computational basis state with $V = V_{\text{MF}} = 0$. As α approaches $1/2$, the concurrence increases, reaching a maximum value. The variance difference, on the other hand, decreases to take on negative values—see the left panel of figure B2. Thus, it is states with structured entanglement such as $|\psi(\alpha)\rangle$ which will not benefit from mean-field corrections. Plotting $V - V_{\text{MF}}$ versus the concurrence of $|\psi(\alpha)\rangle$ in figure B1, we see that this set of states lies on the lower boundary of the envelope of possible values.

Secondly, we consider the two-qubit state produced by rotating qubit 1 of the state $|00\rangle$ about x by an angle θ :

$$|\psi(\theta)\rangle := \cos(\theta/2)|00\rangle - i \sin(\theta/2)|10\rangle. \quad (\text{B.5})$$

The resulting state will always be a product state with zero concurrence; however, the variance difference will depend on θ . Sweeping across θ (see the middle panel of figure B2), we see that $V - V_{\text{MF}}$ grows as θ is increased. A state of this kind lies on the left boundary of the envelope (see figure B1).

Finally, we consider a state that lies on the top boundary of the envelope, with positive values for both the variance difference and the concurrence. We parameterize this by interpolating between a product state with maximum $V - V_{\text{MF}}$ and a state with maximum concurrence while avoiding a Bell state, which is known to lie on the lower boundary:

$$|\psi(\beta)\rangle := \frac{1}{\sqrt{2+2\beta^2}}(|00\rangle + \beta|01\rangle - |10\rangle + \beta|11\rangle). \quad (\text{B.6})$$

The right panel of figure B2 reveals that $V - V_{\text{MF}}$ and $C(|\psi(\beta)\rangle)$ remain non-negative as β is varied. Again, plotting $V - V_{\text{MF}}$ versus the concurrence of $|\psi(\beta)\rangle$ in figure B1 reveals that such a state indeed lies at the top boundary of the envelope.

In the time evolution applications considered in the main text, all initial states are computational basis states. It is thus unlikely that the state evolves into one with large entanglement across the interface with the particular structure required to produce a negative value of $V - V_{\text{MF}}$. This is why, on average, including mean-field corrections enhances the performance of the scheme. More generally, we expect that in practical situations where fragmentation is employed, it is unlikely that highly structured entanglement will form across a fragment interface through evolution. Therefore, it is reasonable to expect that mean-field corrections will still improve performance in general applications.

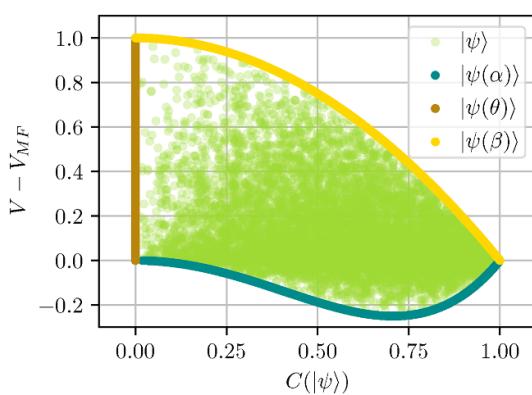


Figure B1. Variance difference $V - V_{MF}$ versus concurrence $C(|\psi\rangle)$ for 10 000 two-qubit states $|\psi\rangle$ randomly generated from the Haar measure. These quantities for specific parameterized states $|\psi(\alpha)\rangle$, $|\psi(\theta)\rangle$, and $|\psi(\beta)\rangle$ (defined in equations (B.4)–(B.6)) are plotted in separate colors. To calculate the variance difference, a single interaction $\hat{\sigma}_z^{(1)}\hat{\sigma}_z^{(2)}$ is considered.

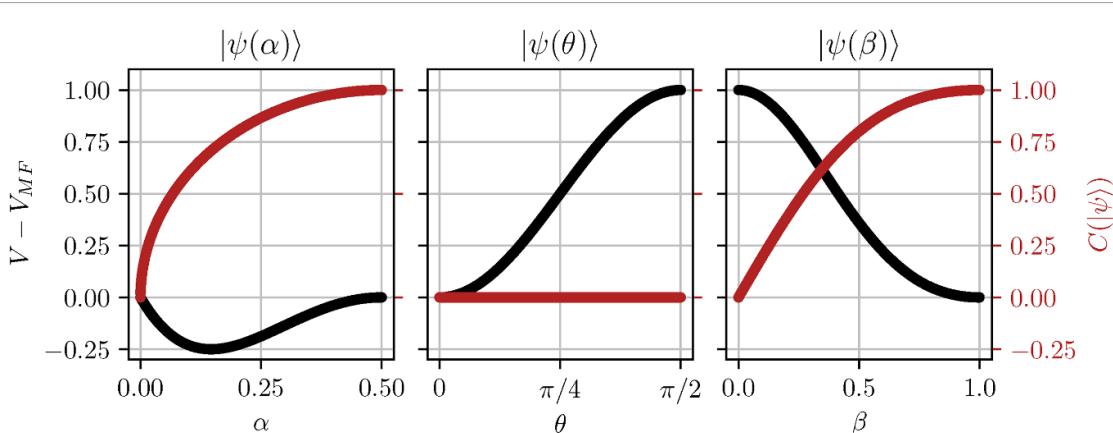


Figure B2. Variance difference $V - V_{MF}$ (left axis) and concurrence $C(|\psi\rangle)$ (right axis) plotted versus a state parameter for specific parameterized two-qubit states, specified by the subplot title. The parameterized states are defined in equations (B.4)–(B.6).

Appendix C. Adding a perturbative transverse field to the ising model

Consider perturbing an Ising-like Hamiltonian with a small transverse field h . We will first derive how the perturbation shifts the ground state energy and the corresponding eigenstate for the case of an unperturbed Hamiltonian containing all Ising-like interactions. Then, we will derive the shifts for the unconventional situation where certain interactions within the Hamiltonian are replaced by mean-field corrective terms. Understanding how these cases differ from each other will illuminate how the ground state solution of a mean-field corrected Hamiltonian strays from that of the full Hamiltonian as h grows, rendering the Ising-like model increasingly quantum.

We write down the Hamiltonian containing all interactions as:

$$H = H_0 + H_I + \lambda V, \quad (C.1)$$

where H_0 contains the intra-fragment interactions:

$$H_0 = - \sum_{\langle i,j \rangle \notin I} J_{ij} \hat{S}_z^{(i)} \hat{S}_z^{(j)}, \quad (C.2)$$

H_I contains the inter-fragment interactions:

$$H_I = - \sum_{\langle i,j \rangle \in I} J_{ij} \hat{S}_z^{(i)} \hat{S}_z^{(j)}, \quad (C.3)$$

V contains the perturbing transverse field:

$$V = -h \sum_i \hat{S}_x^{(i)}, \quad (C.4)$$

and λ is a perturbation parameter. We explicitly separate the interactions that will be replaced by mean-field corrections; these terms comprise H_I .

In contrast, the mean-field corrected Hamiltonian denoted H_{MF} is given by:

$$H_{MF}(|\psi\rangle) = H_0 + H_{I, MF}(|\psi\rangle) + \lambda V, \quad (C.5)$$

where the form of the Hamiltonian now depends on the state of the system due to the mean-field corrections:

$$H_{I, MF}(|\psi\rangle) = - \sum_{(i,j) \in I} J_{ij} \sigma_z^{(i)} \langle \psi | \sigma_z^{(j)} | \psi \rangle. \quad (C.6)$$

This state-dependence must be taken into account when applying perturbation theory.

