



CHORUS

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

Practical Verification of Quantum Properties in Quantum-Approximate-Optimization Runs

M. Sohaib Alam, Filip A. Wudarski, Matthew J. Reagor, James Sud, Shon Grabbe, Zhihui Wang, Mark Hodson, P. Aaron Lott, Eleanor G. Rieffel, and Davide Venturelli

Phys. Rev. Applied **17**, 024026 — Published 9 February 2022

DOI: [10.1103/PhysRevApplied.17.024026](https://doi.org/10.1103/PhysRevApplied.17.024026)

Practical Verification of Quantum Properties in Quantum Approximate Optimization Runs

M. Sohaib Alam,^{1,2,3,*} Filip A. Wudarski,^{2,3,†} Matthew J. Reagor,^{1,‡} James Sud,^{2,3} Shon Grabbe,² Zhihui Wang,^{2,3} Mark Hodson,¹ P. Aaron Lott,^{2,3} Eleanor G. Rieffel,² and Davide Venturelli^{2,3,§}

¹*Rigetti Computing, Berkeley, CA, 94701, USA*

²*Quantum Artificial Intelligence Laboratory (QuAIL),*

NASA Ames Research Center, Moffett Field, CA, 94035, USA

³*USRA Research Institute for Advanced Computer Science (RIACS), Mountain View, CA, 94043, USA*

In order to assess whether quantum resources can provide an advantage over classical computation, it is necessary to characterize and benchmark the non-classical properties of quantum algorithms in a practical manner. In this paper, we show that using measurements in no more than 3 out of the possible 3^N bases, one can not only reconstruct the single-qubit reduced density matrices and measure the ability to create coherent superpositions, but also possibly verify entanglement across all N qubits participating in the algorithm. We introduce a family of generalized Bell-type observables for which we establish an upper bound to the expectation values in fully separable states by proving a generalization of the Cauchy-Schwarz inequality, which may serve of independent interest. We demonstrate that a subset of such observables can serve as entanglement witnesses for QAOA-MaxCut states, and further argue that they are especially well tailored for this purpose by defining and computing an *entanglement potency* metric on witnesses. A subset of these observables also certify, in a weaker sense, the entanglement in GHZ states, which share the \mathbb{Z}_2 symmetry of QAOA-MaxCut. The construction of such witnesses follows directly from the cost Hamiltonian to be optimized, and not through the standard technique of using the projector of the state being certified. It may thus provide insights to construct similar witnesses for other variational algorithms prevalent in the NISQ era. We demonstrate our ideas with proof-of-concept experiments on the Rigetti Aspen-9 chip for ansätze containing up to 24 qubits.

I. INTRODUCTION

At present, there is great interest in designing and validating the implementation of solvers of computational problems that can achieve a *quantum advantage*, i.e. superior performance with respect to other known classical methods. There is considerable research activity in the study of speedup metrics [1] and in the certification and quantification of quantum properties (which we later also refer to as *quantumness*) in the context of variational quantum algorithms [2–4]. It is however of utmost importance in the NISQ era to understand the role of coherence and entanglement, as well as to guarantee that the software-hardware systems that we use for the empirical exploration exploit these resources to outperform classical competitors. This task is distinct from evaluating the fidelity of the experimental results with respect to a simulation, and instead focuses on verifying the key non-classical properties of a solver that may achieve quantum advantage on noisy hardware whose exact mechanics may not be easily described through analytical expressions or even numerical simulations.

Bearing in mind the lively debate over approaches to verify the presence of quantum properties in experimental runs on quantum annealers [5], as well as the recent

literature on quantum volume [6, 7] and on entanglement characterization and detection in NISQ devices [8, 9], in this paper we take a pragmatic approach trying to answer the question of whether any quantum resources survive following the execution of a specific quantum algorithm, applied to a specific problem, run on a specific quantum processing device. Our proposed tests do not necessarily guarantee that the detected quantumness was exploited computationally prior to detection, however the analysis can be coupled with complementary simulations that can evaluate the computational value of the measured resource. While our algorithm of choice is the quantum approximate optimization algorithm [10] (QAOA), applied to the MaxCut problem, the operational character of the study will illuminate what needs to be done in practice for other choices of algorithms, problems and devices.

The QAOA for MaxCut problem (QAOA-MaxCut) can be regarded as the minimization of a 2-local N -qubit cost Hamiltonian $C = \sum_{\langle i,j \rangle} Z_i Z_j$, defined over some set of edges $\langle i, j \rangle$ in a circuit ansatz given by

$$|\psi_{QAOA}\rangle = \left(\prod_{j=1}^p e^{-i\beta_j B} e^{-i\gamma_j C} \right) |+\rangle^{\otimes N}, \quad (1)$$

where $B = \sum_{i=1}^N X_i$ is the *mixer*, and the ansatz parameters $\beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_p$ are to be optimized to minimize the expectation value of C in the state $|\psi_{QAOA}\rangle$. The quantum device used is the Rigetti Aspen-9 system [11, 12], comprising 31 operational transmon qubits with median two-qubit gate fidelity of 97%. Figure 1 provides an overview of our overall approach.

