



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Variational preparation of the thermofield double state of the Sachdev-Ye-Kitaev model

Vincent Paul Su

Phys. Rev. A **104**, 012427 — Published 29 July 2021

DOI: [10.1103/PhysRevA.104.012427](https://doi.org/10.1103/PhysRevA.104.012427)

Variational Preparation of the Sachdev-Ye-Kitaev Thermofield Double

Vincent Paul Su

*Center for Theoretical Physics & Department of Physics,
University of California, Berkeley, CA 94720, USA*

vipasu@berkeley.edu

ABSTRACT: We provide an algorithm for preparing the thermofield double (TFD) state of the Sachdev-Ye-Kitaev model without the need for an auxiliary bath. Following previous work, the TFD can be cast as the approximate ground state of a Hamiltonian, H_{TFD} . Using variational quantum circuits, we propose and implement a gradient-based algorithm for learning parameters that find this ground state, an application of the variational quantum eigensolver. Concretely, we find shallow quantum circuits that prepare the ground state of H_{TFD} for the $q = 4$ SYK model for $N = 8$ Majoranas per side. For $N = 12$, we achieve a variational energy within a percent of the true ground state energy.

1 Introduction

The thermofield double (TFD) is a privileged state in the Anti de-Sitter/Conformal Field Theory (AdS/CFT) correspondence [1], which relates a putative quantum gravity theory in a $D + 1$ dimensional Anti de-Sitter space to a conformal field theory living on the boundary with dimension D . Black holes emit thermal radiation [2], effectively leaving a thermal density matrix on the exterior. It was pointed out by Israel [3] that one could reproduce calculation of observables by considering the thermofield double, analogous to the maximal extension of a Schwarzschild geometry. Later, Maldacena [4] conjectured within the context of AdS/CFT that the TFD of the boundary CFT should correspond to eternal two sided black holes in AdS. The idea that there exists a duality between theories living in dimensions that differ by one is more generally known as holography.

To test this duality, it is interesting to consider the phenomena of traversable wormholes, a striking prediction of AdS/CFT. From a gravity perspective, it is clear that boundaries on opposite sides of the black hole cannot causally communicate with each other. Though there is a spatial wormhole which connects the two exterior regions, one cannot traverse it without falling into the black hole singularity. If Alice and Bob are on opposite sides, they cannot meet unless they jump into the black hole together. A recent development by Gao, Jafferis, and Wall [22] show that a particular coupling on the two boundary theories sources a negative energy shock that renders the wormhole in the TFD state traversable. In other words, Bob could reunite with Alice without being sucked in to the black hole. As a starting point for this protocol, and many other thought experiments in AdS/CFT, one presumes access to the TFD state.

One promising quantum mechanical system for probing AdS/CFT is the Sachdev-Ye-Kitaev (SYK) model [5, 6]. For example, it exhibits conformal symmetry at low energies, with dynamics governed by a Schwarzian action [7]. The same action governs a two dimensional theory of quantum gravity known as Jackiw-Teitelboim gravity [8, 9]. Additionally it has been shown to saturate the chaos bound at low temperatures, a sign of maximal scrambling also shared by black holes [10, 11].

In Ref. [12], the authors construct eternal traversable wormhole solutions in nearly-AdS₂ and show that the low energy limit of two coupled SYK models has an identical action. One key result is that they show the TFD for the SYK model can be well approximated by the ground state of the two sided hamiltonian with a small interaction.

In this work, we consider the task of state preparation of the TFD of the SYK model on Noisy Intermediate-Scale Quantum (NISQ) [13] devices. The more general task of preparing the TFD of arbitrary theories was considered in Ref. [14]. Again, the strategy was to construct a Hamiltonian whose ground state encodes the TFD structure. Though the Hamiltonian in Eq. (3.21) of Ref. [12] can be thought of as a slightly specialized version of the construction in Ref. [14], we will use it in this work for its relative simplicity. Both approaches consider the use of an auxiliary bath to adiabatically cool the system to its ground state. Here, we instead take a variational approach, starting with a quantum circuit ansatz whose parameters are tunable. This obviates the need for an auxiliary system.

A similar approach was used to construct the TFD of the Ising model [15]. Shortly

after, the TFD of the critical transverse field Ising model (TFIM) on 6 qubits was implemented on ion traps [16]. Our approach will have two significant features for holography. One necessary ingredient for AdS/CFT is a large- N (or large central charge) expansion. For the TFIM, larger system sizes simply realize the continuum limit of a $c = 1/2$ theory, whereas the large N limit of the SYK model has a gravitational dual. Second, our variational approach does not rely on classical computers to perform optimization of the quantum circuit, providing a path to a scalable procedure for larger qubit systems.

In this paper, we will use variational quantum circuits to prepare the TFD of the SYK model. In Section 2, we review the physics of the thermofield double of the SYK model as presented in Ref. [12]. The TFD preparation approximately boils down to a ground state problem of a modified Hamiltonian, with the approximation becoming exact in a large N limit. In Section 3, we provide an overview of variational circuits and how to optimize them. In Section 4, we share the results of our variational procedure applied to TFD preparation, making use of the excellent quantum simulation package `Yao.jl` [17]. We successfully prepare the approximate TFD on the $N = 8, q = 4$ variant of SYK models, even in the presence of shot noise, uncertainty that arises from estimating expectation values with a finite number of sample. For $N = 12, q = 4$, we find a state whose energy is within $\sim .2\%$ of the true ground state energy. We conclude with some remarks in Section 5.

2 The Thermofield Double of SYK

The thermofield double (TFD) at inverse temperature β is defined for general quantum systems by entangling two copies of a quantum system in the following way.

$$|\text{TFD}(\beta)\rangle = \frac{1}{Z_\beta^{1/2}} \sum_n e^{-\beta E_n/2} |n\rangle_L |\bar{n}\rangle_R \quad (2.1)$$

where n labels the energy eigenstates of a Hamiltonian that defines one copy of the system and the bar denotes CPT conjugation. Z_β is a normalization constant. The subscripts L, R correspond to the “left” and “right” systems from the context of two sided black holes that appear in Ref. [4]. Upon tracing out either the left or right system, we are left with a thermal state on the remaining system. The TFD can be thought of as a particular purification of a thermal state on a single system.

In the following section, we will motivate the thermofield double of a particular quantum mechanical system, the SYK model. In Section 2.1, we discuss briefly the SYK model and some of the physics related to traversable wormholes. In Section 2.2, we review recent literature on the preparation of the TFD, with specific application to the SYK model.

2.1 The SYK Model and its Gravity Dual

The SYK model at its core involves N Majorana fermions with all-to-all q -body interactions. Here we consider $q = 4$ for concreteness. The Hamiltonian can be written simply as follows

$$H_{\text{SYK}} = \sum_{ijkl} \mathcal{J}_{ijkl} \psi_i \psi_j \psi_k \psi_l, \quad (2.2)$$

$$\mathcal{J}_{ijkl} \sim \mathcal{N}\left(0, \frac{12J^2}{N^3}\right), \quad (2.3)$$

where the strength of the couplings of \mathcal{J}_{ijkl} are i.i.d. Gaussian. They are distributed normally about 0 with a width set by a parameter J . Here and below we work in units of $J = 1$.

The SYK model was first proposed as a quantum model of holography by Kitaev [6] with dynamics that had an analytic solution in the large N and large βJ limit. More recently, its connection to gravity has been studied in the duality to nearly AdS_2 spacetimes. At large N , the SYK model is typically solved with a saddle point approximation. The path integral is usually rewritten in terms of the fermion bilinear $G(\tau_1, \tau_2) = \frac{1}{N} \sum_j \langle \psi^j(\tau_1) \psi^j(\tau_2) \rangle$ and its self-energy Σ with a disorder averaging over the coupling strengths. This leads to the following Euclidean action.

$$S_E = -N \left\{ \log \text{Pf}(\partial_\tau - \Sigma) - \frac{1}{2} \int d\tau_1 d\tau_2 \left[\Sigma(\tau_1, \tau_2) G(\tau_1, \tau_2) - \frac{J^2}{q} G(\tau_1, \tau_2) \right] \right\} \quad (2.4)$$

The equations of motion obtained from this action are also referred to as the Schwinger-Dyson equations. At low temperatures, there is an emergent reparameterization symmetry of $\tau \rightarrow f(\tau)$ which is explicitly broken, leading to a Schwarzian action [18–20]

$$S \sim \int d\tau \{f(\tau), \tau\} = \int d\tau \left[\frac{f'''}{f'} - \frac{3}{2} \left(\frac{f''}{f'} \right)^2 \right], \quad (2.5)$$

where primes are derivatives with respect to τ .

