# Measurement-induced entanglement phase transitions in variational quantum circuits

Roeland Wiersema,[1,2] Cunlu Zhou,[3,4,1] Juan Felipe Carrasquilla,[1,2,5] and Yong Baek Kim[5]

[1] *Vector Institute, MaRS Centre, Toronto, Ontario, M5G 1M1, Canada*
[2] *Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada*
[3] *Department of Physics and Astronomy and Center for Quantum Information and Control, CQuIC, University of New Mexico, Albuquerque, New Mexico 87131, USA*
[4] *Department of Computer Science, University of Toronto, Ontario, M5T 3A1, Canada*
[5] *Department of Physics, University of Toronto, Ontario M5S 1A7, Canada*

(Dated: June 18, 2022)

Variational quantum algorithms (VQAs), which classically optimize a parametrized quantum circuit to solve a computational task, promise to advance our understanding of quantum many-body systems and improve machine learning algorithms using near-term quantum computers. Prominent challenges associated with this family of quantum-classical hybrid algorithms are the control of quantum entanglement and quantum gradients linked to their classical optimization. Known as the barren plateau phenomenon, these quantum gradients may rapidly vanish in the presence of volume-law entanglement growth, which poses a serious obstacle to the practical utility of VQAs. Inspired by recent studies of measurement-induced entanglement transition in random circuits, we investigate the entanglement transition in variational quantum circuits endowed with intermediate projective measurements. Considering the Hamiltonian Variational Ansatz (HVA) for the XXZ model and the Hardware Efficient Ansatz (HEA), we observe a measurement-induced entanglement transition from volume-law to area-law with increasing measurement rate. Moreover, we provide evidence that the transition belongs to the same universality class of random unitary circuits. Importantly, the transition coincides with a "landscape transition" from severe to mild/no barren plateaus in the classical optimization. Our work paves an avenue for greatly improving the trainability of quantum circuits by incorporating intermediate measurement protocols in currently available quantum hardware.

Keywords: Variational quantum circuits, quantum entanglement, entanglement phase transitions, barren plateaus

Controlling quantum entanglement has been identified as a critical element in the development of quantum computing. A prominent example is variational quantum algorithms, which are designed for quantum simulations and machine learning in near-term quantum computers [1]. How to efficiently use quantum entanglement resources impacts the performance of such quantum algorithms. Another important element in a quantum algorithm is the measurement, which can be performed at intermediate steps during a quantum computation and directly affects the overall entanglement structure of the quantum circuit. As there now exists quantum hardware that allows intermediate measurements [2, 3], it stands to reason that such measurements may offer yet another resource for controlling entanglement and improving VQAs.

Recently, there has been great progress in understanding the quantum entanglement evolution in random unitary quantum circuits with intermediate projective measurements. In random unitary circuits, the time evolution is governed by unitaries drawn from a random distribution without specifying any Hamiltonian. In these circuits, the nearby two-qubit gates locally entangle qubits, which generally leads to volume-law entanglement growth. When such a system is measured at randomly selected locations throughout the circuit, entanglement is destroyed globally. One might expect that this leads to a simple decrease in the coefficient of the entanglement growth volume law; however, this is not the case. The competition between local entanglement creation and non-local entanglement destruction induces a phase transition in the entanglement growth from a volume to an area law at a critical measurement rate $p_c$ [4–14]. Moreover, it appears that this critical behavior is universal, independent of the specific implementation of both the unitary or measurement dynamics. A significant amount of theoretical understanding has been gained about the properties of entanglement phase transitions in random unitary circuits [9, 15] by mapping such systems to well-defined statistical mechanics models. So far, most of the studies on measurement-induced phase transitions are focused on random unitary circuits or circuits with quantum chaotic dynamics. But do these transitions also take place in circuits of practical interest such as the variational quantum circuits used in quantum chemistry, quantum many body simulations, and quantum machine learning?

In this letter, we connect two highly active fields of research in condensed matter theory and variational quantum computing, by showing that measurement-induced entanglement phase transitions take place in two prototypical variational quantum circuits used within the Variational Quantum Eigensolver (VQE) algorithm [16]. This hybrid quantum-classical algorithm is used throughout the literature to approximate quantum many-body ground states [17–21] or perform quantum chemistry

simulations [22–26]. Our motivations to investigate the measurement-induced entanglement transitions in variational quantum circuits rest on the following two issues. Most of the quantum ground states of interacting many-body systems follow the area-law entanglement (up to a logarithmic correction). However, ballistic growth of entanglement in time evolution implies that circuits used in VQE can rapidly develop much more entanglement than what may be needed to efficiently simulate these ground states of interest [20, 27, 28]. The second issue is the evaluation of the quantum gradient, which is used to minimize a cost function in VQAs. It is known that quantum circuits that approximate a 2-design have exponentially decaying quantum gradients, localized on so-called barren plateaus, which pose a significant hurdle for variational quantum algorithms [29–31]. It has also been shown that there is a close relation between entanglement scaling and barren plateaus, hence it is natural to consider constraining the amount of entanglement during parts of the variational optimization as a useful strategy for increasing the trainability of variational circuits [32–34].

We anticipate that the inclusion of interspersed measurements in the variational quantum circuits may offer an alternative way to control their quantum entanglement, which can be used for more efficient quantum simulations, as well as to overcome the issue of barren plateau in the evaluation of quantum gradients. Below we numerically show that the measurement-induced entanglement phase transition in the variational quantum circuits coincides with a "landscape transition", a change from a landscape with severe barren plateaus to a landscape with mild or no barren plateaus. This suggests that VQE with intermediate projective measurements can potentially be used to avoid barren plateaus and improve current optimization strategies. In deriving our results, we also provide a modified parameter shift rule for calculating the quantum gradients with intermediate projective measurements.

*Measurement–induced entanglement phase transitions.*—We consider a chain of $N$ qubits with spatial periodic boundary conditions. To quantify the entanglement we consider the bipartite von Neumann entanglement entropy $S(N,p)$ between two halves of an $N$-qubit circuit of depth $L$. At each discrete time step $1 \leq d \leq L$, we apply a layer of local unitaries consisting of a combination of single qubit and two-local quantum gates between nearest neighbor qubits, which rapidly increase the entanglement in the chain. After each layer, we apply projective measurements onto the computational basis $\{|0\rangle, |1\rangle\}$ on each qubit with probability $p$. A single measurement on a state $\rho = |\psi\rangle\langle\psi|$ then results in a state

$$\rho = \frac{\Pi_i \rho \Pi_i}{\mathrm{Tr}\{\Pi_i \rho\}}, \tag{1}$$

where $\Pi_i = |i\rangle\langle i|$ are the projectors onto $\sigma^z$ basis. These measurements destroy entanglement at any length scale due to a local projection onto a single state [6]. As a result, the unitary dynamics locally entangles nearest neighbor qubits, whereas measurements globally destroy entanglement between different subsystems. This competition induces a dynamical phase transition between a volume and area law regime of entanglement scaling at a critical measurement rate $p_c$.

Although the critical point $p_c$ can vary between different types of random unitary dynamics and measurement schemes, the critical exponent characterizing the correlation length scale divergence $\xi \propto (p - p_c)^{-\nu}$ appears to be the same for different models at $\nu \approx 4/3$. This critical exponent can be derived by considering toy models and mapping the projective dynamics to a two-dimensional percolation model, which is exactly solvable [6, 9, 10, 15].