C.1. Perturbing the full Hamiltonian

In this section, we follow the degenerate perturbation theory procedure outlined in [56]. When all interactions are included, the unperturbed Hamiltonian is given by the sum of H_0 and H_I . The eigendecomposition of this operator defines the zeroth order eigenenergies and eigenstates:

$$H_0 + H_I = \sum_k E_k^{(0)} |k^{(0)}\rangle \langle k^{(0)}|. \quad (C.7)$$

The unperturbed Hamiltonian is classical, with computational basis states for eigenstates. It also possesses \mathbb{Z}_2 symmetry, such that each eigenstate $|x\rangle$ and its ‘flipped’ version $|\bar{x}\rangle$ are degenerate, including the ground state $|x^*\rangle$. It is necessary to find the correct linear combinations of the degenerate states to properly calculate the higher order corrections. Typically, this is accomplished by diagonalizing the perturbing Hamiltonian V in the subspace of degenerate eigenstates, which yields the proper zeroth order eigenstates as well as the first order energy corrections, $\{E_k^{(1)}\}$. The particular V of this problem—a global transverse field—vanishes in the subspace of $|x\rangle$, $|\bar{x}\rangle$, requiring the first order energy corrections $\{E_k^{(1)}\}$ to vanish but providing no insight into the correct zeroth order eigenstates.

To calculate the first order eigenstate corrections, the correction is split into one within the degenerate space D_k and one in the space outside D_k , which we define as \bar{D}_k . For the latter correction, we have:

$$P_{\bar{D}_k} |k^{(1)}\rangle = \sum_{m \notin D_k} \frac{\langle m^{(0)} | V | k^{(0)} \rangle}{E_k^{(0)} - E_m^{(0)}} |m^{(0)}\rangle, \quad (C.8)$$

where $P_{\bar{D}_k}$ is the projects out the degenerate subspace. Following [56], the correction within the degenerate subspace is given by:

$$P_{D_k} |k^{(1)}\rangle = \sum_{\substack{k' \in D_k, \\ k' \neq k}} \frac{P_{D_k} |k'(0)\rangle}{E_k^{(1)} - E_{k'}^{(1)}} \langle k'(0) | V P_{\bar{D}_k} \frac{1}{E_k^{(0)} - H_0 - H_I} P_{\bar{D}_k} V | k^{(0)} \rangle. \quad (C.9)$$

The first order energy difference in the denominator of the first fraction of this expression is singular, because the degeneracy within D_k has not been lifted by the first order energy correction. However, this apparent issue provides us with the correct zeroth order eigenstates required to cancel the singular denominator: $\{|k^{(0)}\rangle\}$ must be selected to diagonalize the object $W := V P_{\bar{D}_k} (E_k^{(0)} - H_0 - H_I)^{-1} P_{\bar{D}_k} V$ within the subspace of degenerate states, such that the quantity $\langle k'(0) | W | k^{(0)} \rangle$ vanishes for all $k' \neq k$ within D_k . With this choice for $\{|k^{(0)}\rangle\}$, the first order eigenstate correction within the degenerate space vanishes, and the full first order eigenstate correction takes the form:

$$|k^{(1)}\rangle = \sum_{m \notin D_k} \frac{\langle m^{(0)} | V | k^{(0)} \rangle}{E_k^{(0)} - E_m^{(0)}} |m^{(0)}\rangle. \quad (C.10)$$

One can show that $\langle x | W | x \rangle = \langle x | W | \bar{x} \rangle = \langle \bar{x} | W | x \rangle = \langle \bar{x} | W | \bar{x} \rangle$ using the \mathbb{Z}_2 symmetry of the unperturbed Hamiltonian. Thus, the correct zeroth order eigenstates are the symmetric and antisymmetric superpositions of $|x\rangle$ and $|\bar{x}\rangle$, given by $|\pm_x\rangle = (1/\sqrt{2})(|x\rangle \pm |\bar{x}\rangle)$.

C.2. Perturbing the mean-field corrected Hamiltonian

Throughout this derivation, we will take advantage of the particular form of $H_{I, MF}(|\psi\rangle)$ to make simplifications.

Consider a modified Schrödinger equation that takes into account the state-dependence of $H_{MF}(|\psi\rangle)$:

$$(H_0 + H_{I, MF}(|k_{MF}\rangle) + \lambda V)|k_{MF}\rangle = E_{k, MF}|k_{MF}\rangle. \quad (\text{C.11})$$

In conventional perturbation theory, the eigenstates and eigenvalues ($|k_{MF}\rangle$ and $E_{k, MF}$, respectively) are expanded about their unperturbed counterparts, $|k_{MF}^{(0)}\rangle$ and $E_{k, MF}^{(0)}$. However, the modified Schrödinger equation written above is no longer an eigenvalue problem; the dependence on $|k_{MF}\rangle$ is non-linear. In the spirit of conventional perturbation theory, we will proceed in the usual manner, taking special care to include the non-linearity. Expanding $|k_{MF}\rangle$ and $E_{k, MF}$ in orders of λ :

$$|k_{MF}\rangle = |k_{MF}^{(0)}\rangle + \lambda|k_{MF}^{(1)}\rangle + \lambda^2|k_{MF}^{(2)}\rangle + \dots, \quad (\text{C.12})$$

$$E_{k, MF} = E_{k, MF}^{(0)} + \lambda E_{k, MF}^{(1)} + \lambda^2 E_{k, MF}^{(2)} + \dots. \quad (\text{C.13})$$

Replacing $|k_{MF}\rangle$ and $E_{k, MF}$ in the modified Schrödinger equation:

$$\begin{aligned} & \left(H_0 + H_{I, MF} \left(|k_{MF}^{(0)}\rangle + \lambda|k_{MF}^{(1)}\rangle + \lambda^2|k_{MF}^{(2)}\rangle + \dots \right) + \lambda V \right) \left(|k_{MF}^{(0)}\rangle + \lambda|k_{MF}^{(1)}\rangle + \lambda^2|k_{MF}^{(2)}\rangle + \dots \right) \\ &= \left(E_{k, MF}^{(0)} + \lambda E_{k, MF}^{(1)} + \lambda^2 E_{k, MF}^{(2)} + \dots \right) \left(|k_{MF}^{(0)}\rangle + \lambda|k_{MF}^{(1)}\rangle + \lambda^2|k_{MF}^{(2)}\rangle + \dots \right). \end{aligned}$$

Now, we can isolate and equate the various orders of λ to calculate the perturbations. If we first consider the zeroth order, we recover the unperturbed spectrum:

$$\left(H_0 + H_{I, MF} \left(|k_{MF}^{(0)}\rangle \right) \right) |k_{MF}^{(0)}\rangle = E_{k, MF}^{(0)}|k_{MF}^{(0)}\rangle. \quad (\text{C.14})$$