* malam@usra.edu

† filip.a.wudarski@nasa.gov

‡ matt@rigetti.com

§ dventurelli@usra.edu

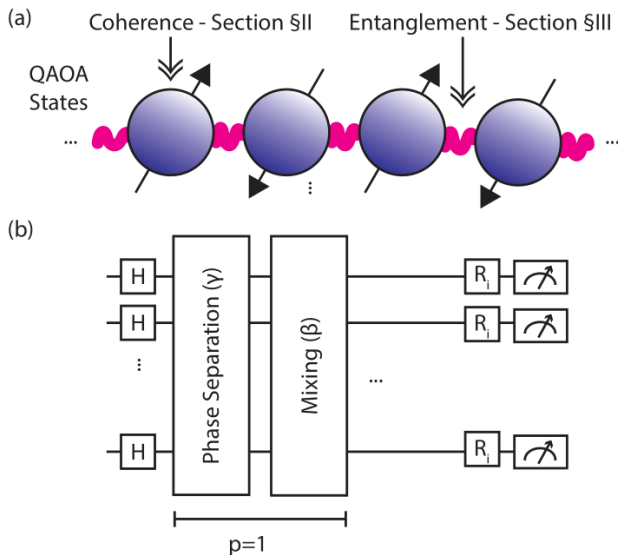


FIG. 1. Practical Verification of Quantum Properties in QAOA: (a) QAOA leverages quantum coherent superposition and entanglement to perform combinatorial optimization. The goal of this work is to develop probes of Coherence (Sect. II) and Entanglement (Sect. III) in states generated by the QAOA algorithm. (b) The specific ansatz studied here consists of an initial register state preparation via transversal Hadamard gates, followed by p applications of phase separation and mixing Hamiltonians. To reveal coherence and entanglement properties, additional single-qubit gates rotate the measurement basis before reading out the state of the register. We analyze experimental results for linear chain problems up to twenty-four qubits and a single round ($p = 1$) of QAOA.

The \mathbb{Z}_2 symmetry of the MaxCut cost Hamiltonian ensures that QAOA-MaxCut states prepared via Eq. (1) inherit the symmetry so that the amplitude of some computational basis state is the same as that of its ones' complement, i.e. $\langle x | \psi_{QAOA} \rangle = \langle \bar{x} | \psi_{QAOA} \rangle$, where $|\bar{x}\rangle = X^{\otimes N} |x\rangle$. Although non-entangled states can exhibit such a symmetry as well, Eq. (1) with a 2-local cost Hamiltonian C ensures that the state $|\psi_{QAOA}\rangle$ can possess entanglement at every value of p . In the limit of very large p and assuming a single MaxCut solution, QAOA would yield a solution of the form $\frac{1}{\sqrt{2}}(|x\rangle + |\bar{x}\rangle)$, where both the bitstrings x and \bar{x} specify the same 'cut' through the set of edges, and just differ in how they label each partition. This kind of entanglement is strongly reminiscent of the GHZ state, and in Section III the similarity between the witnesses constructed for GHZ states as well as QAOA-MaxCut states becomes obvious. Interestingly, we observe that there is a correlation between the value of p for which QAOA-MaxCut saturates the maximum expectation value of the entanglement witness, and the value of p for which it saturates the maximum expectation value of the cost Hamiltonian, suggesting

that the entanglement certified by these observables is actively participating in the computational optimization. We note that earlier works [13, 14] have considered the relationship of the underlying \mathbb{Z}_2 symmetry of the cost Hamiltonian and QAOA performance, and hope that our present work also sheds light on how entanglement is employed as a resource in QAOA.

By probing single-qubit reduced density matrices (SQRDMs), we are able to identify which qubits in the chip participate in the computation with the ability to create coherent superpositions of the classical bit values 0 and 1. Moreover, using the same measurement data as required to construct the SQRDMs, we are able to infer the expectation values of observables that serve as entanglement witnesses for some quantum state(s). The ease with which these witnesses can be measured make them a practical toolkit for a quantum programmer to quickly verify if any entanglement has been generated in the executed circuit. Although such a procedure could work for arbitrary circuits in principle, we demonstrate that the introduced family of witnesses is particularly well suited to detecting entanglement in QAOA circuits. We do so by explicitly showing that such observables do indeed serve as witnesses for QAOA-MaxCut states, and by introducing and computing an *entanglement potency* metric on witnesses, which quantifies the fractional volume of states whose entanglement is detected by the given witness. We compute this metric for QAOA and Haar random states.