An analogous phenomenon occurs in Jackiw-Teitelboim gravity [8, 9], a theory of quantum gravity in two dimensions. After using the equations of motion, one is left with a boundary degree of freedom describing a cutout of AdS space. This field also enjoys an emergent reparameterization symmetry which is broken by the physical $SL(2, \mathbb{R})$ symmetries of AdS [21]. In fact, the Schwarzian is the lowest order expression in derivatives that is local and invariant under $SL(2, \mathbb{R})$.

Given this duality between the SYK model and JT gravity, it is natural to consider the TFD state of the SYK model, involving two separate copies of the theory, and ask whether this corresponds to a wormhole on the gravity side. Indeed in Ref. [12], this question was studied in great detail. By considering interactions of the form in Ref. [22], they show that the low energy actions are equivalent for both JT and the SYK model. On the SYK side, they show that the ground state of the coupled model is well approximated by the TFD state. For more discussion on the overlap, see Section 4.4 of [12]. On the gravity side, they find AdS_2 solutions whose boundaries remain in causal contact for all times. A more direct test of wormhole traversability in the SYK model was analyzed in Ref. [23] closer to the original protocol proposed in [22]. They calculate the two-sided

correlation functions, characterizing the signature of teleportation and comparing with the gravitational expectation. As a starting point, however, they require the TFD state, whose preparation is the aim of this work.

2.2 The TFD as a Ground State

Given the special role that the TFD plays in quantum systems, one question that arises is how to actually prepare such a state. The general task was considered in Ref. [14]. The idea is to construct operators that annihilate the TFD state in Eq. (2.1) and then take the Hamiltonian to be the occupation number of those modes. For example, it is straightforward to show that the following operator annihilates the TFD state.

$$d \equiv e^{-\beta(H_L^0 + H_R^0)/4} (\mathcal{O}_L - \Theta \mathcal{O}_R \Theta^{-1}) e^{\beta(H_L^0 + H_R^0)/4}, \quad (2.6)$$

where $H_{L/R}^0$ are the Hamiltonian of the individual left/right systems and Θ is the CPT operator.

Intuitively, these are operators which measure left-right correlation for different operators. However, they are taken in the Heisenberg picture with evolution in imaginary time, making their explicit construction difficult. Motivated by holography, the authors of [14] consider the limit where the TFD Hamiltonian should be dominated by the Hamiltonians of the original systems. Schematically,

$$H_{\text{simple}} = H_L^0 + H_R^0 + \sum_k c_k (\mathcal{O}_{L,k}^\dagger - \mathcal{O}_{R,k}^*) (\mathcal{O}_{L,k} - \mathcal{O}_{R,k}^T) \quad (2.7)$$

In short, one takes the Hamiltonians of the left and right systems with an interaction term to favor the correct entanglement pattern. One may view this interaction as a Lagrange multiplier. Note that the only operators in this simplified Hamiltonian that involve both the left and right systems are local two point operators.

More specifically for the SYK model, Ref. [12] considered a very similar Hamiltonian to prepare the TFD.

$$H_{\text{TFD}} = H_{L,\text{SYK}} + H_{R,\text{SYK}} + H_{\text{int}}, \quad (2.8)$$

$$H_{\text{int}} = i\mu \sum_j \psi_L^j \psi_R^j, \quad (2.9)$$

where the factor of i ensures the hermiticity. In the large $N, \beta J$ limit with small μ , they show that the ground state approaches the TFD. The strength of μ also determines the effective “temperature”¹ of the resulting TFD state, which scales as $T \sim \mu^{2/3}$ [12].

The analysis for small μ primarily follows from the low energy limit of the SYK model which develops a nearly conformal symmetry, in which case the interaction term can be viewed as a relevant deformation. Its contribution to the action can then be replaced by conformal two point correlators as in Eq. (3.22) of [12]. This is similar to the assumption made in the “simple” Hamiltonian above, in which the coupling term is small relative to

¹This refers to the temperature of the thermal density matrix of one side. The two-sided TFD state is pure.

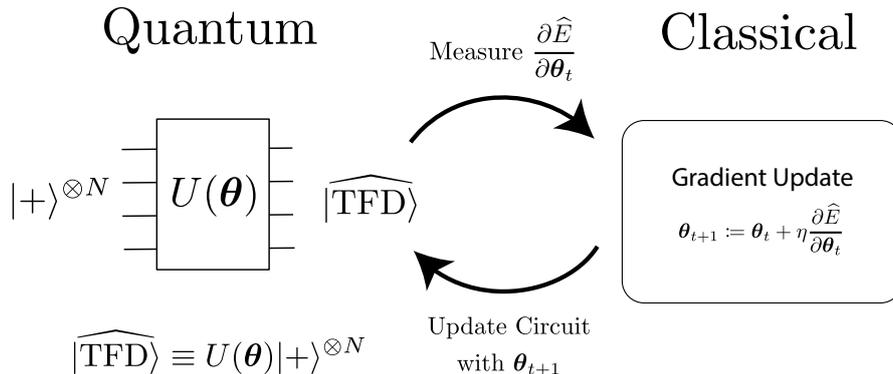


Figure 1: A schematic representation of hybrid quantum classical computation. Here we would like to learn the best parameters for our quantum circuit. Measurements from the circuit inform how the classical computer produces new parameters to modify the quantum circuit. An extremely important aspect of our procedure is that the gradient information is obtained directly from measurements of the quantum circuit, circumventing the need for calculating gradients of matrices of exponential size on a classical computer.

the Hamiltonian of the individual systems. As μ tends to 0, the conformal approximation becomes better, and the fidelity of the ground state of H_{TFD} with the TFD state becomes 1. Outside of this conformal approximation, when μ becomes too large, the fidelity decays exponentially with N [12].

In Ref. [24], the authors consider preparing the ground state of the coupled SYK models by cooling with a bath. There they distinguish between the TFD of the two decoupled systems and the ground state of H_{TFD} by referring to the latter as the “SYK wormhole”. In this work, we make the simplification that these are the same, i.e. that we are in a classical regime. Although current systems are far from having a large N number of qubits, our procedure could be repeated as the quantum hardware advances. We will leave investigation of quantum corrections to future direction and work under the assumption that we wish to prepare the ground state of H_{TFD} .

To summarize this section, we have reduced the problem of preparing the SYK thermofield double to a problem of finding the ground state of a particular Hamiltonian, H_{TFD} . In contrast to previous work that suggest the use of an external bath to cool the system, we will take a variational approach without any additional degrees of freedom.

3 Preparing the TFD on a Near Term Quantum Device

Tasked with finding the ground state of H_{TFD} , we turn to introducing the necessary ingredients for preparing it on a near term device using variational circuits. We follow the hybrid classical-quantum model where the quantum device is augmented by a classical computer that keeps track of the variational parameters of the circuit. For an overview, see Figure 1.

We emphasize that the role of the classical computer is extremely minimal when using the parameter shift rule to compute gradients.

In the context of NISQ devices, variational algorithms are useful for several reasons. First, quantum algorithm design is extremely difficult. Piecing together quantum gates to make a meaningful computation is as unintuitive as writing a calculator with access to only AND and OR gates. Second, circuit depth is a critical limiting factor in the NISQ era. Gate compilation, or translation between universal gate sets, leads to prohibitively large overhead in terms of gate count. Variational algorithms with gates that are native to the architecture keep the circuit depth minimal.

With these considerations, we use gradient descent to find the ground state of H_{TFD} on a hardware efficient ansatz [25]. The use of variational circuits to find ground states is often referred to as the Variational Quantum Eigensolver (VQE), originally applied in the context of quantum chemistry [26]. However, for the systems of application, the optimization has typically been carried out by classical methods, often involving storing the circuit in memory. Though this work is fully numerical, the proposed scheme on a quantum device would not need any classical optimization.