Moving to higher orders, it is helpful to write the explicit form of $H_{I, MF}(|\psi\rangle)$ to properly count the orders of λ . The first order terms take the form:

$$\begin{aligned} & H_0|k_{MF}^{(1)}\rangle + V|k_{MF}^{(0)}\rangle - \sum_{\langle i, j \rangle \in I} J_{ij} \left(\sigma_z^{(i)} \langle k_{MF}^{(0)} | \sigma_z^{(j)} | k_{MF}^{(0)} \rangle \right) |k_{MF}^{(1)}\rangle \\ & - \sum_{\langle i, j \rangle \in I} J_{ij} \left(\sigma_z^{(i)} \langle k_{MF}^{(1)} | \sigma_z^{(j)} | k_{MF}^{(0)} \rangle + \sigma_z^{(i)} \langle k_{MF}^{(0)} | \sigma_z^{(j)} | k_{MF}^{(1)} \rangle \right) |k_{MF}^{(0)}\rangle \\ &= E_{k, MF}^{(0)}|k_{MF}^{(1)}\rangle + E_{k, MF}^{(1)}|k_{MF}^{(0)}\rangle. \end{aligned} \quad (\text{C.15})$$

The operators $\sigma_z^{(j)}$ are diagonal in the computational basis $\{|k_{MF}^{(0)}\rangle\}$. By definition, corrections to $|k_{MF}^{(0)}\rangle$ such as $|k_{MF}^{(1)}\rangle$ will be orthogonal to $|k_{MF}^{(0)}\rangle$; therefore, terms of the form $\langle k_{MF}^{(1)} | \sigma_z^{(j)} | k_{MF}^{(0)} \rangle$ will vanish. The first order equation thus simplifies to:

$$H_0|k_{MF}^{(1)}\rangle + V|k_{MF}^{(0)}\rangle - \sum_{\langle i, j \rangle \in I} J_{ij} \left(\sigma_z^{(i)} \langle k_{MF}^{(0)} | \sigma_z^{(j)} | k_{MF}^{(0)} \rangle \right) |k_{MF}^{(1)}\rangle = E_{k, MF}^{(0)}|k_{MF}^{(1)}\rangle + E_{k, MF}^{(1)}|k_{MF}^{(0)}\rangle. \quad (\text{C.16})$$

To determine the first order energy shift, we project the above equation with the unperturbed eigenstate $\langle k_{MF}^{(0)} |$:

$$\begin{aligned} & \langle k_{MF}^{(0)} | H_0 | k_{MF}^{(1)} \rangle + \langle k_{MF}^{(0)} | V | k_{MF}^{(0)} \rangle - \sum_{\langle i, j \rangle \in I} J_{ij} \langle k_{MF}^{(0)} | \sigma_z^{(i)} | k_{MF}^{(1)} \rangle \langle k_{MF}^{(0)} | \sigma_z^{(j)} | k_{MF}^{(0)} \rangle \\ &= E_{k, MF}^{(0)} \langle k_{MF}^{(0)} | k_{MF}^{(1)} \rangle + E_{k, MF}^{(1)} \langle k_{MF}^{(0)} | k_{MF}^{(0)} \rangle. \end{aligned} \quad (\text{C.17})$$

The terms originating from $H_{I, MF}(|\psi\rangle)$ as well as the first term on the right-hand side of the equation will vanish due to the previously discussed fact that $|k_{MF}^{(1)}\rangle$ will be orthogonal to $|k_{MF}^{(0)}\rangle$. Let us examine the first term on the left-hand side, making use of the definition of the unperturbed spectrum:

$$\begin{aligned} \langle k_{MF}^{(0)} | H_0 | k_{MF}^{(1)} \rangle &= \langle k_{MF}^{(0)} | E_{k, MF}^{(0)} - H_{I, MF} \left(|k_{MF}^{(0)}\rangle \right) | k_{MF}^{(1)} \rangle \\ &= E_{k, MF}^{(0)} \langle k_{MF}^{(0)} | k_{MF}^{(1)} \rangle + \sum_{\langle i, j \rangle \in I} J_{ij} \langle k_{MF}^{(0)} | \sigma_z^{(i)} | k_{MF}^{(1)} \rangle \langle k_{MF}^{(0)} | \sigma_z^{(j)} | k_{MF}^{(0)} \rangle \\ &= 0. \end{aligned}$$

Thus, this term vanishes for the same reason, and the first order energy shift from conventional perturbation theory is recovered:

$$E_{k,\text{MF}}^{(1)} = \langle k_{\text{MF}}^{(0)} | V | k_{\text{MF}}^{(0)} \rangle. \quad (\text{C.18})$$

Of course, this energy correction vanishes for all k , in the same way that the first order energy shift of the full Hamiltonian vanishes.

To determine the first order eigenstate shift $|k_{\text{MF}}^{(1)}\rangle$, we project the first order equation given in equation (C.15) with the unperturbed eigenstate $\langle m_{\text{MF}}^{(0)} |$, where $m \neq k$:

$$\begin{aligned} \langle m_{\text{MF}}^{(0)} | H_0 | k_{\text{MF}}^{(1)} \rangle + \langle m_{\text{MF}}^{(0)} | V | k_{\text{MF}}^{(0)} \rangle - \sum_{\langle i,j \rangle \in I} J_{ij} \langle m_{\text{MF}}^{(0)} | \sigma_z^{(i)} | k_{\text{MF}}^{(1)} \rangle \langle k_{\text{MF}}^{(0)} | \sigma_z^{(j)} | k_{\text{MF}}^{(0)} \rangle \\ = E_{k,\text{MF}}^{(0)} \langle m_{\text{MF}}^{(0)} | k_{\text{MF}}^{(1)} \rangle + E_{k,\text{MF}}^{(1)} \langle m_{\text{MF}}^{(0)} | k_{\text{MF}}^{(0)} \rangle. \end{aligned} \quad (\text{C.19})$$

The only term guaranteed to vanish is the second term on the right-hand side, as $\langle m_{\text{MF}}^{(0)} | k_{\text{MF}}^{(0)} \rangle = 0$ for $m \neq k$. The term $\langle m_{\text{MF}}^{(0)} | H_0 | k_{\text{MF}}^{(1)} \rangle$ can also be simplified, with careful attention to the definition of the zeroth order energy $E_{m,\text{MF}}^{(0)}$:

$$\begin{aligned} \langle m_{\text{MF}}^{(0)} | H_0 | k_{\text{MF}}^{(1)} \rangle &= \langle m_{\text{MF}}^{(0)} | E_{m,\text{MF}}^{(0)} - H_{\text{I},\text{MF}}(|k_{\text{MF}}^0\rangle) | k_{\text{MF}}^{(1)} \rangle \\ &= E_{m,\text{MF}}^{(0)} \langle m_{\text{MF}}^{(0)} | k_{\text{MF}}^{(1)} \rangle + \sum_{\langle i,j \rangle \in I} J_{ij} \langle m_{\text{MF}}^{(0)} | \sigma_z^{(i)} | k_{\text{MF}}^{(1)} \rangle \langle k_{\text{MF}}^{(0)} | \sigma_z^{(j)} | k_{\text{MF}}^{(0)} \rangle. \end{aligned} \quad (\text{C.20})$$

Notice that the mean-field terms in the expression above precisely cancel with the mean-field terms remaining in equation (C.19). Simplifying:

$$\langle m_{\text{MF}}^{(0)} | V | k_{\text{MF}}^{(0)} \rangle = (E_{k,\text{MF}}^{(0)} - E_{m,\text{MF}}^{(0)}) \langle m^{(0)} | k^{(1)} \rangle. \quad (\text{C.21})$$