Central to the investigations on phase transitions induced by measurements is the concept of steady state entanglement dynamics [6, 7]. Given a circuit with a number of qubits $N$, we are primarily interested in the late time behavior when $d \to \infty$. In this infinite depth (long time) limit we expect the system to evolve into a steady state, characterized by a typical value of entanglement entropy that depends on the measurement rate $p$, but not the dynamics at finite times. In order to characterize this regime, we can investigate the average entanglement entropy as a function of depth for different values of $p$. For the moderate system sizes considered in this work, we observe steady state entanglement dynamics at $d = 16$.

*The Variational Quantum Eigensolver.*—Instead of considering Haar random circuits, Floquet dynamics or Random Clifford circuits, we turn to quantum circuits widely used in VQAs, e.g., in VQE algorithms [16]. In this hybrid quantum-classical algorithm we consider a quantum circuit $U(\boldsymbol{\theta})$ parametrized by a set of parameters $\boldsymbol{\theta}$. By invoking the variational principle $E_{\text{ground}} \leq E(\boldsymbol{\theta}) = \langle H \rangle_\theta$, one can use a classical optimization routine to minimize the energy of a Hamiltonian $H$ with respect to the parametrized wave function $\rho(\boldsymbol{\theta}) = U(\boldsymbol{\theta})\,|0\rangle\langle 0|\,U^\dagger(\boldsymbol{\theta})$ and approximate the ground state of $H$. As with other variational methods, the choice of ansatz $U(\boldsymbol{\theta})$ is crucial since the ground states must be reachable from the initial state by application of this unitary. There exists a variety of proposals, including the so-called Hamiltonian Variational Ansatz (HVA) [17–21] and the Hardware efficient Ansatz (HAE) [22, 35, 36]. The former exploits the structure of the Hamiltonian for the unitary ansatz design, whereas the latter aims to provide a hardware-friendly parameterization with enough degrees of freedom to capture a variety of states. For our numerical study, we investigate the projective dynamics of the XXZ-chain HVA and the HEA, whose circuits are depicted in fig. 1. Notice that the dynamics in the HVA is specified by a Hamiltonian in contrast to random unitaries. The HVA for the XXZ model is of particular

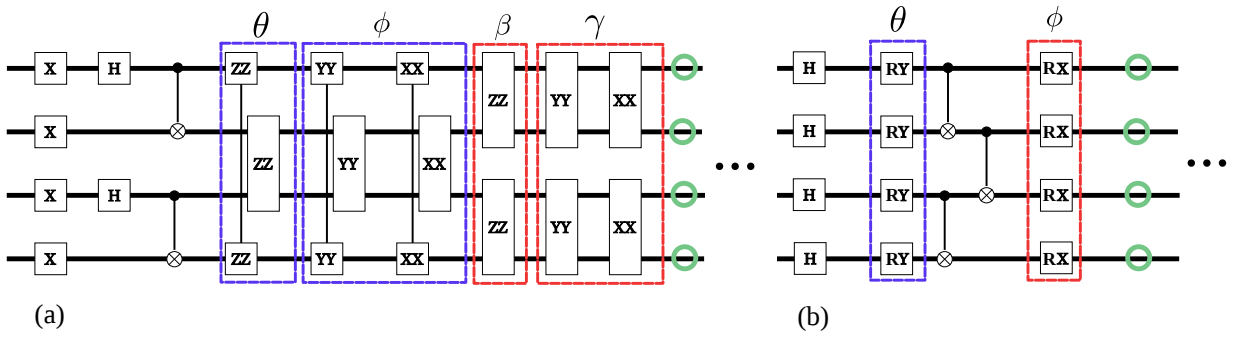



FIG. 1. Schematic depiction of the circuits studied in this letter. (a) For the XXZ-HVA, we prepare a Bell state on the even sites and alternatingly apply ZZ, YY and XX two-qubit rotations on odd and even bonds in the chain, which corresponds to the unitary rotations generated by the terms in the of the Hamiltonian $H_{\mathrm{XXZ}} = \sum_{i=1}^{N} \left[\sigma_i^x \sigma_{i+1}^x + \sigma_i^y \sigma_{i+1}^y + \Delta \sigma_i^z \sigma_{i+1}^z\right]$. For the odd (even) bonds, the ZZ rotations are parametrized by $\theta$ ($\beta$) whereas the YY and XX rotations are parametrized by $\phi$ ($\gamma$). (b) The initial state in the HEA consists of the equal superposition followed by $L$ layers of low-depth entangling unitaries. These unitaries consists of $N$ Pauli-Y rotations on each qubit, a chain of nearest neighbor CNOTs and $N$ Pauli-X rotations on each qubit. All $2N$ rotations are controlled by individual parameters $\theta_{i,l}, \phi_{i,l}$, where $i = 1, \ldots, N$ and $l = 1, \ldots, L$. After each layer, we perform a projective measurement according to eq. (1) with probability $p$ on each qubit (indicated by the green circles here), bringing the average number of measurements in the circuits to $NLp$.

interest as it represents the simulation of a non-trivial interacting model of fermions, which is Bethe-ansatz integrable and it is still an open question if Bethe-ansatz integrable models also show measurement-induced entanglement phase transitions [9]. For the HEA, it is expected that the behavior is close to that of random circuits [30].

For each circuit, we measure the bipartite entanglement entropy of the output state of the circuit. Note that we are not creating a mixed state by remixing the individual intermediate measurement states into a single density matrix. Instead, we measure the entanglement entropy of each pure state in the ensemble individually. The average $S(p, N)$ is obtained by averaging over $3 \times 10^3$ circuit realizations with all circuit parameters sampled uniformly in $(0, 2\pi)$ and measurements sampled uniformly with probability $p$. Due to the difficulty in simulating large systems, we restrict ourselves to $N = 6, 8, \ldots, 18$.

*Finite size scaling analysis.*—Since phase transitions only occur in the thermodynamic limit $N \to \infty$, we have to take care of the finite-size effects in analyzing our numerical data. To account for finite-size effects, we fit the scaling form [5, 6, 9]

$$S(N, p, \nu) - S(N, p_c, \nu) = f(N^{1/\nu}(p - p_c)) \qquad (2)$$

where $f$ is a scaling function, to get a data collapse of the individual circuits of size $N$. To determine $p_c$ and $\nu$, we minimize a Chi-squared statistic between the scaling form above and the data, and use a statistical bootstrap to verify the integrity of the fit. In fig. 2 we find critical exponents close to the previously mentioned value of $\nu \approx 4/3$. To extrapolate the critical exponent to the thermodynamic limit, we do a linear fit of $\nu$ as a function of $1/N'$ where $N_{\max}/2 \leq N' \leq N_{\max}$ is the largest value of $N$ in the data set. The intercept then gives us the value of $\nu$ for $N' \to \infty$ [5]. The details of our statistical estimation procedure are outlined in the Supplementary Materials A..

In addition to the finite scaling analysis, we can investigate the quantum mutual information,

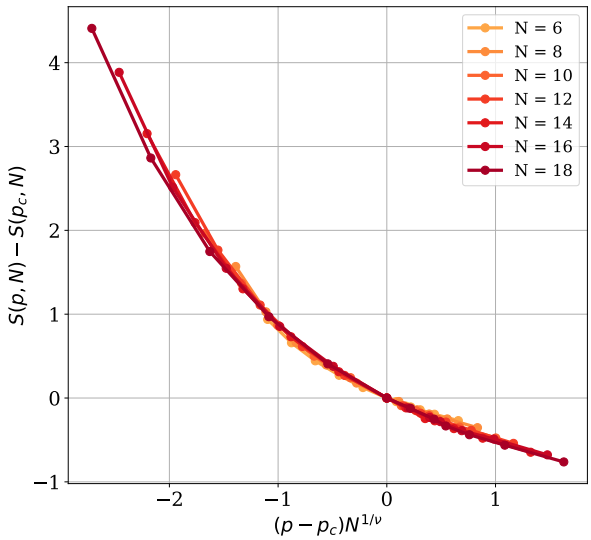
$$I(A, B) = S_A(N, p) + S_B(N, p) - S_{A \cup B}(N, p), \qquad (3)$$

between qubits $A$ and $B$ separated by a distance $r$, which we expect to peak at a critical point due to subsystem correlations becoming non-negligible. From these data, we find similar critical measurement rates $p_c \approx 0.25$ and $p_c \approx 0.5$ for the XXZ-HVA and HEA, respectively. In Supplementary Materials B. we give further details on this procedure.