In Section 3.1, we discuss the Jordan Wigner transformation to encode H_{TFD} , expressed in terms of Majoranas, into Pauli operators acting on qubits. We then discuss in Section 3.2 our choice of *parameterized quantum circuits* (PQCs), circuits whose sequence of gates are fixed but have tunable parameters, such as angles of a Pauli rotation. Finally, in Section 3.3, we review how the parameter shift rule can be used to optimize our PQC with limited classical computation.

3.1 Jordan-Wigner Transformation

By far, qubits are the dominant computing paradigm in NISQ devices. In order to run our algorithm, we will need to express our Hamiltonian in terms of qubits rather than Majorana fermions. We review the Jordan-Wigner transformation that precisely enables this.²

The Jordan-Wigner transformation is a duality mapping between spinless fermionic operators to non-local Pauli operators. For Majorana fermions, which are their own antiparticles, this mapping can be written as

$$\psi_l = \left(\prod_{j=1}^{\tilde{l}-1} \sigma_j^z \right) \sigma_l^{\alpha_l} \quad (3.1)$$

where $\tilde{l} = \lfloor (l+1)/2 \rfloor$ and α_l denotes x if l is even and y otherwise. In short, each of the Majorana operators gets mapped to a Pauli string. Products appearing in the SYK Hamiltonian likewise get mapped to products of these Pauli strings.

Equipped with this redefinition, we can express H_{TFD} in terms of qubits, which we temporarily refer to as H_{qubit} .³ Since H_{TFD} is a sum over products of Majorana operators, H_{qubit} will be a sum over Pauli operators. Since the remainder of the paper will deal with

²See Ref. [27] for an excellent review.

³The OpenFermion package [28] was used to automate this conversion.

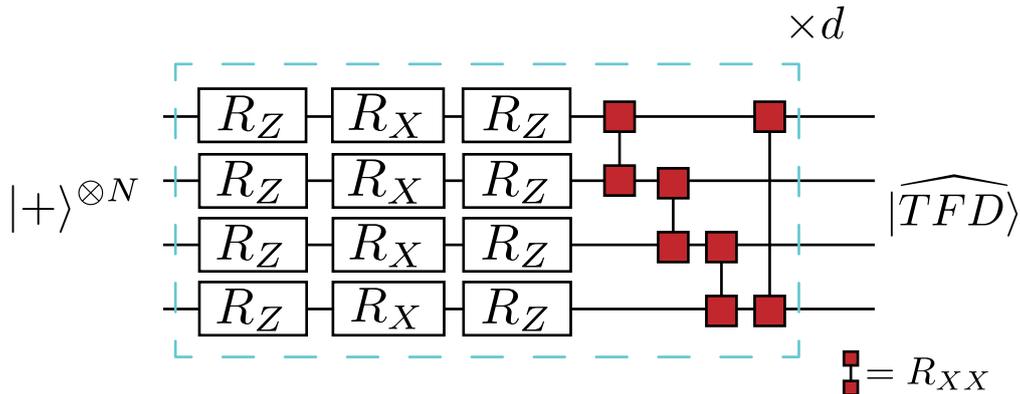


Figure 2: A simple parameterized quantum circuit (PQC) on $N = 4$ qubits composed of single qubit rotations and nearest neighbor two qubit gates. Each of the rotations is described by a continuous parameter θ . On the diagram, $R_X \equiv \exp(-\frac{i}{2}\theta X)$. Parameters are not necessarily shared between gates. This sequence of gates is repeated d times, indicating the depth of the circuit. The PQC acts on a fixed input state and produces our estimate of the TFD state. We can thus switch between viewing the variational object as either the circuit $U(\boldsymbol{\theta})$ or the state as in Eq. (3.3). This is the form of ansatz we consider in the Section 4. One can also replace the R_{XX} gate with other 2 qubit entanglers and get similar results.

qubits, we can drop this distinction. Similarly H_L will refer to the SYK Hamiltonian after applying the Jordan-Wigner transformation.

The number of qubits required is N . In terms of our original system, N was the number of Majorana fermions on a single copy of the theory. In the thermofield double, defined on two copies of the theory, we have a total of $2N$ Majoranas. However, the Hilbert space dimension of each Majorana fermion is $2^{1/2}$. For a total of $2N$ Majorana fermions, this is equivalent to N qubits. Thus the parameter N denotes the number of Majoranas on a single “side” of the TFD as well as the number of qubits in our discussion.

3.2 Parameterized Quantum Circuits

In short, parameterized quantum circuits are a variational ansatz for a quantum circuit, generally with a fixed gate layout. The variational parameters will typically correspond to rotation angles for a subset of gates. For example, on a single qubit, one can define rotation about the z axis by an angle θ as

$$R_Z(\theta) \equiv e^{-\frac{i}{2}\theta Z} \quad (3.2)$$

where Z is the Pauli Z matrix. This definition can also be extended to Pauli operators which act on multiple qubits by simply replacing σ_i in the exponential by the appropriate Pauli string.

For a given input state, one may then view this PQC ansatz as describing a class of states. Like any variational algorithm, we know that the energy of the state produced in this manner is lower bounded by the true ground state energy.

Our choice of PQC will be alternating layers of arbitrary single qubit rotations followed by two qubit entanglers. This is then repeated d times indicating the depth of the circuit. Motivated by experiment, we take our entangling gate to be the R_{XX} gate, though this could be swapped depending on available hardware.

Phrased in this way, both our circuit and, consequently, the resulting state on N qubits are parameterized by the angles $\boldsymbol{\theta}$

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|+\rangle^{\otimes N}, \quad (3.3)$$

where $U(\boldsymbol{\theta})|0\rangle$ describes our circuit acting on a fixed input state in the computational basis. The energy would then simply be the expectation value of the Hamiltonian in this state.

$$\widehat{E}(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle. \quad (3.4)$$

In the next section we will describe the procedure for how to optimize the parameters with respect to the energy.

3.3 Gradients from the Parameter Shift Rule

To find the ground state, we aim to minimize the variational energy and compare to results obtained by exact diagonalization. The minimization happens by taking gradients with respect to the expectation value of the energy in the state $|\psi(\boldsymbol{\theta})\rangle$.⁴

Intuitively, one could estimate gradients by perturbing the angles of the variational circuit by an amount ϵ and recording whether the energy increased or decreased. This is often referred to as the method of finite differences. However, this approximation works best as ϵ tends to 0. For near term devices where noise is an issue, this method is in tension with getting reliable gradient estimates.

Instead, we make use of the parameter shift rule [29, 30]. Procedurally, one still varies each of the angles. However, the shifts in angles are finite rather than infinitesimal. For example, take the PQC described in the previous section⁵. Let us assume that the total number of gate parameters is K . The i -th entry of the gradient would be given by the following expression.

$$\frac{\partial \widehat{E}(\theta_1, \dots, \theta_i, \dots, \theta_K)}{\partial \theta_i} = \frac{1}{2} \left[\widehat{E}(\theta_1, \dots, \theta_i + \frac{\pi}{2}, \dots, \theta_K) - \widehat{E}(\theta_1, \dots, \theta_i - \frac{\pi}{2}, \dots, \theta_K) \right]. \quad (3.5)$$

⁴Precisely (in the Schrodinger picture) what we mean is the gradient of an observable in the state prepared by the circuit. In the Heisenberg picture, we are measuring an observable that is a function of the circuit parameters.

⁵For additional discussion of the parameter shift rule see Appendix A

Let us make a few remarks about the practicality of this method. The first feature worth noting is that the gradient of the quantum circuit is actually given by the same quantum circuit, just with different angles. No additional hardware is needed beyond the size of the original variational circuit. Second, the Hamiltonian is a sum of Pauli terms. By linearity, one can apply this rule to each of the terms. To save on the number of circuit runs, the Pauli terms can be grouped into commuting sets. Finally, since the gradient information is calculated by the quantum device, the only classical computation required is storing gradient information, which scales as $O(K)$ rather than $O(2^N)$ as would be required for optimizing the circuit classically. This is critical for applying this algorithm to larger qubit systems, which will be important for reaching the large N limit of the SYK model, where the dual gravity picture is valid.