Therefore, the first order correction to the eigenstate is identical to that of the full Hamiltonian with no mean-field corrections:

$$\begin{aligned} |k_{\text{MF}}^{(1)}\rangle &= \sum_m |m_{\text{MF}}^{(0)}\rangle \langle m_{\text{MF}}^{(0)} | k_{\text{MF}}^{(1)} \rangle \\ &= \sum_{m \notin D_k} |m_{\text{MF}}^{(0)}\rangle \frac{\langle m_{\text{MF}}^{(0)} | V | k_{\text{MF}}^{(0)} \rangle}{E_{k,\text{MF}}^{(0)} - E_{m,\text{MF}}^{(0)}}. \end{aligned} \quad (\text{C.22})$$

C.3. First order eigenstate overlap

In this section, we derive the overlap between an eigenstate of an Ising-like Hamiltonian H with a small transverse field and the corresponding eigenstate of a mean-field corrected version of the Hamiltonian, H_{MF} , to leading order in perturbation theory. This overlap is approximately correct as long as the strength of the transverse field $|h|$ is small enough that first order perturbation theory is a good description of the eigenstates.

To first order, an eigenstate of each Hamiltonian is given by:

$$|\psi_k\rangle = \mathcal{N} \left(|k^{(0)}\rangle + |k^{(1)}\rangle \right), \quad (\text{C.23})$$

$$|\psi_{k,\text{MF}}\rangle = \mathcal{N}_{\text{MF}} \left(|k_{\text{MF}}^{(0)}\rangle + |k_{\text{MF}}^{(1)}\rangle \right), \quad (\text{C.24})$$

where \mathcal{N} , \mathcal{N}_{MF} are normalization factors. The zeroth order eigenstates are normalized by definition, so re-normalization is required when any higher order corrections are included.

The overlap between the two states is given by their inner product:

$$\langle \psi_k | \psi_{k,\text{MF}} \rangle = \mathcal{N}^* \mathcal{N}_{\text{MF}} \left(\langle k^{(0)} | k_{\text{MF}}^{(0)} \rangle + \langle k^{(1)} | k_{\text{MF}}^{(1)} \rangle \right), \quad (\text{C.25})$$

where the cross-terms $\langle k^{(1)} | k_{\text{MF}}^{(0)} \rangle$, $\langle k^{(0)} | k_{\text{MF}}^{(1)} \rangle$ have been dropped due to the fact that the first order corrections $|k^{(1)}\rangle$, $|k_{\text{MF}}^{(1)}\rangle$ lie outside the degenerate subspace of their corresponding zeroth order states $|k^{(0)}\rangle$ and $|k_{\text{MF}}^{(0)}\rangle$, which both live in the same degenerate subspace D_k .

As argued in previous sections, the zeroth order eigenstates $|k_{\text{MF}}^{(0)}\rangle$ are computational basis states $|x\rangle$, while the zeroth order eigenstates $|k^{(0)}\rangle$ are superpositions of two computational basis states related by \mathbb{Z}_2 symmetry, $|\pm_x\rangle = (1/\sqrt{2})(|x\rangle \pm |\bar{x}\rangle)$. Without loss of generality, we will assume $|k^{(0)}\rangle$ is the positive superposition $|+_x\rangle$, as is the case for the ground state given our conventions. The overlap between zeroth order eigenstates follows directly from these definitions: $\langle k^{(0)}|k_{\text{MF}}^{(0)}\rangle = \frac{1}{\sqrt{2}}$.

To find the overlap between first order corrections, we rewrite these corrections as a sum over degenerate subspaces D_m rather than over zeroth order eigenstates. This is possible because of the \mathbb{Z}_2 symmetry of each Hamiltonian, allowing the full set of eigenstates to be split into degenerate pairs.

First, we consider $|k^{(1)}\rangle$. The correction can be expressed abstractly as a weighted sum over states within each degenerate subspace D_m :

$$\begin{aligned} |k^{(1)}\rangle &= \sum_{D_m \neq D_k} c(m, k) |D_m\rangle \\ &= \sum_{D_y \neq D_x} c(y, x) |D_y\rangle. \end{aligned} \quad (\text{C.26})$$

In the second line, the labels are changed to explicitly highlight the relationship to computational basis states x, y . Using the previous definition in equation (21):

$$\begin{aligned} c(y, x) |D_y\rangle &= \frac{1}{\Delta E_{x,y}} (\langle +_y | V | +_x \rangle |+_y\rangle + \langle -_y | V | +_x \rangle |-_y\rangle) \\ &= \frac{1}{\Delta E_{x,y}} \left[\frac{1}{2} (V_{yx} + V_{y\bar{x}} + V_{\bar{y}x} + V_{\bar{y}\bar{x}}) |+_y\rangle + \frac{1}{2} (V_{yx} + V_{y\bar{x}} - V_{\bar{y}x} - V_{\bar{y}\bar{x}}) |-_y\rangle \right] \\ &= \frac{1}{\Delta E_{x,y}} (V_{yx} + V_{y\bar{x}}) |+_y\rangle, \end{aligned} \quad (\text{C.27})$$

where $\Delta E_{x,y}$ is the energy difference between D_x and D_y , and $V_{xy} = \langle x | V | y \rangle$ is an element of the matrix V in the computational basis. We have made use of properties of V (namely, $V_{yx} = V_{\bar{y}\bar{x}}$ and $V_{y\bar{x}} = V_{\bar{y}x}$) to simplify.

Rewriting $|k_{\text{MF}}^{(1)}\rangle$:

$$\begin{aligned} |k_{\text{MF}}^{(1)}\rangle &= \sum_{D_m \neq D_k} c_{\text{MF}}(m, k) |D_{m,\text{MF}}\rangle \\ &= \sum_{D_y \neq D_x} c_{\text{MF}}(y, x) |D_{y,\text{MF}}\rangle. \end{aligned} \quad (\text{C.28})$$

The differing zeroth order states leads to a different expression for the weighted sum:

$$\begin{aligned} c_{\text{MF}}(y, x) |D_{y,\text{MF}}\rangle &= \frac{1}{\Delta E_{x,y}} (\langle y | V | x \rangle |y\rangle + \langle \bar{y} | V | x \rangle |\bar{y}\rangle) \\ &= \frac{1}{\Delta E_{x,y}} (V_{yx} |y\rangle + V_{\bar{y}x} |\bar{y}\rangle). \end{aligned} \quad (\text{C.29})$$

The inner product between first order corrections is now straightforward:

$$\begin{aligned} \langle k^{(1)} | k_{\text{MF}}^{(1)} \rangle &= \sum_{D_y \neq D_x} c^*(y, x) c_{\text{MF}}(y, x) \langle D_y | D_{y,\text{MF}} \rangle \\ &= \sum_{D_y \neq D_x} \frac{1}{(\Delta E_{x,y})^2} \langle +_y | (V_{yx} + V_{y\bar{x}}) (V_{yx} |y\rangle + V_{\bar{y}x} |\bar{y}\rangle) \\ &= \frac{1}{\sqrt{2}} \sum_{D_y \neq D_x} \frac{1}{(\Delta E_{x,y})^2} (V_{yx} + V_{y\bar{x}})^2 \\ &= \frac{1}{\sqrt{2}} \langle k^{(1)} | k^{(1)} \rangle. \end{aligned} \quad (\text{C.30})$$