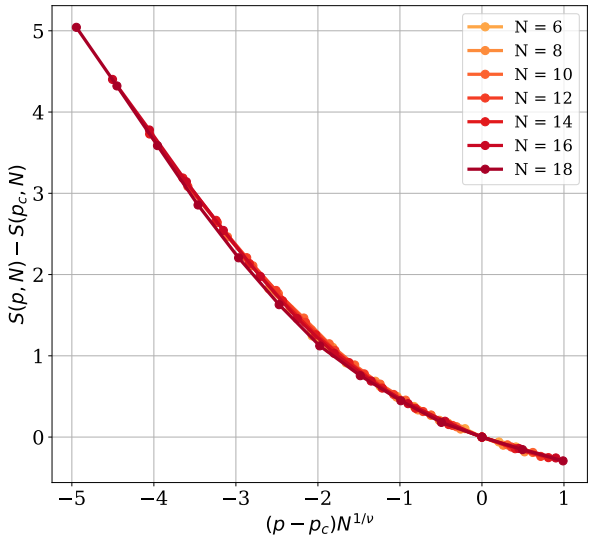
*Projective gradients and barren plateaus.*—For variational quantum algorithms,one prepares a parametrized quantum circuit on a quantum computer and feeds the output state to a classical optimizer which updates the parameters by minimizing the energy. Gradient descent is the most widely used classical optimization method for such general non-convex optimization problems but is plagued by barren plateaus. As a result, a variety of recent works are aimed at finding ways to circumvent these regions where optimization is hard [30, 33, 37–41]. Here, we investigate barren plateau problems under the influence of projective measurements, more specifically the variance of the gradients in the XXZ-HVA and the HEA with intermediate projective measurements.

Although, there exists a variety of methods for calculating gradients in parametrized quantum circuits [42–49], none of these works consider quantum gradients through a circuit undergoing projective measurements.

In [50] the quantum natural gradient [51] is extended to quantum channels. Additionally, in [52] measurement-based VQE is investigated, but only in the context of the




(a) XXZ-HVA




(b) HAE

FIG. 2. Data collapse of the average entanglement entropies. (a) For the XXZ-HVA, we find $p_c = 0.25$ and $\nu \approx 1.22 \pm 0.24$. (b) For the HEA, we find $p_c \approx 0.5$ and $\nu \approx 1.26 \pm 0.23$. The error bars are calculated as the difference between the critical exponent in the thermodynamic extrapolation and the finite-size data collapse. The entropies are averaged over $3 \times 10^3$ circuit realizations.

work by Briegel [53] where an entangled state is prepared and measurement is directly part of the algorithm.

In Supplementary Materials C., we provide a detailed analytical derivation of a quantum gradient in a circuit undergoing $M$ measurements. The result is an equation for calculating projective gradients of the form,

$$\mathrm{Tr}\{(\partial_{\theta_l}\rho)O\} = \sum_i^{2^M} \frac{1}{2}\left(\langle O\rangle^{(i),+}\frac{p^{(i),+}}{p^{(i)}} - \langle O\rangle^{(i),-}\frac{p^{(i),-}}{p^{(i)}}\right). \tag{4}$$

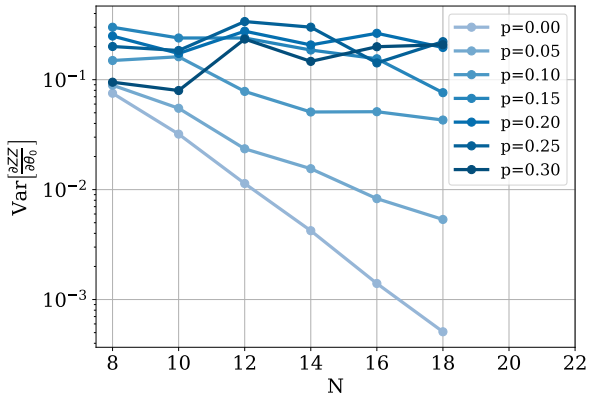
Here, $\rho \equiv \rho(\boldsymbol{\theta}) = \sum_i^{2^M} p^{(i)}\rho(\boldsymbol{\theta})^{(i)}$ is a mixture of pure states $\rho(\boldsymbol{\theta})^{(i)}$ that are the result of intermediate measurements. The bitstrings $(i) = (i_1, \ldots, i_M) \in \{0,1\}^M$ indicate the measurement outcomes of the projective measurements. The probabilities $p^{(i)}$ and $p^{(i),\pm}$ are the probabilities of observing outcomes $(i)$ on the circuit with and without $\theta_l$ shifted by $\pm\pi/2$, respectively. In this work, we have computed those weights exactly and we will explore the scalability of their estimation in future work. The expectation values $\langle O\rangle^{(i),\pm}$ correspond to the expectation value of $O$ if the circuit has the $\theta_l$ parameter shifted by $\pm\pi/2$.

The average entanglement entropy of sufficiently random quantum states is given by the Page entropy [54], which follows a volume law entanglement scaling. With intermediate projective measurements, volume-law states can be broken down into states with less entanglement, i.e., area-law/logarithmic-law states which are not Haar random. As a results, we do not expect barren plateaus to occur for these states [32, 34]. To investigate this effect, we consider the same circuit settings discussed in fig. 2 and examine the projective gradients with respect to the expectation value of $H = Z_0 Z_1$. We calculate the projective gradients for the first circuit parameter ($\theta$ in the first parametrized layer in both the HVA and HEA, see fig. 1) using eq. (4). We consider a depth $d = 16$ circuit for system sizes $N = 8, \ldots 18$. In fig. 3a and fig. 3b, we observe that the gradient variances in both the XXZ-HVA and HEA transition from exponentially decaying to a constant as the measurement rate increases. This transition, coincides with the critical measurement rate for the volume-area law transition. Therefore, we see that the measurement-induced entanglement phase transitions induces a landscape transition in the circuit from mild/severe barren plateaus to no barren plateaus. This landscape transition can serve as the motivation for a projective gradient VQE algorithm where the early optimization of the circuit is done with projective gradients to escape barren plateaus due to initialization. An important caveat to consider is that intermediate measurements on the circuit change the nature of the optimization problem, since we are effectively minimizing the mixture

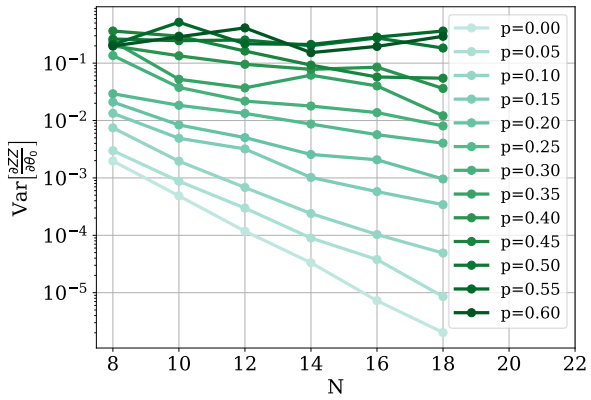
$$E_{\text{ground}} \leq E_{\text{int}}(\boldsymbol{\theta}) = \sum_i^{2^M} p^{(i)}\,\mathrm{Tr}\left\{\rho^{(i)}(\boldsymbol{\theta})H\right\} \tag{5}$$

Hence we require that the measurement rate is either annealed or set to zero after a number of optimization

steps to ensure that the final state is pure. We leave the exploration of this class of algorithms as future work.