4 Numerical Results

In this section, we apply our variational algorithm to find the ground state of H_{TFD} as in Eq. (2.8). To summarize the procedure, we start by sampling the random couplings, \mathcal{J}_{ijkl} of the SYK model on N Majorana fermions. This amounts to selecting the couplings of the Hamiltonians H_L and H_R . Once those are fixed, we construct H_{TFD} , which consists of the sum of left and right Hamiltonians and a bilinear coupling between the two. For concreteness, we fix $\mu = .01$ for every sampling of the Hamiltonian. We then apply the Jordan-Wigner mapping that maps the Majorana operators to Pauli strings, so our final Hamiltonian acts on qubits. Our variational state is defined through the PQC described in Section 3.2 acting on the all $|+\rangle$ state, with rotation angles initialized to 0. We minimize the energy by taking derivatives with respect to the circuit parameters, which are obtained from the quantum circuit itself by the parameter shift rule. The circuit parameters are then updated by a modified version of gradient descent known as Adam [31] that takes into account the gradient information of previous iterations.⁶

With this procedure, we find the ground state of the H_{TFD} for $N = 8$ for generic instantiations of the couplings. For more details on the exact numerical procedure, we make the code publicly available.⁷ Simulations were carried out via the quantum simulation package `Yao.jl` [17]. Excitingly, as few as $d = 2$ layers can achieve the ground state for some coupling values. For $N = 12$, there is a very near 4-fold degeneracy of the ground state energy, but our variational approach reaches below the fourth excited state.

Finally, to bring this even one step closer to physical implementation, in Section 4.2 we include the effects of *shot noise* in our training. The gradients in the parameter shift rule are given by expectation values. In a simulation we can calculate exact expectation values. However, in a hardware experiment, these expectation values would have to be estimated from a finite number of measurements, or shots. One might worry that this type of sampling noise leads to gradients that are too noisy to find the true energy minima, though we find this is not the case.

⁶We make use of the `Flux.jl` package [32] for implementation of Adam.

⁷<https://github.com/vipasu/SYK-TFD>

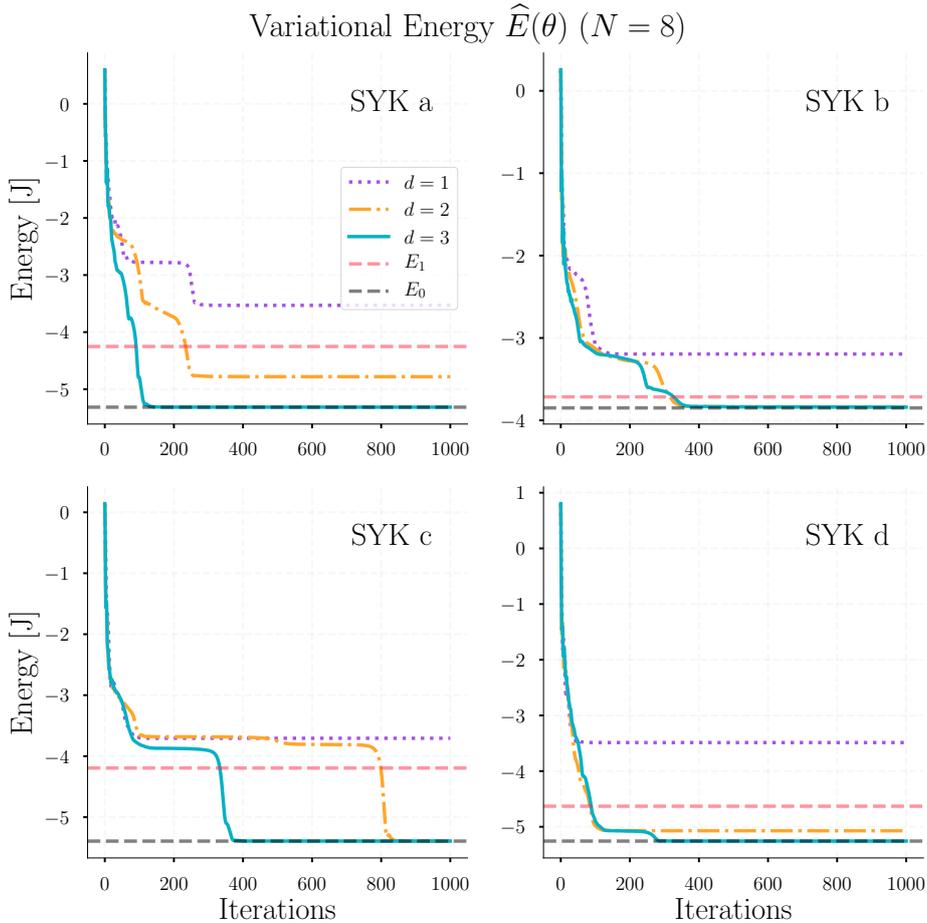


Figure 3: We plot the variational energy achieved by our parameterized circuit for four instantiations of the SYK model on $N = 8$ qubits. For each of the Hamiltonians, we plot the energies of the ground state and first excited state. Energy is expressed in units of the parameter $J = 1$ that sets the overall strength of the Hamiltonian. We show the variational energy for our PQC ansatz with different depths. As few as 2 layers can reach a variational energy below the first excited state in some cases. We could generically achieve the ground state with 3 layers of our PQC.

4.1 Results with Analytic Gradients

We start by carrying out our procedure with exact knowledge of the gradients. This gives us a baseline sense of whether this procedure can work in principle. We find that quite generally we are able to achieve energies below the first excited state for a relatively small system size of $N = 8$ qubits, or Majoranas per side, indicating high overlap with the ground state. For $N = 12$, the case is a bit more subtle since the size of the gap between the ground state and first excited state shrinks substantially. For these system sizes, the energy can be computed via exact diagonalization. For larger system sizes, one could use Krylov methods to extract low lying spectra. Apart from the energy, we can also check the

parity symmetry by computing the difference $H_L - H_R$.

Since we are simulating the circuit, we have access to exact expectation values. Because the parameter shift rule relies on expectation values for the gradient, we therefore also get perfect knowledge of the gradient. Thus, this section serves as a proof of principle. In a real experiment, one would have to deal with possibly faulty gates and using measurements to estimate expectation values. These are not well characterized, so it is difficult to predict their effect on this gradient based method. We run our simulations again with finite number of shots in Section 4.2.

For $N = 8$, fewer than 500 gradient iterations with depth $d = 3$ worked well enough to produce the ground state of several instantiations of the SYK Hamiltonian. In Figure 3, we plot the variational energy as a function of gradient updates for different depths of the circuit described in Section 3.2. This is often referred to as a training curve. As few as $d = 2$ layers was enough to reach energies below the first excited state.

An additional necessary, though insufficient, condition that the ground state has been achieved is parity symmetry. This is measured by the difference of $H_L - H_R$. One can straightforwardly verify that this should annihilate the TFD state as in Eq. (2.1). When using a symmetric two qubit entangler such as R_{XX} and starting from a parity symmetric state, we find that parity is maintained throughout. This is because the circuit is identical on both sides, so any circuit updates via the gradient are also symmetric. While other two qubit entangling gates, such as CNOT and controlled- R_Z gates could also achieve low variational energies, the variational states did not always obey the parity constraint.

For $N = 12$ qubits, we can run the same procedure, simply adding more terms to the Hamiltonian and expanding our PQC to include more qubits. For sufficiently large depth d , our procedure finds a state whose energy matches the ground state energy to within $\sim .2\%$. These results are summarized in Figures 4 and 5. We show a similar training curve for the larger system size in Figure 4. Due to a near degeneracy, the first few excited states visually overlap, so we plot the fourth excited state for comparison. Additionally, we compute the spectrum in Figure 5 to show the promise of gradient methods. Without taking into account any of the problem structure, this variational method is able to navigate a relatively large space of states.

Though it was not necessary for this example, one could in principle use the square of this quantity as a Lagrange multiplier in the energy minimization. Since $(H_L - H_R)^2$ would also be a sum of Pauli strings, this would likewise be amenable to the gradient techniques described in Section 3.3.

To recap, we have shown that a simple quantum circuit with reasonable depth can prepare the approximate TFD for the SYK model on $N = 8$ qubits. For $N = 12$ qubits, we can numerically simulate a quantum circuit that approaches the ground state energy, though more work would be required to bring it to a form implementable on a NISQ device. Larger system sizes were not readily comparable via exact diagonalization, so one would likely need to find a way to verify the variational state, for example by the parity check and possibly by matching thermal correlators. As an additional benefit of this variational computation, we have an explicit circuit that can be implemented on hardware for the system sizes considered.