Finally, the overlap between eigenstates to first order is given by:

$$\begin{aligned} \langle \psi_k | \psi_{k,\text{MF}} \rangle &= \frac{1}{\sqrt{2}} \mathcal{N}^* \mathcal{N}_{\text{MF}} \left(\langle k^{(0)} | k^{(0)} \rangle + \langle k^{(1)} | k^{(1)} \rangle \right) \\ &= \frac{1}{\sqrt{2}} \frac{\mathcal{N}_{\text{MF}}}{\mathcal{N}}. \end{aligned} \quad (\text{C.31})$$

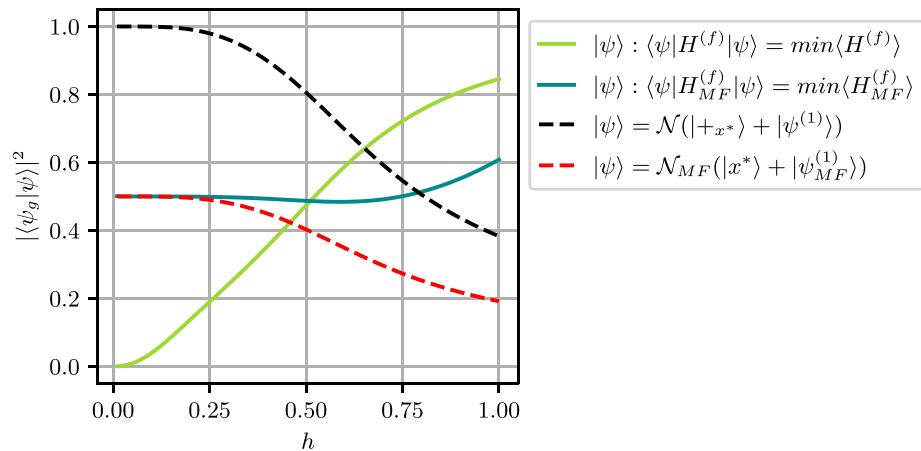


Figure C1. Ising-like model of $N = 6$ with all-to-all J_{ij} sampled from a Gaussian distribution centered at zero with width 1.0. Sweeping across transverse field strength h , the ground state $|\psi_g\rangle$ is computed using exact diagonalization of H . Fidelity is plotted between $|\psi_g\rangle$ and the minimum energy state of $H^{(f)}$ (light green), the minimum energy state of $H_{MF}^{(f)}$ (dark teal), ground state first order perturbation theory for H (black), and ground state first order perturbation theory for $H_{MF}^{(f)}$ (red).

Consider the form of the normalization coefficient \mathcal{N} and \mathcal{N}_{MF} :

$$\begin{aligned} \mathcal{N} &= \sqrt{\frac{1}{1 + \langle k^{(1)} | k^{(1)} \rangle}} \\ &= \sqrt{\frac{1}{1 + \sum_{D_y \neq D_x} (V_{yx} + V_{y\bar{x}})^2 / (\Delta E_{x,y})^2}} \end{aligned} \quad (C.32)$$

$$\begin{aligned} \mathcal{N}_{MF} &= \sqrt{\frac{1}{1 + \langle k_{MF}^{(1)} | k_{MF}^{(1)} \rangle}} \\ &= \sqrt{\frac{1}{1 + \sum_{D_y \neq D_x} (|V_{yx}|^2 + |V_{y\bar{x}}|^2) / (\Delta E_{x,y})^2}}. \end{aligned} \quad (C.33)$$

In the computational basis, the matrix elements of the perturbing transverse field V are real and all have the same sign (with the sign depending on the sign of h). Thus, any product of two matrix elements will be a positive number or zero. Likewise, the squared energy difference $(\Delta E_{x,y})^2$ is positive. The denominator of \mathcal{N} contains two additional factors from the cross-terms of $(V_{yx} + V_{y\bar{x}})^2$ that do not appear in \mathcal{N}_{MF} ; however, these cross-terms vanish as only V_{yx} or $V_{y\bar{x}}$ can be nonzero for the same y . It follows that $\mathcal{N}_{MF} = \mathcal{N}$, and:

$$\langle \psi_k | \psi_{k,MF} \rangle = \frac{1}{\sqrt{2}}. \quad (C.34)$$

These results are evaluated numerically for a small system in figure C1. Here, we consider $N = 6$ qubits split into two fragments with $N_f = 3$. The qubits interact all-to-all, with J_{ij} drawn from a Gaussian distribution centered at zero with width 1.0. Three Hamiltonians are considered: the full Hamiltonian H , the fragmented Hamiltonian $H^{(f)}$ which neglects all interactions crossing the fragment interface, and finally the mean-field corrected Hamiltonian $H_{MF}^{(f)}$, which replaces interface interactions by mean-field terms. Sweeping across the transverse field strength h , the minimum energy state is calculated and compared to the ground state of H , $|\psi_g\rangle$. To calculate the minimum energy state of $H_{MF}^{(f)}$ (which is state dependent and thus cannot be computed by solving a linear eigenvalue problem), we perform gradient descent directly on the state vector, minimizing the energy cost. Additionally, the ground states of H and $H_{MF}^{(f)}$ are approximately computed using first order perturbation theory.

As predicted, the minimum energy state of $H_{MF}^{(f)}$ has fidelity equal to 0.5 with $|\psi_g\rangle$ for small h . As h is increased beyond 0.25, the system enters a regime where first order perturbation theory is no longer sufficient to describe the state. Finally, we note that for small h , the minimum energy state of $H^{(f)}$ has negligible overlap with $|\psi_g\rangle$: when J_{ij} interactions are totally neglected, it is unlikely that the same $|x^*\rangle$ will minimize the energy.

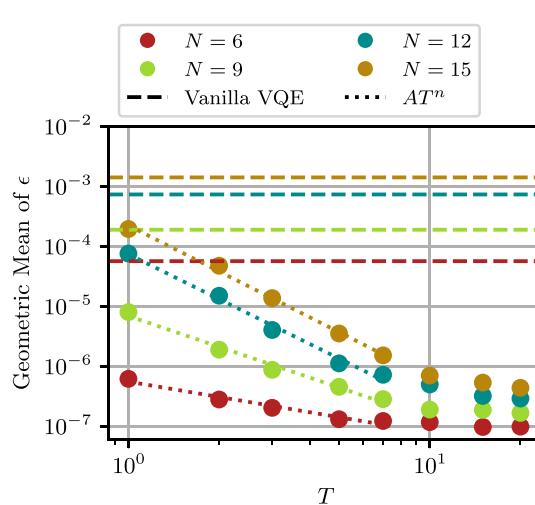


Figure D1. The geometric average of the final error as a function of batch size T . The results are presented for batch sizes T ranging from 1 to 20, with color indicating the number of qubits N in the full system. As a comparison, the results for vanilla VQE are plotted as dashed lines. The data for $T < 10$ is fit to a power law (with fitting parameters A and n), and the fit for each system size N given by a dotted line. Increasing T reduces the geometric mean of the final percent error, but even a single, fragmented pre-trained solution ($T = 1$) provides an order of magnitude reduction in averaged error or more.