(a) XXZ-HVA




(b) HAE

FIG. 3. Variance of the projective gradients taken with respect to the first parameter in the circuit ($\theta$ in the first parametrized layer in both the HVA and HEA, see fig. 1). The variances are estimated over $10^3$ samples where for each data point, we randomly choose measurements with probability $p$ and uniformly sample the gate parameters. The gradient is then calculated exactly from eq. (4). For the 1D HVA-XXZ circuits with depth $d = 16$ (a) and the 1D HAE circuit with depth $d = 16$ (b) the gradient variance becomes constant as the measurement rate $p$ increases.

*Outlook.*—In this work, we demonstrated the existence of a measurement-induced entanglement transition in variational quantum circuits which coincides with a "landscape transition" in the behavior of quantum gradients. As mentioned earlier, the exponentially-vanishing quantum gradients in presence of volume-law entanglement growth, the so-called barren plateau, is a serious obstacle in the applications of variational quantum circuits. Our work suggests that the intermediate projective measurements may provide a useful knob to control the barren plateau issue. Inclusion of the measurement protocol in the quantum-classical hybrid algorithm would be a timely development given that quantum computing hardware companies like IBM and Honeywell now allow their users to perform mid-circuit measurements, enabling the real-time logic required for performing these algorithms in an experimental setting [2, 3]. In particular, the Hamiltonian variational quantum circuits considered in this work could be implemented in the quantum hardware. For the projective gradient VQE, the exponential sum in eq. (4) currently inhibits the number of measurements that can be performed in practice. A detailed analysis of when and how a projective circuit optimization can be practical and "advantageous" would be an excellent topic of future study.

For a practical implementation of a projective gradient VQE algorithm, note that the scheme we provided here is quite general and many extensions and modifications are possible. For instance, the projective measurements used in this work can be replaced by general Positive Operator Value Measures (POVM) or parametrized measurements. Additionally, we have focused on one-dimensional quantum circuits where the measurement-induced entanglement transition belongs to the same universality class as in the random unitary circuits. It would be interesting to consider moderately sized quantum circuits with a two-dimensional topology, and see if a similar phase transition appears there and investigate the universality class.

*Acknowledgements.*—We would like to thank Henry Yuen for the discussions during the course of this project. Y.B.K. is supported by the NSERC of Canada and the Center for Quantum Materials at the University of Toronto. J.C. acknowledges support from NSERC, the Shared Hierarchical Academic Research Computing Network (SHARCNET), Compute Canada, Google Quantum Research Award, and the CIFAR AI chair program. Resources used in preparing this research were provided, in part, by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute www.vectorinstitute.ai/#partners. C.Z. acknowledges support from the U.S. National Science Foundation under Grant No. 2116246 and the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator, and the Postgraduate Affiliate Award from the Vector Institute.

[1] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational Quantum Algorithms. *Nature Reviews Physics*, 3(9):625–644, September 2021.

[2] Honeywell Quantum. Mid-circuit measurements on the

System Model H1. https://www.honeywell.com/us/en/company/quantum/quantum-computer/, 2021. [Online; accessed 07/18/2021].

[3] IBM Quantum. Mid-circuit Measurements Tutorial. https://quantum-computing.ibm.com/lab/docs/iql/manage/systems/midcircuit-measurement/, 2021. [Online; accessed 07/18/2021].

[4] Amos Chan, Rahul M. Nandkishore, Michael Pretko, and Graeme Smith. Unitary-projective entanglement dynamics. *Phys. Rev. B*, 99:224307, 6 2019.

[5] Brian Skinner, Jonathan Ruhman, and Adam Nahum. Measurement-Induced Phase Transitions in the Dynamics of Entanglement. *Phys. Rev. X*, 9:031009, 07 2019.

[6] Yaodong Li, Xiao Chen, and Matthew P. A. Fisher. Measurement-driven entanglement transition in hybrid quantum circuits. *Phys. Rev. B*, 100:134306, 10 2019.

[7] Yaodong Li, Xiao Chen, and Matthew P. A. Fisher. Quantum Zeno effect and the many-body entanglement transition. *Phys. Rev. B*, 98:205136, 11 2018.

[8] Xiangyu Cao, Antoine Tilloy, and Andrea De Luca. Entanglement in a fermion chain under continuous monitoring. *SciPost Phys.*, 7:24, Aug 2019.

[9] Yimu Bao, Soonwon Choi, and Ehud Altman. Theory of the phase transition in random unitary circuits with measurements. *Phys. Rev. B*, 101:104301, 3 2020.

[10] Stefanie Czischek, Giacomo Torlai, Sayonee Ray, Rajibul Islam, and Roger G. Melko. Simulating a measurement-induced phase transition for trapped ion circuits, 2021. arXiv:2106.03769.

[11] Maxwell Block, Yimu Bao, Soonwon Choi, Ehud Altman, and Norman Yao. The Measurement-induced Transition in Long-range Interacting Quantum Circuits, 2021. arXiv:2104.13372.

[12] Michele Coppola, Emanuele Tirrito, Dragi Karevski, and Mario Collura. Growth of entanglement entropy under local projective measurements, 2021. arXiv:2109.10837.

[13] Alberto Biella and Marco Schiró. Many-Body Quantum Zeno Effect and Measurement-Induced Subradiance Transition. *Quantum*, 5:528, August 2021.

[14] Xhek Turkeshi, Alberto Biella, Rosario Fazio, Marcello Dalmonte, and Marco Schiró. Measurement-induced entanglement transitions in the quantum ising chain: From infinite to zero clicks. *Phys. Rev. B*, 103:224210, Jun 2021.

[15] Chao-Ming Jian, Yi-Zhuang You, Romain Vasseur, and Andreas W. W. Ludwig. Measurement-induced criticality in random quantum circuits. *Phys. Rev. B*, 101:104302, 3 2020.

[16] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):4213, Jul 2014.

[17] Wen Wei Ho and Timothy H. Hsieh. Efficient variational simulation of non-trivial quantum states. *SciPost Phys.*, 6:29, 2019.

[18] Chris Cade, Lana Mineh, Ashley Montanaro, and Stasja Stanisic. Strategies for solving the Fermi-Hubbard model on near-term quantum computers. *Phys. Rev. B*, 102:235122, Dec 2020.

[19] David Wierichs, Christian Gogolin, and Michael Kastoryano. Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer. *Phys. Rev. Research*, 2:043246, Nov 2020.

[20] Roeland Wiersema, Cunlu Zhou, Yvette de Sereville, Juan Felipe Carrasquilla, Yong Baek Kim, and Henry Yuen. Exploring Entanglement and Optimization within the Hamiltonian Variational Ansatz. *PRX Quantum*, 1:020319, 12 2020.

[21] Joris Kattemölle and Jasper van Wezel. Variational quantum eigensolver for the Heisenberg antiferromagnet on the kagome lattice, 2021. arXiv:2108.02175.

[22] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.