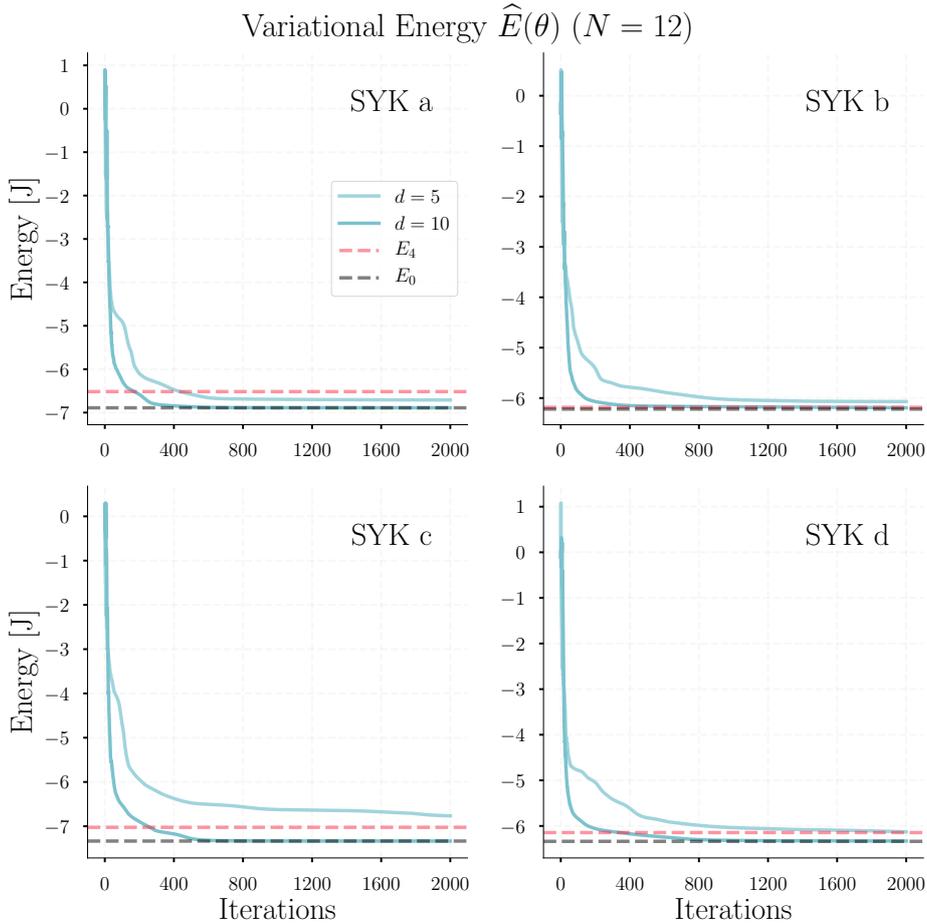


Figure 4: Variational energy achieved on $N = 12$. There is a near degeneracy in the spectrum, which is shown in Figure 5. The depth required was not significantly more relative to $N = 8$. For our ansatz, a PQC of depth d reaches a strict subset of states compared to depth $d+1$ since the last layer of rotations could have 0 angle. However, higher depth circuits did not always have lower variational energy at every training iteration, hinting to the subtleties in navigating the optimization landscape.

For larger system sizes, we again stress that the variational procedure can still be implemented on quantum devices. This is in line with the ultimate goal of finding ways to use quantum computers to probe physics we cannot already access on a classical computer. This hinges on the fact that all of the gradient information came from the quantum circuit itself. Thus, if the quantum hardware can carry out the ansatz circuit, it can also provide the gradient information necessary to perform this procedure.

4.2 Learning in the Presence of Shot Noise

At the core of this algorithm is the fact that gradients can be obtained from the quantum circuit by measuring expectation values. Within a simulation, the expectation values can

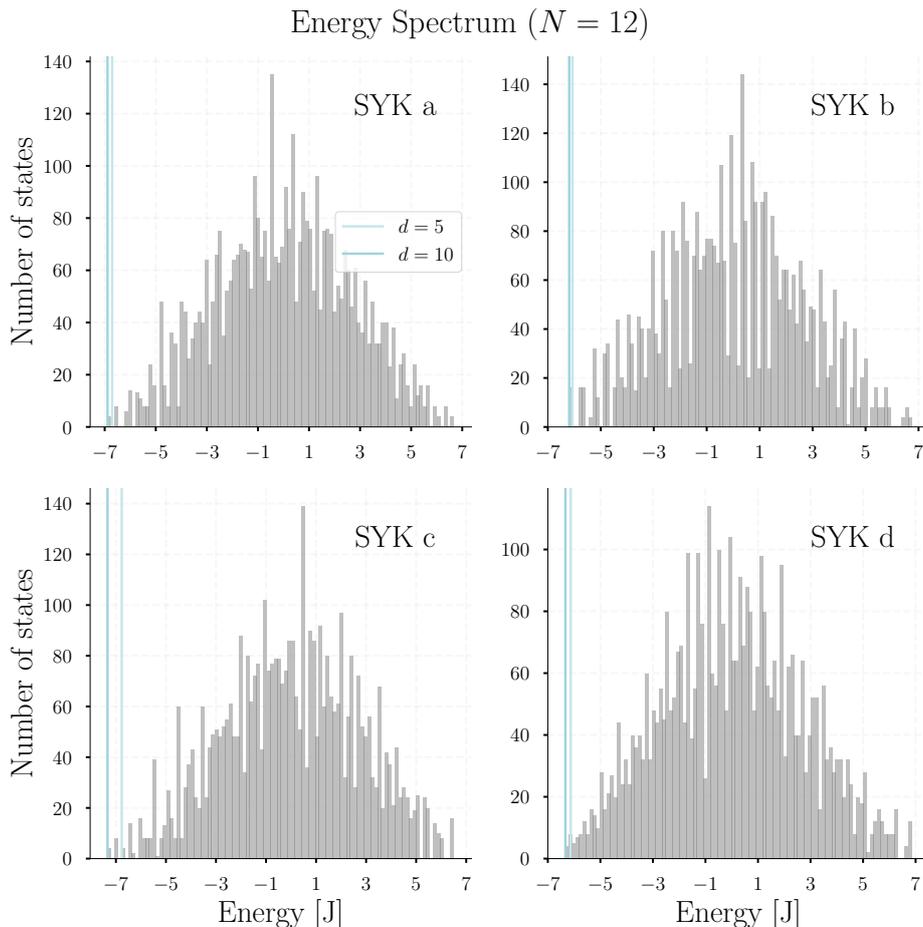


Figure 5: Spectrum obtained by exact diagonalization in a Hilbert space of dimension 2^{12} . Vertical lines denote the minimum energy achieved by PQCs of varying depth. Interestingly, there is a degeneracy of states near the ground state energy. This is an indication that the more general form of H_{TFD} in Ref. [14] could be more apt.

be computed exactly. On a physical device, however, one must take into account the noise introduced by having to average a finite number of measurements, or *shots*. This effect is called shot noise. In this section, we will demonstrate that our algorithm can find the ground state in the presence of this noise.

Let us precisely distinguish the noise we account for here. On a quantum device, the expectation value of an observable requires several measurements on the same circuit. With a finite number of measurements, you may not converge to the true expectation value. This is distinct from the notion that running the same circuit might implement different unitaries if there are uncalibrated errors. Since we are using expectation values of the quantum circuit to produce gradient estimates, both of these will be important for a physical implementation of our algorithm.