Appendix D. Batched pre-training

We employ a batched approach to increase the probability of successful pre-training a PQC. To set up a single pre-training attempt, the full PQC is randomly split into fragments with at most $N_f = 3$ and two auxiliary registers, producing fragmented circuits of size $N_{f+a} = 5$ or smaller, which can comfortably be optimized using classical resources. To produce the pre-trained results presented in figures 8 and 9, a number $T = 10$ of such partitioned circuits are generated for the same problem Hamiltonian and optimized in parallel according to algorithm 1. After their initial optimization, the loss associated with each of the T sets of pre-trained parameters is estimated for the full circuit. This requires T additional loss measurements, a modest overhead when the value T is kept small relative to the required number of iterations. The set of pre-trained parameters producing the smallest loss is selected as the initial starting point for the full circuit optimization.

It is worth mentioning that for large problem Hamiltonians treated using quantum resources, the batched optimization of fragmented circuits can be done classically in parallel prior to using the quantum hardware. Only after pre-training would it be necessary to use the quantum hardware in order to estimate the loss of the pre-trained parameters. The role of batch size T in optimization success is explored numerically in appendix D.1.

D.1. The dependence of VQE performance on batch size

The role of batch size T is explored in figure D1. Here, we consider the geometric average of the final error after fragmented pre-training, contrasting the results to those of vanilla VQE (which utilizes a combination of random and near identity initialization for brickwork and all-to-all layers, respectively) plotted as dashed lines. On average, even pre-training a single set of fragmented circuits reduces the final error by an order of magnitude or more. The performance gain due to fragmented pre-training grows even larger as T is increased, and notably, the amount of error reduction grows with system size. These results indicate that increasing the batch size T will continue to reduce the error on average. In general, larger batch sizes are necessary to produce the same average error as the system size is increased. This can be understood by the fact that the number of possible ways to partition the system can increase with N , although exact scaling depends heavily on how the fragmentation is constrained. Empirically, the scaling of averaged final error is piece-wise. For small T , the scaling decays according to a power law; the data for $T < 10$ is fit to a power law and the resulting fit is provided by the dotted lines in figure D1. At roughly $T = 10$, the averaged error begins to plateau, and increasing T beyond 10 provides diminishing returns.

ORCID iDs

Abigail McClain Gomez  <https://orcid.org/0009-0002-0090-0941>
Taylor L Patti  <https://orcid.org/0000-0002-4242-6072>
Anima Anandkumar  <https://orcid.org/0000-0002-6974-6797>
Susanne F Yelin  <https://orcid.org/0000-0003-1655-9151>

References

- [1] DiVincenzo D P and Loss D 1999 Quantum computers and quantum coherence *J. Magn. Magn. Mater.* **200** 202–18
- [2] Denchev V S and Pandurangan G 2008 Distributed quantum computing: a new frontier in distributed systems or science fiction? *ACM SIGACT News* **39** 77–95
- [3] Gyongyosi L and Imre S 2019 A survey on quantum computing technology *Comput. Sci. Rev.* **31** 51–71
- [4] Birman K P 1993 The process group approach to reliable distributed computing *Commun. ACM* **36** 37–53
- [5] Attiya H and Welch J 2004 *Distributed Computing: Fundamentals, Simulations and Advanced Topics* (Wiley)
- [6] Kshemkalyani A, Singh M and Kshemkalyani A D 2008 *Distributed Computing: Principles, Algorithms and Systems* 1st edn (Cambridge University Press)
- [7] Hajibaba M and Gorgin S 2014 A review on modern distributed computing paradigms: cloud computing, jungle computing and fog computing *J. Comput. Inform. Technol.* **22** 69
- [8] Cheng B *et al* 2023 Noisy intermediate-scale quantum computers *Front. Phys.* **18** 21308
- [9] Cuomo D, Caleffi M and Sara Cacciapuoti A 2020 Towards a distributed quantum computing ecosystem *IET Quantum Commun.* **1** 3–8
- [10] Caleffi M, Sara Cacciapuoti A and Bianchi G 2018 Quantum internet: from communication to distributed computing! *Proc. 5th ACM Int. Conf. on Nanoscale Computing and Communication* pp 1–4
- [11] Sara Cacciapuoti A, Caleffi M, Tafuri F, Cataliotti F S, Gherardini S and Bianchi G 2020 Quantum internet: networking challenges in distributed quantum computing *IEEE Netw.* **34** 137–43
- [12] Wootters W K and Zurek W H 1982 A single quantum cannot be cloned *Nature* **299** 802–3
- [13] Xiao-Song M *et al* 2012 Quantum teleportation over 143 kilometres using active feed-forward *Nature* **489** 269–73
- [14] Van Meter R and Devitt S J 2016 The path to scalable distributed quantum computing *Computer* **49** 31–42
- [15] Qiao C, Zhao Y, Zhao G and Hongli X 2022 Quantum data networking for distributed quantum computing: opportunities and challenges *IEEE INFOCOM 2022—IEEE Conf. on Computer Communications Workshops (INFOCOM WKSHPS)* pp 1–6
- [16] Ferrari D, Sara Cacciapuoti A, Amoretti M and Caleffi M 2021 Compiler design for distributed quantum computing *IEEE Trans. Quantum Eng.* **2** 1–20
- [17] Ying M and Feng Y 2009 An algebraic language for distributed quantum computing *IEEE Trans. Comput.* **58** 728–43
- [18] Yimsiriwattana A and Lomonaco S J Jr 2004 Generalized GHZ states and distributed quantum computing (arXiv:quant-ph/0402148)
- [19] Liang Lim Y, Beige A and Chuan Kwek L 2005 Repeat-until-success linear optics distributed quantum computing *Phys. Rev. Lett.* **95** 030505
- [20] Liu X, Hu X-M, Zhu T-X, Zhang C, Xiao Y-X, Miao J-L, Zhong-Wen O, Liu B-H, Zhou Z-Q, Li C-F and Guo G-C 2023 Distributed quantum computing over 7.0 km (arXiv:2307.15634 [quant-ph])
- [21] Yimsiriwattana A and Samuel J L 2004 Distributed quantum computing: a distributed Shor algorithm *Quantum Information and Computation II (SPIE)* vol 5436, ed E Donkor, A R Pirich and H E Brandt (International Society for Optics and Photonics) pp 360–72
- [22] Zhang Z and Zhuang Q 2021 Distributed quantum sensing *Quantum Sci. Technol.* **6** 043001
- [23] Saleem Z H, Tomesh T, Perlin M A, Pranav G and Martin S 2022 Divide and conquer for combinatorial optimization and distributed quantum computation (arXiv:2107.07532 [quant-ph])
- [24] Parekh R, Ricciardi A, Darwish A and DiAdamo S 2021 Quantum algorithms and simulation for parallel and distributed quantum computing *2021 IEEE/ACM 2nd Int. Workshop on Quantum Computing Software (QCS)* pp 9–19
- [25] Fitzi M, Gisin N and Maurer U 2001 Quantum solution to the Byzantine agreement problem *Phys. Rev. Lett.* **87** 217901
- [26] Gavoille C, Kosowski A and Markiewicz M 2009 What can be observed locally? Round-based models for quantum distributed computing (arXiv:0903.1133 [quant-ph])
- [27] Avron J, Casper O and Rozen I 2021 Quantum advantage and noise reduction in distributed quantum computing *Phys. Rev. A* **104** 052404
- [28] Censor-Hillel K, Fischer O, Le Gall F, Leitersdorf D and Oshman R 2022 Quantum distributed algorithms for detection of cliques, 2022. (arXiv:2201.03000 [quant-ph])
- [29] Beals R, Brierley S, Gray O, Harrow A W, Kutin S, Linden N, Shepherd D and Stather M 2013 Efficient distributed quantum computing *Proc. R. Soc. A* **469** 20120686
- [30] Cirac J I, Ekert A K, Huelga S F and Macchiavello C 1999 Distributed quantum computation over noisy channels *Phys. Rev. A* **59** 4249–54
- [31] Van Meter R, Ladd T D, Fowler A G and Yamamoto Y 2010 Distributed quantum computation architecture using semiconductor nanophotonics *Int. J. Quantum Inform.* **8** 295–323
- [32] Gyongyosi L and Imre S 2021 Scalable distributed gate-model quantum computers *Sci. Rep.* **11** 5172
- [33] Sundaram R G, Gupta H and Ramakrishnan C R 2021 Efficient distribution of quantum circuits *35th Int. Symp. on Distributed Computing (DISC 2021) (Leibniz Int. Proc. in Informatics (LIPIcs)* vol 209) ed S Gilbert (Leibniz-Zentrum für Informatik) p 41
- [34] Cicconetti C, Conti M and Passarella A 2022 Resource allocation in quantum networks for distributed quantum computing *2022 IEEE Int. Conf. on Smart Computing (SMARTCOMP)* pp 124–32
- [35] Ngoenriang N, Minrui X, Supittayapornpong S, Niyato D, Han Y and Xuemin S 2022 Optimal stochastic resource allocation for distributed quantum computing (arXiv:2210.02886 [quant-ph])
- [36] Daei O, Navi K and Zomorodi-Moghadam M 2020 Optimized quantum circuit partitioning *Int. J. Theor. Phys.* **59** 3804–20
- [37] Houshmand M, Mohammadi Z, Zomorodi-Moghadam M and Houshmand M 2020 An evolutionary approach to optimizing teleportation cost in distributed quantum computation *Int. J. Theor. Phys.* **59** 1315–29