[23] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P. Lanyon, Peter Love, Ryan Babbush, Alán Aspuru-Guzik, Rainer Blatt, and Christian F. Roos. Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator. *Phys. Rev. X*, 8:031022, Jul 2018.

[24] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi. Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm. *Phys. Rev. X*, 8:011021, Feb 2018.

[25] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Edward Farhi, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Matthew P. Harrigan, Alan Ho, Sabrina Hong, Trent Huang, William J. Huggins, Lev Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Erik Lucero, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Hartmut Neven, Murphy Yuezhen Niu, Thomas E. O'Brien, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Doug Strain, Kevin J. Sung, Marco Szalay, Tyler Y. Takeshita, Amit Vainsencher, Theodore White, Nathan Wiebe, Z. Jamie Yao, Ping Yeh, and Adam Zalcman. Hartree-Fock on a superconducting qubit quantum computer. *Science*, 369(6507):1084–1089, 2020.

[26] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis. Scalable Quantum Simulation of Molecular Energies. *Phys. Rev. X*, 6:031007, Jul 2016.

[27] Hyungwon Kim and David A. Huse. Ballistic Spreading of Entanglement in a Diffusive Nonintegrable System. *Phys. Rev. Lett.*, 111:127205, Sep 2013.

[28] Adam Nahum, Jonathan Ruhman, Sagar Vijay, and Jeongwan Haah. Quantum Entanglement Growth under Random Unitary Dynamics. *Phys. Rev. X*, 7:031016, Jul 2017.

[29] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, Nov 2018.

[30] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1791, Mar 2021.

[31] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. Noise-Induced Barren Plateaus in Variational Quantum Algorithms, 2021. arXiv:2007.14384.

[32] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. Entanglement Induced Barren Plateaus, 2020. arXiv:2010.15968.

[33] Taylor L. Patti, Khadijeh Najafi, Xun Gao, and Susanne F. Yelin. Entanglement Devised Barren Plateau Mitigation, 2020. arXiv:2012.12658.

[34] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3:010313, Jan 2022.

[35] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Phys. Rev. A*, 92:042303, 2015.

[36] Harper R. Grimsley, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications*, 10(1):3007, Jul 2019.

[37] Tyler Volkoff and Patrick J Coles. Large gradients via correlation in random parameterized quantum circuits. *Quantum Science and Technology*, 6(2):025008, jan 2021.

[38] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, December 2019.

[39] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D. Lukin. Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices. *Phys. Rev. X*, 10:021067, Jun 2020.

[40] Andrea Skolik, Jarrod R. McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib. Layerwise learning for quantum neural networks, 2020. arXiv:2006.14904.

[41] Arthur Pesah, M. Cerezo, Samson Wang, Tyler Volkoff, Andrew T. Sornborger, and Patrick J. Coles. Absence of Barren Plateaus in Quantum Convolutional Neural Networks, 2020. arXiv:2011.02966.

[42] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. Quantum circuit learning. *Phys. Rev. A*, 98:032309, 2018.

[43] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A*, 99:032331, 2019.

[44] David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. General parameter-shift rules for quantum gradients, 2021. arXiv:2107.12390.

[45] Andrea Mari, Thomas R. Bromley, and Nathan Killoran. Estimating the gradient and higher-order derivatives on quantum hardware. *Phys. Rev. A*, 103:012405, Jan 2021.

[46] Artur F. Izmaylov, Robert A. Lang, and Tzu-Ching Yen. Analytic gradients in variational quantum algorithms: Algebraic extensions of the parameter-shift rule to general unitary transformations, 2021. arXiv:2107.08131.

[47] Oleksandr Kyriienko and Vincent E. Elfving. Generalized quantum circuit differentiation rules, 2021. arXiv:2108.01218.

[48] Johannes Jakob Meyer, Johannes Borregaard, and Jens Eisert. A variational toolbox for quantum multiparameter estimation. *npj Quantum Information*, 7(1):89, Jun 2021.

[49] Leonardo Banchi and Gavin E. Crooks. Measuring Analytic Gradients of General Quantum Evolution with the Stochastic Parameter Shift Rule. *Quantum*, 5:386, January 2021.

[50] Bálint Koczor and Simon C. Benjamin. Quantum natural gradient generalised to non-unitary circuits, 2019. arXiv:1912.08660.

[51] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum Natural Gradient. *Quantum*, 4:269, May 2020.

[52] Ryan R. Ferguson, Luca Dellantonio, Karl Jansen, Abdulrahim Al Balushi, Wolfgang Dür, and Christine A. Muschik. A measurement-based variational quantum eigensolver, 2020. arXiv:2010.13940.

[53] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. Van den Nest. Measurement-based quantum computation. *Nature Physics*, 5(1):19–26, Jan 2009.

[54] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71:1291–1294, 1993.

# Supplemental Materials

## FINITE-SCALING ANALYSIS AND DATA COLLAPSE

The correlation length $\xi$ of a system quantifies the length scale over which parts of a system are correlated. When a system undergoes a continuous phase transition, the correlation length diverges. Phase transitions only occur in the thermodynamic limit, and hence simulations of finite-sized systems will contain artifacts that have to be accounted for in order to capture the correct behavior [1]. In particular, for a finite system the correlation length $\xi$ cannot become infinite and is cut off at $L^d$, the maximum volume of a finite $d$-dimensional system. To account for this effect, we can perform a finite-scaling analysis.

The entanglement entropy as a function of measurement rate is conjectured to follow a volume law for $p < p_c$, a constant plus logarithmic correction at $p = p_c$ and area law for $p > p_c$ [2–4]. We can therefore construct a scaling form of the entanglement entropy as

$$S(N, p, \nu) = S(N, p_c, \nu) + f(N^{1/\nu}(p - p_c)) \tag{1}$$

where $S(N, p, \nu)$ denotes the von Neumann entropy at measurement rate $p$ and $f$ is a scaling function. The critical exponent $\nu$ determines the scaling of the entanglement entropy near $p_c$. If this scaling form is correct, we should be able to account for finite-size effects and all the data can be appropriately rescaled to match a single curve representing $f$ with a proper choice of $\nu$.

To determine the critical exponents, we fit a 5th-degree polynomial $g$ to our data using a Nelder-Mead optimization [5] and minimize the $\chi^2$-statistic

$$\chi^2 = \sum_i \frac{(S(N_i, p_i, \nu) - \tilde{S}(N_i, p_i, \nu))^2}{\Delta S}. \tag{2}$$

Here, $\tilde{S}(N_i, p_i, \nu)$ is estimated from the data and $S(N_i, p_i, \nu)$ is the proposed scaling form from eq. (1). $\Delta S$ is the standard deviation of the von Neumann entropies which arises due to the fluctuations induced by the randomized measurements and their outcomes. From the unscaled data, we determine a set of potential critical points $p_c$ and fit the above $\chi^2$-statistic to determine $\nu$. We then report the values of $p_c$ and $\nu$ that provided the best fit.

To verify the stability of the fit, we perform a statistical bootstrapping procedure to estimate the error bars on the fitted critical exponent $\nu$. We take $K_{\text{boot}} = 100$, where each data set consists of $K$ samples obtained by sampling from the entire data set of $3 \times 10^3$ data points with replacement. The final obtained error bars on $\nu$ are $\approx 0.01$.