A standard method to construct an entanglement witness [15] for some state $|\Psi\rangle$ is to construct an operator out of a projector

$$W = \alpha \mathbb{I} - |\Psi\rangle \langle \Psi| \quad (2)$$

where α is the maximum fidelity of the state $|\Psi\rangle$ with those in some set of states that we wish to certify against. However, for a parametric family of states such as QAOA-MaxCut states, the construction in Eq. (2) not only depends on the choice of parameters, but may also require measurements in a large number of different bases as well as terms to estimate the expectation values of. In previous work, it has been shown that for certain N -qubit pure states, there exists a witness that requires $2N - 1$ measurements [16, 17]. Here, we construct observables that require no more than 3 bases measurements for any value of N , and at most a polynomial number of terms to be estimated given the measurement data. While the theorems we prove only establish that bounds by fully separable states can be violated, we also provide numerical evidence that a subset of such observables can certify against bi-separable (and by extension k -separable, for $k \geq 2$) states, and therefore witness genuine N -partite entanglement.

The remainder of this article is organized as follows. In Section II, we discuss the measurement of coherences at the single-qubit level, and their relevance as a measure of quantumness. We also present experimental results from the Rigetti Aspen-9 quantum device. Section III

contains the main focus of our paper, wherein we discuss the construction of a family of entanglement witnesses, and establish several of their properties using a mix of analytical and numerical results. In particular, we analytically prove an upper bound to the absolute value of the expectation of a large class of observables in fully separable states using a generalized Cauchy-Schwarz inequality, which we also prove as a lemma, as well as a lower bound to the maximum eigenvalue for a subset of such observables, thus proving that they can serve as witnesses for some entangled states. We also identify a particular class of observables that are especially relevant for QAOA-MaxCut, and prove that they serve as witnesses for such states prepared via the ring cost Hamiltonian. In addition, we also provide numerical evidence for the view that similar observables constructed for other cost Hamiltonians also serve as witnesses for the corresponding QAOA-MaxCut states. We also demonstrate the relevance of the constructed observables to QAOA-MaxCut through the calculation of an entanglement potency metric, which we introduce. We finish this section by presenting proof-of-concept experimental results from the Rigetti Aspen-9 quantum chip. We conclude in Section IV, discussing future avenues of exploration that our work opens up.

II. QUANTUM COHERENCE

The most iconic property of a quantum system is the ability to be in a *coherent superposition* of multiple reference states. This basic distinctive property of quantum information processing and its direct manifestations (e.g. tunneling) is believed to be essential to achieving any kind of quantum advantage [18]. Formally, a system is said to be in coherent superposition with respect to a basis of states, if its density matrix is non-diagonal when expressed in that basis [19]. In combinatorial optimization, the specification of both the problem variables as well as the solution, or an algorithm’s approximation to it, are efficiently representable with a polynomial number of classical bits. A quantum algorithm for combinatorial optimization might in general exploit the coherent superposition of all 2^N computational basis states. However, probing these coherences require a prohibitively exponential experimental cost.

Without requiring a full tomographic reconstruction of the many-body density matrix of a quantum state, we propose to use a method which allows us to extract information about off-diagonal elements of all single-qubit reduced density matrices (SQRDMs) expressed in the computational basis, which we later use as a threshold figure of merit to determine the extent of quantumness. If sizeable values of these elements are measured at any time during the execution of an algorithm, it can be concluded that the individual qubits at that time still hold the potential to exploit single-particle quantum effects for the purpose of information processing. A generalization of

the arguments to multi-particle coherent effects (i.e. co-tunneling) would be straightforward in principle, though with an exponentially greater experimental cost.

A. Efficient SQRDM Tomography

Suppose that we have a single qubit quantum state, either pure or mixed, represented by the density matrix

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}, \quad (3)$$

for which we identify the off-diagonal elements $\rho_{12} = \rho_{21}^*$ as the “coherences”. Here, we choose to use the absolute value of coherence, i.e. $C_\rho = |\rho_{12}|$, as our metric of interest. Since $\rho \geq 0$ and $\text{Tr}(\rho) = 1$, the maximum possible value attainable for C_ρ is $\frac{1}{2}$, and is saturated for example by the $|+\rangle$ state. Each single qubit state can be experimentally reconstructed via state tomography [20–22] using measurements in the Pauli X, Y and Z bases, which allow us to determine expectation values of Pauli matrices from the outcome statistics and represent the state as

$$\rho = \frac{1}{2}(\mathbb{I} + \langle X \rangle X + \langle Y \rangle Y + \langle Z \rangle Z) \quad (4)$$

One could argue that only two measurements suffice to measure $C_\rho = \frac{1}{2}|\langle X \rangle + i\langle Y \rangle|$. However, since the measurement protocol is inevitably noisy, it is more reasonable to first tomographically reconstruct the state using maximum likelihood estimation (MLE) [23], which would guarantee that the measured state is a legitimate quantum object, i.e. being normalized $\text{Tr}(\rho) = 1$, and positive semi-definite $\rho \geq 0$, and then compute the required expectation values in this state to measure C_ρ .