Amazingly, we found that the variational algorithm was able to succeed in the presence

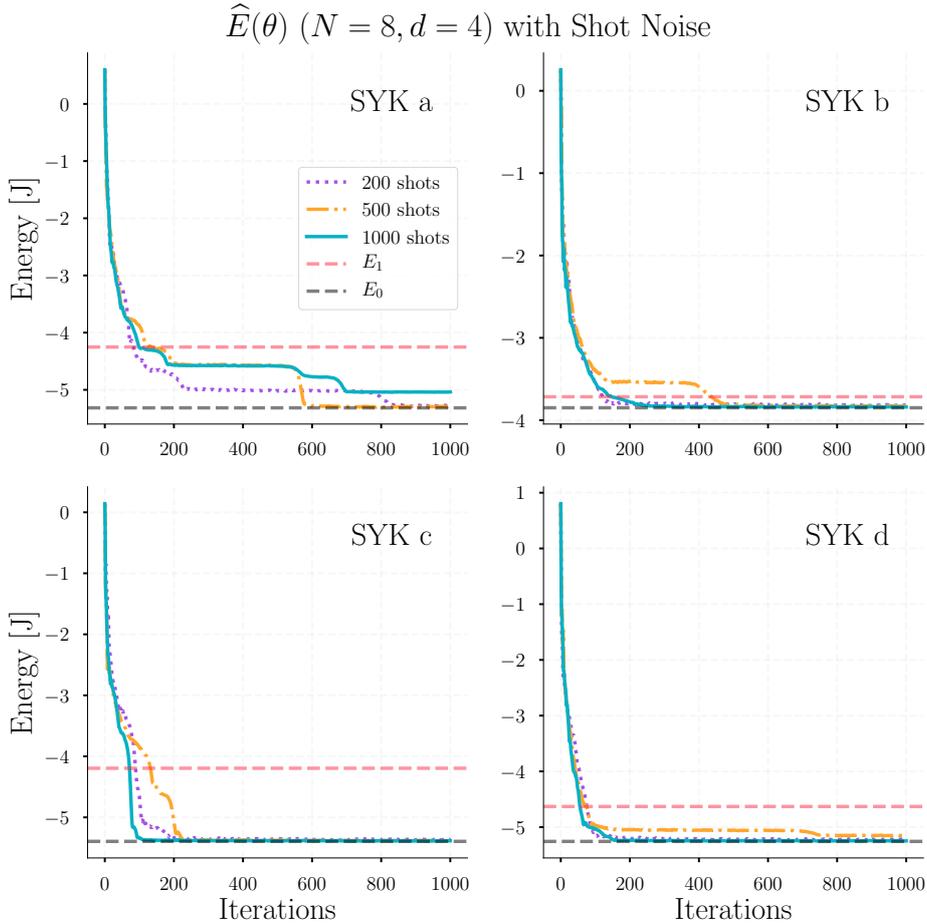


Figure 6: Variational energy achieved when taking shot noise into account. First we note that we had to augment the circuit depth from 3 to 4 to get comparable performance. However, this is not a fundamental obstacle to achieving energies well below the first excited state. Another interesting point is that performance was not monotonic with number of shots. This also reflects the difficulty of the optimization landscape.

of shot noise. One could worry that the errors in the gradient would render the algorithm unable to find the ground state energy. Indeed for the same depth $d = 3$, the algorithm did not converge to the ground state energy. However, by introducing one more layer to our ansatz, we show in Figure 6 that shot noise is not a fundamental obstacle. Interestingly, more shots does not always result in better performance.

It is not a priori obvious that adding more layers is a counter to the effects of shot noise. While the deeper circuit will certainly capture a larger class of states, there is not much evidence that the gradients would be any more noise resilient. In fact, intuition based on [33] would indicate that deeper circuits are more difficult to train. This could be due to our choice of initialization of trivial angles rather than random ones.

In Figure. 7, we plot the parity violation as a function of iteration. Unlike the previous

$H_L - H_R (N = 8)$ with Shot Noise

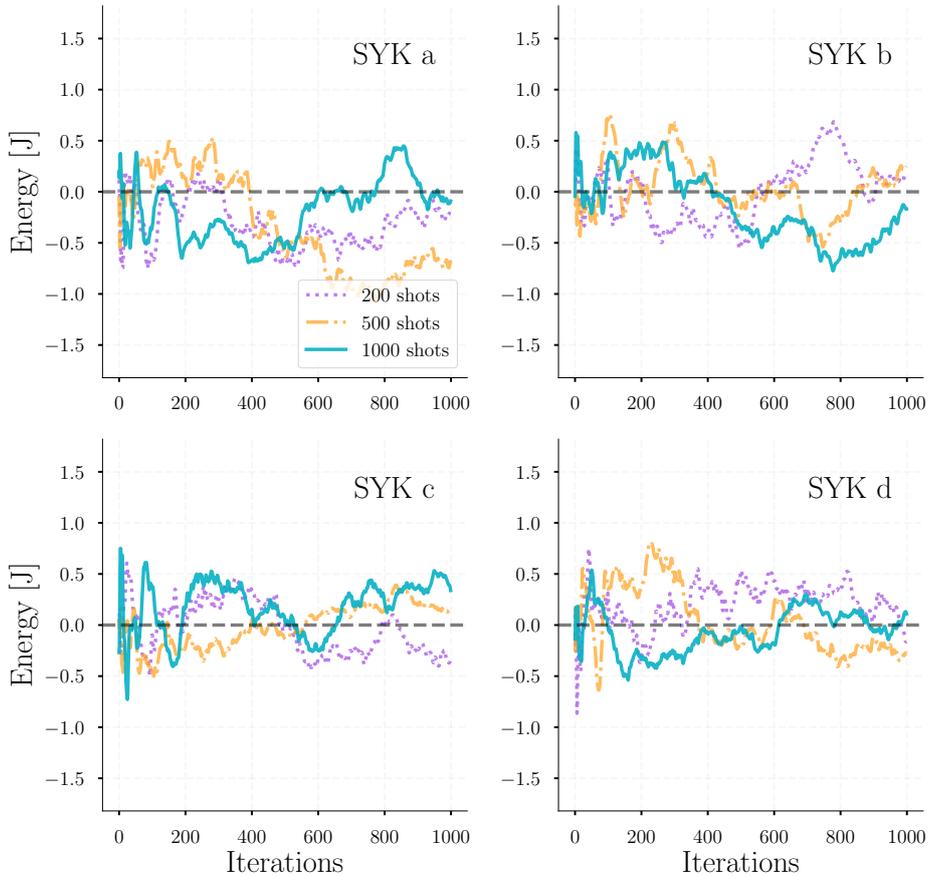


Figure 7: In the presence of shot noise, the parity constraint is violated even with a circuit that is parity symmetric. Even as the ground state energy is approached, the resulting state is not necessarily parity symmetric. An additional term to the Hamiltonian (the objective) that penalizes larger values of $|H_L - H_R|$ could be employed to ensure parity.

section with analytic gradients, this did remain trivially satisfied throughout training. As the ground state energy is approached, this quantity should tend to 0. The final value we observe is small but nonzero. One could add another penalty term to ensure parity either in conjunction with the energy or in an alternating fashion.

5 Discussion

The TFD is a central ingredient for many experiments in recent years with applications to chaos, scrambling, teleportation and more [10, 34, 35]. While most works assume the existence of the preparation, we provide a concrete realization for a chaotic Hamiltonian, without the need for an external bath.

In this work, we presented a preparation scheme for holographic probes that can be implemented on a quantum computer *today*. While we are far from the large N limit, this is an important stepping stone for marching theory towards experiment and vice versa. We were able to achieve near optimal energies for systems up to $N = 12$ qubits, beyond which exact comparison starts becoming much more computationally intense. On the theory side, studying $1/N$ corrections to the classical limit of Ref. [12] would help us understand how far we are from a true holographic state with a geometric dual.

Variational computing had been employed previously in constructing the TFD for the Ising model [15, 16], but this approach drastically cuts down on the classical part of the computation. By getting gradient information directly from the quantum hardware rather than doing the variational optimization by grid search, one can scale this approach to larger system sizes. An additional benefit is that the gradient obtained in this way will factor in the inherent noise due to physical implementation, whereas classical optimization assumes idealized gates.

We will conclude with some connections and additional directions for future work.

Hardware Implementation

It would be extremely interesting to prepare the TFD with this variational procedure on a quantum device. It would also enable us to run the traversable wormhole protocol. We make some comments to further close the gap between theory and experiment.

Our choice of PQC was rather uninspiring. We elaborate on why this design choice is a feature and not a bug. As alluded to in Section 3, gate compilation is a critical issue for ensuring that a circuit can run on a near term device.