[38] Cuomo D, Caleffi M, Krsulich K, Tramonto F, Agliardi G, Prati E and Sara C A 2023 Optimized compiler for distributed quantum computing *ACM Trans. Quantum Comput.* **4** 1–29

[39] Eddins A, Motta M, Gujarati T P, Bravyi S, Mezzacapo A, Hadfield C and Sheldon S 2022 Doubling the size of quantum simulators by entanglement forging *PRX Quantum* **3** 010309

[40] Gentinetta G, Metz F and Carleo G 2024 Overhead-constrained circuit knitting for variational quantum dynamics *Quantum* **8** 1296

[41] Barratt F, Dborin J, Bal M, Stojevic V, Pollmann F and Green A G 2021 Parallel quantum simulation of large systems on small NISQ computers *npj Quantum Inf.* **7** 79

[42] Das P, Tannu S S, Nair P J and Qureshi M 2019 A case for multi-programming quantum computers *Proc. 52nd Annual IEEE/ACM Int. Symp. on Microarchitecture (ACM)* pp 291–303

[43] Park G, Zhang K, Kwangmin Y and Korepin V 2022 Quantum multi-programming for Grover’s search (arXiv:2207.14464 [quant-ph])

[44] Strecka J and Jascur M 2015 A brief account of the ising and ising-like models: Mean-field, effective-field and exact results (arXiv:1511.03031)

[45] McCaskey A, Dumitrescu E, Liakh D and Humble T 2018 Hybrid programming for near-term quantum computing systems 2018 *IEEE Int. Conf. on Rebooting Computing (ICRC)* (IEEE) pp 1–12

[46] Britt K A and Humble T S 2017 High-performance computing with quantum processing units *ACM J. Emerg. Technol. Comput. Syst.* **13** 39:1–39:13

[47] Fedorov D A, Peng B, Govind N and Alexeev Y 2022 VQE method: a short survey and recent developments *Mater. Theory* **6** 2

[48] Pfeuty P 1970 The one-dimensional ising model with a transverse field *Ann. Phys., NY* **57** 79–90

[49] Stinchcombe R B 1973 Ising model in a transverse field. I. Basic theory *J. Phys. C: Solid State Phys.* **6** 2459–83

[50] Noiri A, Takeda K, Nakajima T, Kobayashi T, Sammak A, Scappucci G and Tarucha S 2022 A shuttling-based two-qubit logic gate for linking distant silicon quantum processors *Nat. Commun.* **13** 5740

[51] Bennett C H, Brassard G, Crépeau C, Jozsa R, Peres A and Wootters W K 1993 Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels *Phys. Rev. Lett.* **70** 1895–9

[52] Bouwmeester D, Pan J-W, Mattle K, Eibl M, Weinfurter H and Zeilinger A 1997 Experimental quantum teleportation *Nature* **390** 575–9

[53] Parkinson J B and Farnell D J J 2010 *An Introduction to Quantum Spin Systems* vol 816 (Springer)

[54] Kim K *et al* 2011 Quantum simulation of the transverse Ising model with trapped ions *New J. Phys.* **13** 105003

[55] Schauss P 2018 Quantum simulation of transverse Ising models with Rydberg atoms *Quantum Sci. Technol.* **3** 023001

[56] Sakurai J J and Napolitano J 2020 *Modern Quantum Mechanics* 3rd edn (Cambridge University Press)

[57] Wouters S, Jiménez-Hoyos C A, Sun Q and Chan G K-L 2016 A practical guide to density matrix embedding theory in quantum chemistry *J. Chem. Theory Comput.* **12** 2706–19

[58] Heß B A, Marian C M, Wahlgren U and Gropen O 1996 A mean-field spin-orbit method applicable to correlated wavefunctions *Chem. Phys. Lett.* **251** 365–71

[59] Bobbio A, Gribaldo M and Telek M 2008 Analysis of large scale interacting systems by mean field method 2008 5th Int. Conf. on Quantitative Evaluation of Systems pp 215–24

[60] Sargent M, Scully M O and Lamb W E Jr 1978 *Laser Physics* (Avalon Publishing)

[61] Gilmore R 1974 Baker–Campbell–Hausdorff formulas *J. Math. Phys.* **15** 2090–2

[62] Sachdev S 1999 Quantum phase transitions *Phys. World* **12** 33

[63] Dziarmaga J 2005 Dynamics of a quantum phase transition: Exact solution of the quantum ising model *Phys. Rev. Lett.* **95** 245701

[64] Suzuki S, Inoue J-I and Chakrabarti B K 2012 *Quantum Ising Phases and Transitions in Transverse Ising Models* vol 862 (Springer)

[65] Bergholm V *et al* 2022 PennyLane: automatic differentiation of hybrid quantum-classical computations (arXiv:1811.04968 [physics, physics:quant-ph])

[66] Liang Y-C, Yeh Y-H, Mendonça P E M F, Yan Teh R, Reid M D and Drummond P D 2019 Quantum fidelity measures for mixed states *Rep. Prog. Phys.* **82** 076001

[67] Thiago R de O, Marcio F C and Felipe F F 2014 Monogamy of entanglement of formation *Phys. Rev. A* **89** 034303