We can extrapolate our result to the thermodynamic limit by fitting the data for $N' = N_{\max}/2$ to $N' = N_{\max}$ and plotting the resulting values for $\nu$ against $1/N'$ [2]. By doing a linear fit on the resulting data, we obtain

$$\tilde{\nu}(N') = a\frac{1}{N'} + b \tag{3}$$

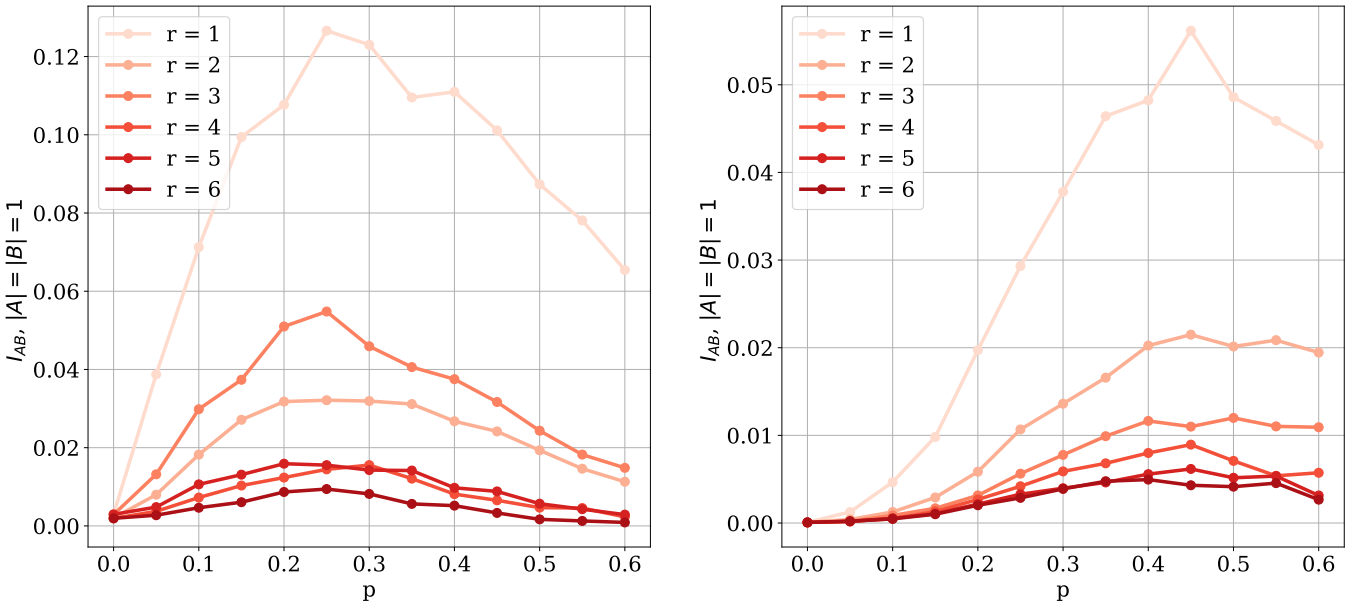
and so the intercept $b$ corresponds to the value of $\nu$ in the thermodynamic limit, since $\lim_{N' \to \infty} 1/N' = 0$. When fitting the data, we weigh the errors by the standard errors obtained in the statistical bootstrap described above.

## MUTUAL INFORMATION

The quantum mutual information can be used to quantify subsystem correlations, and subsequently detect phase transitions since we expect correlations to divergence at criticality [2, 3, 6]. As additional confirmation that the critical values $p_c$ estimated from the prior analysis are correct, we calculate the quantum mutual information as,

$$I(A,B) = S_A(N,p) + S_B(N,p) - S_{A\cup B}(N,p) \tag{4}$$

Here, we take the same approach as in [3], and take $A$ and $B$ to be two single qubit subsystems $|A| = 1$ and $|B| = 1$. We then vary the distance $r$ between qubit $A$ and $B$, to determine the effect of the distance on the subsystem correlations. In fig. 1, we observe two broad peaks around the previously found values $p_c \approx 0.25$ and $p_c \approx 0.5$ for the XXZ-HVA and HAA, respectively.




(a) XXZ-HVA

(b) HAE

FIG. 1. Quantum mutual information between two qubits $A$ and $B$ separated by a distance $r$ on a chain of length 16. The mutual information is averaged over $3\times10^3$ samples, where each sample corresponds to a random circuit realization, as described in the main text.

## PROJECTIVE GRADIENTS

Let $|\psi\rangle$ be a quantum state of an $n$-qubit system with corresponding density operator $|\psi\rangle\langle\psi| = \rho \in L(\mathbb{C}^{2^n})$. A projective measurement can transform the state as

$$\rho \mapsto \frac{\Pi\rho\Pi}{\mathrm{Tr}\{\Pi\rho\}}, \tag{5}$$

where $\Pi$ is a projector onto an eigenbasis of some Hermitian observable $O$ and is therefore Hermitian itself. Since $\Pi$ is a projector it satisfies $\Pi^2 = \Pi$. The normalization constant $\mathrm{Tr}\{\Pi\rho\}$ gives the overlap of the state $\rho$ with the basis onto which $\Pi$ projects the state. We can write a parameterized state as

$$|\psi(\theta_{1:k})\rangle = \overleftarrow{\prod_{k=1}} U(\theta_k)\,|\psi_0\rangle\,, \tag{6}$$

where we use the shorthand notation $\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_k)$ and $\overleftarrow{\prod}$ indicates that the product is ordered from right to left.

Given a set of $M$ intermediate measurements performed on the circuit, we are interested in minimizing the energy

$$E_{\mathrm{int}}(\boldsymbol{\theta}) = \sum_{i}^{2^M} p_M^{(i)}\,\mathrm{Tr}\Big\{\rho^{(i)}(\boldsymbol{\theta})H\Big\} \tag{7}$$

where $H = \sum_n O_n$ is the Hamiltonian. Here we consider projective measurements with 2 outcomes (the extension to more general measurements follows the same logic), hence $p^{(i)}$ is the probability of obtaining the bitstring outcomes $(i) = (i_1, \ldots, i_M) \in \{0,1\}^M$ for the intermediate measurements. The variational energy is thus the statistical mixture of all possible variational energies given the pure state $\rho^{(i)}(\boldsymbol{\theta})$ that results from the circuit and its intermediate measurement. Clearly $E_{\mathrm{ground}} \leq E_{\mathrm{int}}(\boldsymbol{\theta})$. Calculating the gradient of eq. (7) involves calculating the gradient for all individual states in the mixture and all terms $O_n$ in the Hamiltonian.

We construct the gradient update rule inductively by calculating the resulting states for 1, 2 and $M$ measurements applied in the circuit. Consider an initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$, to which we apply a unitary matrix $U(\theta_1)$ followed by a projective measurement $\Pi_1$,

$$\rho_1(\theta_1) = \frac{\Pi_1 U(\theta_1)\rho_0 U^\dagger(\theta_1)\Pi_1}{\mathrm{Tr}\{\Pi_1 U(\theta_1)\rho_0 U^\dagger(\theta_1)\Pi_1\}} \tag{8}$$

$$= \frac{\Pi_1 U(\theta_1)\rho_0 U^\dagger(\theta_1)\Pi_1}{p_1(\theta_1)}. \tag{9}$$

where we omitted the label $(i)$ for now and use the subscript to simply indicate the number of measurements. Next, we add an additional unitary and measurement,

$$\rho_2(\theta_1,\theta_2) = \frac{\Pi_2 U(\theta_2)\rho_1(\theta_1)U^\dagger(\theta_2)\Pi_2}{\mathrm{Tr}\{\Pi_2 U(\theta_2)\rho_1(\theta_1)U^\dagger(\theta_2)\Pi_2\}} \tag{10}$$

$$= \frac{\Pi_2 U(\theta_2)\Pi_1 U(\theta_1)\rho_0 U^\dagger(\theta_1)\Pi_1 U^\dagger(\theta_2)\Pi_2}{\mathrm{Tr}\{\Pi_2 U(\theta_2)\Pi_1 U(\theta_1)\rho_0 U^\dagger(\theta_1)\Pi_1 U^\dagger(\theta_2)\Pi_2\}} \tag{11}$$

$$= \frac{\Pi_2 U(\theta_2)\Pi_1 U(\theta_1)\rho_0 U^\dagger(\theta_1)\Pi_1 U^\dagger(\theta_2)\Pi_2}{p_2(\theta_1,\theta_2)}. \tag{12}$$