The beauty in our simple circuit is that there was no optimization of the gates that indicated this would be optimal, or even remotely good, at preparing the TFD. Instead, it was built out of gates from a standard gate set that are perceived to be easy. We expect that hardware realizations of our procedure would use a variational circuit that is tailored to what is easily implementable. For fixed gate depth, this prioritizes the number of “layers” rather than worrying about compiling a specific set of gates. We repeated the experiments with different choice of two qubit entanglers and achieved similar performance.

Of course, a issue with NISQ devices is the noise. Our simulation did not include the effects of noise. However, an additional benefit of the parameter shift rule is that the quantum gradients do not make assumptions on the form of the unitary being implemented as a classical optimizer would. Namely, the quantum gradients as measurements yield the gradient for the noisy gate. This eliminates the need to characterize and calibrate systematic offsets. Of course, it would be great to verify the gradient procedure working in the presence of noisy gates on a physical device. An additional hardware consideration is the tradeoff between circuit depth providing a more rich ansatz and the additional noise introduced with each layer.

Our approach suggests two ways to implement this in the near term. The first is to use the simulations to obtain the optimal angles and plug them into a quantum computer. The more interesting case is to repeat the full procedure and learn on the quantum device,

in regimes we can no longer simulate. Section 4.2 provides evidence that shot noise is not a fundamental obstacle to this procedure working in practice.

Machine Learning

In this work, we started with a particularly simple form of variational ansatz. Using an off the shelf gradient descent procedure, we could achieve the ground state energy of H_{TFD} for system sizes up to $N = 12$. Both of these seem achievable in the NISQ era, but obviously if their circuits can be compressed, this would dramatically decrease the time to successful implementation. We propose two engineering approaches to tackling this problem. The first involves modifying the choice of ansatz with better circuit design, about which little is known theoretically.⁸ The second involves trying more sophisticated optimization methods.

On the optimization front, we present two ideas for investigation. The first is gaining a better understanding of the optimization landscapes of quantum circuits. Random PQCs have been shown to suffer from the problem of vanishing gradients [33], rendering the gradient-based navigation much less useful. There have been recent proposals for training schemes that get around this, such as training layer by layer [37].

The second, more ambitious approach would be to use reinforcement learning methods to automate the circuit design. This has obvious applications to gate compilation as well, but a successful implementation would circumvent the initial choice of circuit to compile altogether. Reinforcement learning has been successfully used in the context of autonomously preparing quantum states of floquet systems [38].

The idea of circuits whose architecture themselves is also variable is something that has emerged in the quantum chemistry community. See Ref. [39] and references therein. The algorithm is driven by constructing the circuit by greedily adding gates from a fixed gate set that result in a lower energy state.

Whether from a theoretical perspective or a practical implementation view, there is clear room for improvement when it comes to variational circuit design.

General TFD Preparation

It would also be interesting to see how to incorporate the more general approach of TFD preparation as in Ref. [14]. Formulating H_{TFD} in terms of additional operators would provide an alternate gradient for achieving the TFD state, one whose optimization landscape might be easier to exploit. An additional theoretical result they show is the existence of a gap in the spectrum of H_{TFD} in their construction even for larger system sizes. As evidenced in Figure 5, there was a near degeneracy of the ground state energy. Having a finite gap for larger system sizes would very likely help our variational optimization procedure.

An alternative approach to preparing approximate TFD states was discussed in Ref. [40] in an ansatz called the Product Spectrum Ansatz (PSA). The PSA is also variational, consisting of a product of mixed states which are then locally entangled with arbitrary 2 qubit gates. A comparison with the ground state method approach in this work is provided

⁸See [36] for one idea to compare the *expressibility* of different quantum circuits.

in [40]. It would be interesting to see whether one approach would work better in the context of a near term quantum device.

Simulating AdS/CFT

The role of the SYK model in simulating AdS/CFT was initially proposed in [41]. Towards an experimental realization of SYK simulation, a proposal using highly controllable ultracold gases was put forth in [42]. For a review of such holographic quantum matter proposals, see [43]. Numerical evolution of the SYK model was also carried out using classical resources recently in an effort to study scrambling and chaos [44].

With access to the TFD state, this gives a very interesting state to study time dynamics on a quantum computer. There, we would be interested in simulating time evolution using digital quantum gates. One promising avenue for digital Hamiltonian simulation is qubitization. This approach was taken for the SYK model in [45]. One could combine the evolution with this preparation method to carry out the traversable wormhole experiments.

Acknowledgments

It is a pleasure to thank Unpil Baek, Ning Bao, Raphael Bousso, Nicholas Cinko, Ben Freivogel, Tim Hsieh, Bill Huggins, Jin-Guo Liu, Xiu-Zhe Luo, Hugo Marrochio, Bryan O’Gorman, Xiao-Liang Qi, Pratik Rath, and K. Birgitta Whaley for interesting and useful discussion. Additionally, I thank the group at the NASA Quantum AI Lab for stimulating discussion and hospitality last summer. I am supported in part by the Berkeley Center for Theoretical Physics, by the National Science Foundation (award number PHY-1521446), and by the U.S. Department of Energy under contract DE-AC02-05CH11231 and award DE-SC0019380. I gratefully acknowledge support by the NSF GRFP under Grant No. DGE 1752814.

References

- [1] J. M. Maldacena, *The Large N limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231 [[hep-th/9711200](#)].
- [2] S. W. Hawking, *Particle creation by black holes*, *Communications in Mathematical Physics* **43** (1975) 199.
- [3] W. Israel, *Thermo-field dynamics of black holes*, *Physics Letters A* **57** (1976) 107.
- [4] J. Maldacena, *Eternal black holes in anti-de sitter*, *Journal of High Energy Physics* **2003** (2003) 021021.
- [5] S. Sachdev and J. Ye, *Gapless spin-fluid ground state in a random quantum heisenberg magnet*, *Phys. Rev. Lett.* **70** (1993) 3339.
- [6] A. Kitaev, *A simple model of quantum holography*, KITP Program: Entanglement in Strongly-Correlated Quantum Matter, May, 2015.
- [7] J. Maldacena, D. Stanford and Z. Yang, *Diving into traversable wormholes*, *Fortschritte der Physik* **65** (2017) 1700034.

- [8] R. Jackiw, *Lower Dimensional Gravity*, *Nucl. Phys. B* **252** (1985) 343.
- [9] C. Teitelboim, *Gravitation and Hamiltonian Structure in Two Space-Time Dimensions*, *Phys. Lett. B* **126** (1983) 41.
- [10] J. Maldacena, S. H. Shenker and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106 [[1503.01409](#)].
- [11] K. Jensen, *Chaos in AdS₂ Holography*, *Phys. Rev. Lett.* **117** (2016) 111601 [[1605.06098](#)].
- [12] J. Maldacena and X.-L. Qi, *Eternal traversable wormhole*, [[1804.00491](#)].
- [13] J. Preskill, *Quantum computing in the nisq era and beyond*, *Quantum* **2** (2018) 79.
- [14] W. Cottrell, B. Freivogel, D. M. Hofman and S. F. Lokhande, *How to build the thermofield double state*, *Journal of High Energy Physics* **02** (2019) 058 .
- [15] J. Wu and T. H. Hsieh, *Variational thermal quantum simulation via thermofield double states*, *Physical Review Letters* **123** (2019) 220502.
- [16] D. Zhu, S. Johri, N. M. Linke, K. A. Landsman, N. H. Nguyen, C. H. Alderete et al., *Generation of thermofield double states and critical ground states with a quantum computer*, *Proc. Nat. Acad. Sci.* **117** (2020) 25402 [[1906.02699](#)].
- [17] X.-Z. Luo, J.-G. Liu, P. Zhang and L. Wang, *Yao.jl: Extensible, efficient framework for quantum algorithm design*, *Quantum* **4** (2020) 341.
- [18] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, *Phys. Rev. D* **94** (2016) 106002 [[1604.07818](#)].
- [19] A. Kitaev and S. J. Suh, *The soft mode in the Sachdev-Ye-Kitaev model and its gravity dual*, *JHEP* **05** (2018) 183 [[1711.08467](#)].
- [20] J. Maldacena, D. Stanford and Z. Yang, *Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space*, *PTEP* **2016** (2016) 12C104 [[1606.01857](#)].
- [21] G. Sárosi, *AdS₂ holography and the SYK model*, *PoS Modave2017* (2018) 001 [[1711.08482](#)].
- [22] P. Gao, D. L. Jafferis and A. C. Wall, *Traversable wormholes via a double trace deformation*, *Journal of High Energy Physics* **12** (2017) 151.
- [23] P. Gao and D. L. Jafferis, *A Traversable Wormhole Teleportation Protocol in the SYK Model*, [[1911.07416](#)].
- [24] J. Maldacena and A. Milekhin, *SYK wormhole formation in real time*, *JHEP* **04** (2021) 258 [[1912.03276](#)].
- [25] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow et al., *Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets*, *Nature* **549** (2017) 242246.
- [26] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love et al., *A variational eigenvalue solver on a photonic quantum processor*, *Nature Communications* **5** (2014) 4213.
- [27] M. Nielsen, *The fermionic canonical commutation relations and the jordan-wigner transform*, 2005.
- [28] J. R. McClean, K. J. Sung, I. D. Kivlichan, Y. Cao, C. Dai, E. S. Fried et al., *Openfermion: The electronic structure package for quantum computers*, *Quantum Sci. Technol.* **5** (2020) 03401.