[68] Lasse Buesen F L, Segal D and Khait I 2023 Simulating time evolution on distributed quantum computers *Phys. Rev. Res.* **5** L022003

[69] Cerezo M *et al* 2021 Variational quantum algorithms *Nat. Rev. Phys.* **3** 625–44

[70] Chinzei K, Hoan Tran Q, Maruyama K, Oshima H and Sato S 2024 Splitting and parallelizing of quantum convolutional neural networks for learning translationally symmetric data *Phys. Rev. Res.* **6** 023042

[71] Jinyang Li, Wang Z, Zhirui H, Date P, Li A and Jiang W 2023 A novel spatial-temporal variational quantum circuit to enable deep learning on nisq devices (arXiv:2307.09771)

[72] Peng T, Harrow A W, Ozols M and Xiaodi W 2020 Simulating large quantum circuits on a small quantum computer *Phys. Rev. Lett.* **125** 150504

[73] Grant E, Wossnig L, Ostaszewski M and Benedetti M 2019 An initialization strategy for addressing barren plateaus in parametrized quantum circuits *Quantum* **3** 214

[74] Egger D J, Mareček J and Woerner S 2021 Warm-starting quantum optimization *Quantum* **5** 479

[75] Patti T L, Shehab O, Najafi K and Yelin S F 2022 Markov chain monte carlo enhanced variational quantum algorithms *Quantum Sci. Technol.* **8** 015019

[76] Jun Q and Tejedor J 2022 Classical-to-quantum transfer learning for spoken command recognition based on quantum neural networks *ICASSP 2022—2022 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 8627–31

[77] Dborin J, Barratt F, Wimalaweera V, Wright L and Green A G 2022 Matrix product state pre-training for quantum machine learning *Quantum Sci. Technol.* **7** 035014

[78] Qiuchi Li, Wang B, Zhu Y, Lioma C and Liu Q 2023 Adapting pre-trained language models for quantum natural language processing (arXiv:2302.13812 [quant-ph])

[79] Niu Y-F, Zhang S and Bao W-S 2023 Warm starting variational quantum algorithms with near clifford circuits *Electronics* **12** 347

[80] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O’Brien J L 2014 A variational eigenvalue solver on a photonic quantum processor *Nat. Commun.* **5** 4213

[81] McCaskey A J, Parks Z P, Jakowski J, Moore S V, Morris T D, Humble T S and Pooser R C 2019 Quantum chemistry as a benchmark for near-term quantum computers *npj Quantum Inf.* **5** 1–8

[82] Feniou C, Hassan M, Traoré D, Giner E, Maday Y and Piquemal J-P 2023 Overlap-ADAPT-VQE: practical quantum chemistry on quantum computers via overlap-guided compact Ansätze *Commun. Phys.* **6** 1–11

[83] Méle A A, Mbeng G B, Santoro G E, Collura M and Torta P 2022 Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz *Phys. Rev. A* **106** L060401

[84] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 Barren plateaus in quantum neural network training landscapes *Nat. Commun.* **9** 4812

[85] Holmes Z, Arrasmith A, Yan B, Coles P J, Albrecht A and Sornborger A T 2021 Barren plateaus preclude learning scramblers *Phys. Rev. Lett.* **126** 190501

[86] Ortiz Marrero C, Kieferová M and Wiebe N 2021 Entanglement-induced barren plateaus *PRX Quantum* **2** 040316

[87] Patti T L, Najafi K, Gao X and Yelin S F 2021 Entanglement devised barren plateau mitigation *Phys. Rev. Res.* **3** 033090

[88] Tüysüz C, Clemente G, Crippa A, Hartung T, Kühn S and Jansen K 2023 Classical splitting of parametrized quantum circuits *Quantum Machine Intell.* **5** 34

[89] Opper M and Winther O 1996 Mean field approach to bayes learning in feed-forward neural networks *Phys. Rev. Lett.* **76** 1964–7

[90] Pedro A H-S, Winther O and Kai Hansen L 2002 Mean-field approaches to independent component analysis *Neural Comput.* **14** 889–918

[91] Gao J B, Gunn S R and Harris C J 2003 Mean field method for the support vector machine regression *Neurocomputing* **50** 391–405

[92] Patti T L, Kossaifi J, Anandkumar A and Yelin S F 2022 Variational quantum optimization with multibasis encodings *Phys. Rev. Res.* **4** 033142

[93] Benedetti M, Lloyd E, Sack S and Fiorentini M 2019 Parameterized quantum circuits as machine learning models *Quantum Sci. Technol.* **4** 043001

[94] Yuxuan D, Hsieh M-H, Liu T and Tao D 2020 Expressive power of parametrized quantum circuits *Phys. Rev. Res.* **2** 033125

[95] Liu X, Angone A, Shaydulin R, Safro I, Alexeev Y and Cincio L 2022 Layer vqe: a variational approach for combinatorial optimization on noisy quantum computers *IEEE Trans. Quantum Eng.* **3** 1–20

[96] Zhao J, Schäfer F and Anandkumar A 2022 Zero initialization: initializing neural networks with only zeros and ones (arXiv:2110.12661 [cs])

[97] Zhang S, Qin Z, Zhou Y, Li R, Chunxiao D and Xiao Z 2024 Single entanglement connection architecture between multi-layer bipartite HEA (arXiv:2307.12323)

[98] Poljak S and Tuza Z 1995 Maximum cuts and large bipartite subgraphs *Combinatorial Optimization (DIMACS Series in Discrete Mathematics and Theoretical Computer Science* vol 20 (AMS)) pp 181–244

[99] Kingma D P and Jimmy B 2017 Adam: a method for stochastic optimization (arXiv:1412.6980 [cs])

[100] Paszke A *et al* 2019 Pytorch: an imperative style, high-performance deep learning library *CoRR* (arXiv:1912.01703)

[101] Stein S, Wiebe N, Ding Y, Peng B, Kowalski K, Baker N, Ang J and Li A 2022 EQC: ensembled quantum computing for variational quantum algorithms *Proc. 49th Annual Int. Symp. on Computer Architecture* (ACM) pp 59–71

[102] Khait I, Tham E, Segal D and Brodutch A 2023 Variational quantum eigensolvers in the era of distributed quantum computers (arXiv:2302.14067 [cond-mat, physics:quant-ph])

[103] Adam, Kim M S, Pollmann F and Knolle J 2019 Simulating quantum many-body dynamics on a current digital quantum computer *npj Quantum Inf.* **5** 1–13

[104] Kassal I, Jordan S P, Love P J, Mohseni M and Aspuru-Guzik A 2008 Polynomial-time quantum algorithm for the simulation of chemical dynamics *Proc. Natl Acad. Sci.* **105** 18681–6

[105] Pyrkov A, Aliper A, Bezrukov D, Lin Y-C, Polykovskiy D, Kamya P, Ren F and Zhavoronkov A 2023 Quantum computing for near-term applications in generative chemistry and drug discovery *Drug Discov. Today* **28** 103675

[106] Salehi Ozlem, Glos A and Adam Miszczak J 2022 Unconstrained binary models of the travelling salesman problem variants for quantum optimization *Quantum Inf. Process.* **21** 67

[107] Miranowicz A and Grudka A 2004 Ordering two-qubit states with concurrence and negativity *Phys. Rev. A* **70** 032326