Note how the normalization constant of $\rho_1(\theta_1)$ cancels. Generalizing this to $M$ measurements, we get the general form

$$\rho_M(\theta_1,\ldots,\theta_M) = \left(\overleftarrow{\prod_{m=1}^{M}} \Pi_m U(\theta_m)\right)\rho_0\left(\overrightarrow{\prod_{m=1}^{M}} U^\dagger(\theta_m)\Pi_m\right) p_M^{-1}(\theta_1,\ldots,\theta_M) \tag{13}$$

$$= \tilde{\rho}_M(\theta_1,\ldots,\theta_M)p_M^{-1}(\theta_1,\ldots,\theta_M), \tag{14}$$

where

$$\tilde{\rho}_M(\theta_1,\ldots,\theta_M) = \left(\overleftarrow{\prod_{m=1}^{M}} \Pi_m U(\theta_m)\right)\rho_0\left(\overrightarrow{\prod_{m=1}^{M}} U^\dagger(\theta_m)\Pi_m\right), \tag{15}$$

$$p_M(\theta_1,\ldots,\theta_M) = \mathrm{Tr}\{\tilde{\rho_M}(\theta_1,\ldots,\theta_M)\}, \tag{16}$$

are the unnormalized state and its normalization constant, respectively. We will omit the arguments $(\theta_1, \ldots, \theta_M)$ for now to reduce notational clutter. We are interested in calculating gradients of a cost function that contains a sum of terms of the form $\mathrm{Tr}\{\rho_M O\}$ with respect to the parameter $\theta_l$,

$$\partial_{\theta_l} \mathrm{Tr}\{\rho_M O\} = \mathrm{Tr}\big\{(\partial_{\theta_l}\tilde{\rho}_M)\, p_M^{-1} O\big\} + \mathrm{Tr}\big\{\tilde{\rho}_M \left(\partial_{\theta_l} p_M^{-1}\right) O\big\}. \tag{17}$$

For the derivative of the unnormalized state, we get

$$\mathrm{Tr}\{(\partial_{\theta_l}\tilde{\rho}_M)\, O\} = \langle\psi_0| \left(\overrightarrow{\prod_{m=1}^{M}} U^\dagger(\theta_m)\Pi_m\right) O \left(\overleftarrow{\prod_{m=l+1}^{M}} \Pi_m U(\theta_m)\right) \Pi_l \partial_{\theta_l} U(\theta_l) \left(\overleftarrow{\prod_{m=1}^{l-1}} \Pi_m U(\theta_m)\right) |\psi_0\rangle \tag{18}$$

$$+ \langle\psi_0| \left(\overrightarrow{\prod_{m=1}^{l-1}} U^\dagger(\theta_m)\Pi_m\right) \partial_{\theta_l} U^\dagger(\theta_l)\Pi_l \left(\overrightarrow{\prod_{m=l+1}^{M}} U^\dagger(\theta_m)\Pi_m\right) O \left(\overleftarrow{\prod_{m=1}^{M}} \Pi_m U(\theta_m)\right) |\psi_0\rangle \tag{19}$$

$$= \left\langle\tilde{\psi}_0\right| U^\dagger(\theta_l)\tilde{O}\partial_{\theta_l} U(\theta_l) \left|\tilde{\psi}_0\right\rangle + \left\langle\tilde{\psi}_0\right| \partial_{\theta_l} U^\dagger(\theta_l)\tilde{O} U(\theta_l) \left|\tilde{\psi}_0\right\rangle, \tag{20}$$

where

$$\left|\tilde{\psi}_0\right\rangle = \left(\overleftarrow{\prod_{m=1}^{l-1}} \Pi_m U(\theta_m)\right) |\psi_0\rangle\,, \tag{21}$$

$$\tilde{O} = \left(\overrightarrow{\prod_{m=l+1}^{M}} U^\dagger(\theta_m)\Pi_m\right) O \left(\overleftarrow{\prod_{m=l+1}^{M}} \Pi_m U(\theta_m)\right). \tag{22}$$

If $U(\theta_l)$ is generated by a Pauli operator $A$, then $\partial_{\theta_l} U(\theta_l) = -\frac{i}{2} A U(\theta_l)$ and so we can use the parameter-shift rule [7, 8]

$$-\frac{i}{2}\left\langle\tilde{\psi}_0\right| \left[A, U^\dagger(\theta_l)\tilde{O}U(\theta_l)\right] \left|\tilde{\psi}_0\right\rangle = \frac{1}{2}\Big( \left\langle\tilde{\psi}_0\right| U^\dagger(\theta_l + \tfrac{\pi}{2})\tilde{O}U(\theta_l + \tfrac{\pi}{2}) - U^\dagger(\theta_l - \tfrac{\pi}{2})\tilde{O}U(\theta_l - \tfrac{\pi}{2}) \left|\tilde{\psi}_0\right\rangle \Big) \tag{23}$$

If the state would be properly normalized, then this would provide a strategy for measuring this gradient. However, the final result

$$\mathrm{Tr}\big\{(\partial_{\theta_l}\tilde{\rho}_M)\, p_M^{-1} O\big\} = \frac{1}{2}\Big( \left\langle\tilde{\psi}_0\right| U^\dagger(\theta_l + \tfrac{\pi}{2})\tilde{O}U(\theta_l + \tfrac{\pi}{2}) - U^\dagger(\theta_l - \tfrac{\pi}{2})\tilde{O}U(\theta_l - \tfrac{\pi}{2}) \left|\tilde{\psi}_0\right\rangle \Big) p_M^{-1}, \tag{24}$$

calculates an observable with respect to an unnormalized state, because $p_M^{-1}$ normalizes the state without the parameter-shifted gates. Hence, we need to first normalize the state in order to be able to execute the gradient calculation on the device. We can achieve this by multiplying with the identity

$$\mathrm{Tr}\{(\partial_{\theta_l}\tilde{\rho}_M)\, O\} = \frac{1}{2}\left\langle\tilde{\psi}_0\right| \Bigg( U^\dagger(\theta_l + \tfrac{\pi}{2})\tilde{O}U(\theta_l + \tfrac{\pi}{2}) \times \frac{p_M^{+,l}(\theta_1, \ldots, \theta_l + \frac{\pi}{2}, \ldots, \theta_M)}{p_M^{+,l}(\theta_1, \ldots, \theta_l + \frac{\pi}{2}, \ldots, \theta_M)} \tag{25}$$

$$- U^\dagger(\theta_l - \tfrac{\pi}{2})\tilde{O}U(\theta_l - \tfrac{\pi}{2}) \times \frac{p_M^{-,l}(\theta_1, \ldots, \theta_l - \frac{\pi}{2}, \ldots, \theta_M)}{p_M^{-,l}(\theta_1, \ldots, \theta_l - \frac{\pi}{2}, \ldots, \theta_M)} \Bigg) \left|\tilde{\psi}_0\right\rangle p_M^{-1} \tag{26}$$

$$= \frac{1}{2}\left( \langle O\rangle^+ \frac{p_M^{+,l}}{p_M} - \langle O\rangle^- \frac{p_M^{-,l}}{p_M} \right). \tag{27}$$

Here, $\langle O\rangle^\pm$ is the expectation value of the observable $O$ after the measurements $\{\Pi_1, \ldots, \Pi_M\}$ have been applied to the parameter-shifted circuit.