- [29] K. Mitarai, M. Negoro, M. Kitagawa and K. Fujii, *Quantum circuit learning*, *Physical Review A* **98** (2018) 032309.
- [30] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac and N. Killoran, *Evaluating analytic gradients on quantum hardware*, *Physical Review A* **99** (2019) 032331.
- [31] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, eds., 2015, 1412.6980.
- [32] M. Innes, *Flux: Elegant machine learning with julia*, *Journal of Open Source Software* **3**(25) (2018) 602.
- [33] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush and H. Neven, *Barren plateaus in quantum neural network training landscapes*, *Nature Communications* **9** (2018) 4812.
- [34] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067 [1306.0622].
- [35] P. Hosur, X.-L. Qi, D. A. Roberts and B. Yoshida, *Chaos in quantum channels*, *Journal of High Energy Physics* **2** (2016) 4 .
- [36] S. Sim, P. D. Johnson and A. AspuruGuzik, *Expressibility and entangling capability of parameterized quantum circuits for hybrid quantumclassical algorithms*, *Advanced Quantum Technologies* **2** (2019) 1900070.
- [37] A. Skolik, J. R. McClean, M. Mohseni, P. v. d. Smagt and M. Leib, *Layerwise learning for quantum neural networks*, [2006.14904].
- [38] M. Bukov, *Reinforcement learning for autonomous preparation of floquet-engineered states: Inverting the quantum kapitza oscillator*, *Physical Review B* **98** (2018) 224305 .
- [39] H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes et al., *Qubit-adapt-vqe: an adaptive algorithm for constructing hardware-efficient ansatzes on a quantum processor*, [1911.10205].
- [40] J. Martyn and B. Swingle, *Product spectrum ansatz and the simplicity of thermal states*, *Physical Review A* **100** (2019) 032107.
- [41] L. García-Álvarez, I. L. Egusquiza, L. Lamata, A. del Campo, J. Sonner and E. Solano, *Digital Quantum Simulation of Minimal AdS/CFT*, *Phys. Rev. Lett.* **119** (2017) 040501 [1607.08560].
- [42] I. Danshita, M. Hanada and M. Tezuka, *Creating and probing the sachdevykitae model with ultracold gases: Towards experimental studies of quantum gravity*, *Progress of Theoretical and Experimental Physics* **2017** 8 (2017) 083I01.
- [43] M. Franz and M. Rozali, *Mimicking black hole event horizons in atomic and solid-state systems*, *Nature Rev. Mater.* **3** (2018) 491 [1808.00541].
- [44] B. Kobrin, Z. Yang, G. D. Kahanamoku-Meyer, C. T. Olund, J. E. Moore, D. Stanford et al., *Many-Body Chaos in the Sachdev-Ye-Kitaev Model*, *Phys. Rev. Lett.* **126** (2021) 030602 [2002.05725].
- [45] R. Babbush, D. W. Berry and H. Neven, *Quantum Simulation of the Sachdev-Ye-Kitaev Model by Asymmetric Qubitization*, *Phys. Rev. A* **99** (2019) 040301 [1806.02793].
- [46] A. M. Childs and N. Wiebe, *Hamiltonian simulation using linear combinations of unitary operations*, *Quantum Information and Computation* **12** (2012) pp0901-0924.

A Parameter Shift Rule for a single qubit

The parameter shift rule allows us to compute the gradient of the expectation value of an observable with respect to the parameters of a PQC. The technique is quite remarkable and worth highlighting. First, it is not obvious at all that the output of *any* quantum circuit will give information about how to tune parameters of a variational circuit. Second, of great practical importance, is the amazing fact that it is the *same* variational circuit that gives gradient information about the quantum circuit! Both of these are consequences of the parameter shift rule. We start by discussing a toy example, following the original work of [29]⁹.

For the sake of simplicity, let us consider a single qubit which can be visualized on the Bloch sphere. Our parameterized quantum circuit will consist of a single Pauli rotation gate

$$R_X(\theta) \equiv \exp(iX\theta/2), \quad (\text{A.1})$$

where we use interchangeably the notation X for the Pauli operator σ^x .

Suppose we are interested in minimizing the expectation value of the observable Z . Clearly, the optimal qubit state which minimizes $\langle Z \rangle$ is the $|1\rangle$ state at the south pole of the Bloch sphere. In the language described in previous sections, our circuit is $U(\theta) = R_X(\theta)$ and produces the variational state $|\psi\rangle = U(\theta)|0\rangle$.

To minimize the expectation value of Z , we would like to calculate the gradient with respect to θ .

$$\frac{\partial \langle Z(\theta) \rangle}{\partial \theta} = \frac{\partial \langle 0|U^\dagger(\theta)ZU(\theta)|0\rangle}{\partial \theta}. \quad (\text{A.2})$$

By the product rule,

$$\frac{\partial \langle 0|U^\dagger(\theta)ZU(\theta)|0\rangle}{\partial \theta} = \frac{i}{2} \langle 0|U^\dagger(\theta)(XZ - ZX)U(\theta)|0\rangle. \quad (\text{A.3})$$

Using the Pauli algebra, one can express the commutator $[X, Z]$

$$[X, Z] = i \left(U^\dagger \left(\frac{\pi}{2} \right) ZU \left(\frac{\pi}{2} \right) - U^\dagger \left(-\frac{\pi}{2} \right) ZU \left(-\frac{\pi}{2} \right) \right) \quad (\text{A.4})$$

Plugging back into Eq (A.3),

$$\frac{\partial \langle Z(\theta) \rangle}{\partial \theta} = \frac{1}{2} \left(\langle 0|U^\dagger(\theta)U^\dagger \left(\frac{\pi}{2} \right) ZU \left(\frac{\pi}{2} \right) U(\theta)|0\rangle - \langle 0|U^\dagger(\theta)U^\dagger \left(-\frac{\pi}{2} \right) ZU \left(-\frac{\pi}{2} \right) U(\theta)|0\rangle \right) \quad (\text{A.5})$$

Note that the extra applications of U can be absorbed into the original angle, leaving us with a simple form.

$$\frac{\partial \langle Z(\theta) \rangle}{\partial \theta} = \frac{1}{2} (\langle Z(\theta^+) \rangle - \langle Z(\theta^-) \rangle) \quad (\text{A.6})$$

where $\theta^\pm = \theta \pm \pi/2$. We can see now that the derivative is given by the difference of two expectation values using exactly the same form of the quantum circuit, as promised at the beginning of the section.

⁹See also https://pennylane.ai/qml/glossary/parameter_shift.html

For other gates which are generated by products of Pauli terms, a similar analysis holds. It may seem overly restrictive to consider gates which are generated by the Pauli matrices. However, this was generalized in Ref. [30], where they consider the primary task of evaluating these circuit gradients on actual hardware. They first generalize the parameter shift rule beyond tunable gates which are generated by the Pauli's. It suffices to consider gates whose generators have up to two eigenvalues. As they point out, three of Google's native Xmon gates satisfy this criteria. Alternatively, for circuit architectures which do not fit this structure, one can apply *linear combination of unitaries* [46] using an ancilla qubit to get estimates of the gradient.