For the gradient of the inverse of the normalization constant, we get

$$\mathrm{Tr}\left\{\tilde{\rho}_M \left(\partial_{\theta_l} p_M^{-1}\right) O\right\} = -\langle O\rangle\, p_M^{-1} \partial_{\theta_l} p_M, \tag{28}$$

where we used the normalization constant to define $\langle O\rangle$, the expectation value of $O$ with respect to the measured circuit. The final step is to calculate $\partial_{\theta_l} p_M$:

$$\partial_{\theta_l} p_M = \mathrm{Tr}\{\partial_{\theta_l} \tilde{\rho_M}(\theta_1,\ldots,\theta_M)\} \tag{29}$$

$$= \langle\psi_0| \left(\prod_{m=1}^{\overrightarrow{M-1}} U^\dagger(\theta_m)\Pi_m\right) U^\dagger(\theta_M)\Pi_M U(\theta_M) \left(\prod_{m=l+1}^{\overleftarrow{M-1}} \Pi_m U(\theta_m)\right) \Pi_l \partial_{\theta_l} U(\theta_l) \left(\prod_{m=1}^{\overleftarrow{l-1}} \Pi_m U(\theta_m)\right) |\psi_0\rangle \tag{30}$$

$$+ \langle\psi_0| \left(\prod_{m=1}^{\overrightarrow{l-1}} U^\dagger(\theta_m)\Pi_m\right) \partial_{\theta_l} U^\dagger(\theta_l)\Pi_l \left(\prod_{m=l+1}^{\overrightarrow{M-1}} U^\dagger(\theta_m)\Pi_m\right) U^\dagger(\theta_M)\Pi_M U(\theta_M) \left(\prod_{m=1}^{\overleftarrow{M-1}} \Pi_m U(\theta_m)\right) |\psi_0\rangle \tag{31}$$

$$= \left\langle\tilde{\psi_0}\right| U^\dagger(\theta_l)\tilde{\Pi}_M \partial_{\theta_l} U(\theta_l) \left|\tilde{\psi_0}\right\rangle + \left\langle\tilde{\psi_0}\right| \partial_{\theta_l} U^\dagger(\theta_l)\tilde{\Pi}_M U(\theta_l) \left|\tilde{\psi_0}\right\rangle \tag{32}$$

where

$$\tilde{\Pi}_M = \left(\prod_{m=l+1}^{\overrightarrow{M-1}} U^\dagger(\theta_m)\Pi_m\right) U^\dagger(\theta_M)\Pi_M U(\theta_M) \left(\prod_{m=l+1}^{\overleftarrow{M-1}} \Pi_m U(\theta_m)\right), \tag{33}$$

and $\left|\tilde{\psi_0}\right\rangle$ is the same as in eq. (21). Again, we can apply the parameter-shift rule to obtain

$$\partial_{\theta_l} p_M = \frac{1}{2}\left( \left\langle\tilde{\psi_0}\right| U^\dagger(\theta_l + \frac{\pi}{2})\tilde{\Pi}_M U(\theta_l + \frac{\pi}{2}) - U^\dagger(\theta_l - \frac{\pi}{2})\tilde{\Pi}_M U(\theta_l - \frac{\pi}{2}) \left|\tilde{\psi_0}\right\rangle \right). \tag{34}$$

But these expectation values are simply the normalization constants $p_M^{\pm,l}$, hence the final result becomes

$$\mathrm{Tr}\left\{\tilde{\rho}_M \left(\partial_{\theta_l} p_M^{-1}\right) O\right\} = \langle O\rangle \frac{1}{2}\left(\frac{p_M^{+,l}}{p_M} - \frac{p_M^{-,l}}{p_M}\right). \tag{35}$$

All together, the final projective gradient is then

$$\partial_{\theta_l} \mathrm{Tr}\{\rho_M O\} = \frac{1}{2}\left(\left(\langle O\rangle^+ + \langle O\rangle\right)\frac{p_M^{+,l}}{p_M} - \left(\langle O\rangle^- + \langle O\rangle\right)\frac{p_M^{-,l}}{p_M}\right). \tag{36}$$

We see that if all projectors are the identity projector, $\frac{p_M^{-,l}}{p_M} = 1$, so we get back the old parameter shift rule

$$\partial_{\theta_l} \mathrm{Tr}\{\rho_M O\} = \frac{1}{2}\left(\langle O\rangle^+ - \langle O\rangle^-\right). \tag{37}$$

We now reintroduce the label $i = (i_1,\ldots,i_M)$ to indicate the specific measurement outcome. The probability of the measurement $(i)$ occurring is given by the probability $p_M^{(i)}(\theta_1,\ldots,\theta_M)$. The total mixed state as a result of the measurements is then

$$\rho(\theta_1,\ldots,\theta_M) = \sum_i p_M^{(i)}(\theta_1,\ldots,\theta_M)\rho_M^{(i)}(\theta_1,\ldots,\theta_M) \tag{38}$$

$$= \sum_i \tilde{\rho}_M^{(i)}(\theta_1,\ldots,\theta_M) \tag{39}$$

From eq. (27) we see that the gradient of the mixed state is then simply

$$\mathrm{Tr}\{(\partial_{\theta_l}\rho)O\} = \sum_i^{2^M} \frac{1}{2}\left(\langle O\rangle^{(i),+} \frac{p_M^{(i),+,l}}{p_M^{(i)}} - \langle O\rangle^{(i),-} \frac{p_M^{(i),-,l}}{p_M^{(i)}}\right) \tag{40}$$

Hence the estimator for the gradient corresponds to the average expectation value over intermediate measurements done on parameter-shifted circuits weighted by $p_M^{(i),l}$. Therefore, the projective gradients can be estimated by obtaining

statistics from the measurements done on the vanilla and parameter-shifted circuits. These statistics are obtained when estimating $\langle O \rangle$ and $\langle O \rangle^{\pm}$, respectively. To calculate the projective gradients, we use the TensorFlow-based quantum simulator `Zyglrox` [9].


[1] M. E. J. Newman and G. T. Barkema. *Monte Carlo methods in statistical physics*. Clarendon Press, Oxford, 1999.

[2] Brian Skinner, Jonathan Ruhman, and Adam Nahum. Measurement-Induced Phase Transitions in the Dynamics of Entanglement. *Phys. Rev. X*, 9:031009, 07 2019.

[3] Yaodong Li, Xiao Chen, and Matthew P. A. Fisher. Measurement-driven entanglement transition in hybrid quantum circuits. *Phys. Rev. B*, 100:134306, 10 2019.

[4] Yimu Bao, Soonwon Choi, and Ehud Altman. Theory of the phase transition in random unitary circuits with measurements. *Phys. Rev. B*, 101:104301, 3 2020.

[5] J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 01 1965.

[6] Xiangyu Cao, Antoine Tilloy, and Andrea De Luca. Entanglement in a fermion chain under continuous monitoring. *SciPost Phys.*, 7:24, Aug 2019.

[7] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. Quantum circuit learning. *Phys. Rev. A*, 98:032309, 2018.

[8] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A*, 99:032331, 2019.

[9] R.C. Wiersema. Zyglrox. https://github.com/therooler/zyglrox, 2021.