The above protocol easily generalizes [24] - without additional measurement overhead - to determine the SQRDMs for an N -qubit state produced by a quantum algorithm, such as QAOA. In order to achieve this, we measure all (algorithmically relevant) qubits simultaneously in the X, Y and Z bases and determine ρ_k as in Eq. (4) where now $\langle Q_k \rangle \equiv \langle \mathbb{I}_1 \otimes \dots \otimes Q_k \otimes \dots \rangle$ denotes the expectation value of $Q = X, Y, Z$ acting on the k th qubit. In this, way we partially reconstruct the final state after the measurements and the MLE procedure as

$$\rho_{\text{single-particle}} = \bigotimes_{k=1}^N \rho_k,$$

with all ρ_k being legitimate states. It is important to note that $\rho_{\text{single-particle}}$ is not a reconstruction of the density matrix of the state of the quantum processor before the measurements. Most notably, it is a product state, and carries no information about other quantum features such as entanglement [25] or discord [26], which have been washed out in the course of local measurements. Nevertheless, the detection of a non-zero coherence in a ρ_k

would indicate that superposition was still possible for qubit k before measurement, so that some single-particle quantumness was exploitable. A figure of merit for this kind of limited quantumness could be identified in the “maximum observed single-particle coherence”, i.e. the largest measured value of C_{ρ_k} over both the qubits and the parameters that define the circuit.

B. SQRDMs in QAOA for MaxCut

Now let us demonstrate how the above protocol works for a QAOA algorithm. Let d denote the degree of the qubit (node) in the underlying MaxCut problem graph. The QAOA ansatz is $|\psi(\gamma, \beta)\rangle = e^{-i\beta B} e^{-i\gamma C} |+\rangle^{\otimes N}$ with the mixer $B = \sum_{j=1}^N X_j$ and the cost function $C = \sum_{\langle ij \rangle} Z_i Z_j$. Then, for a single layer ($p = 1$) the SQRDMs can be analytically calculated to be

$$\rho_{QAOA} = \frac{1}{2} \begin{pmatrix} 1 & \cos^d(2\gamma) \\ \cos^d(2\gamma) & 1 \end{pmatrix}, \quad (5)$$

Note the independence from β and the total number of qubits N . Eq. (5) tells us that $C_{\rho_{QAOA}} = \frac{1}{2} |\text{Tr}(\rho_{QAOA}(X + iY))| = \frac{1}{2} |\cos^d(2\gamma)|$. Moreover, the fidelity [27] of this state with an arbitrary single-qubit state of the form Eq. (4) is found to be

$$F(\rho_{QAOA}, \rho) = \frac{1}{2} (1 + \langle X \rangle \cos^d 2\gamma) + \frac{1}{2} (1 - \cos^2 2\gamma)^{1/2} \cdot (1 - |\vec{P}|^2)^{1/2}, \quad (6)$$

where $\vec{P} = (\langle X \rangle, \langle Y \rangle, \langle Z \rangle)$. Note that this fidelity explicitly depends on γ . Of course, when $\vec{P} = (\cos^d 2\gamma, 0, 0)$, the above expression equals one for all γ .

We can similarly compute the fidelity between the ideal QAOA state in Eq. (5) and a classical probabilistic bit

$$\rho_{cl}(\theta) = \begin{pmatrix} \cos^2 \theta & 0 \\ 0 & \sin^2 \theta \end{pmatrix}, \quad (7)$$

In particular, we consider the maximum fidelity over all such probabilistic bits (max over θ), given by

$$\max_{\theta} F(\rho_{QAOA}, \rho_{cl}) = \frac{1}{2} + \frac{1}{2} \sqrt{1 - \cos^2 2\gamma}. \quad (8)$$

States for which the difference between Eqs. (6) and (8) is greater than zero represent resources that cannot be reproduced or employed by any classical device. We can therefore use this difference as a measure of whether quantum resources have been employed by a physical quantum device when running QAOA.

C. Results - Coherences

We use the methods described above to report on the amount of coherences each single qubit can build during

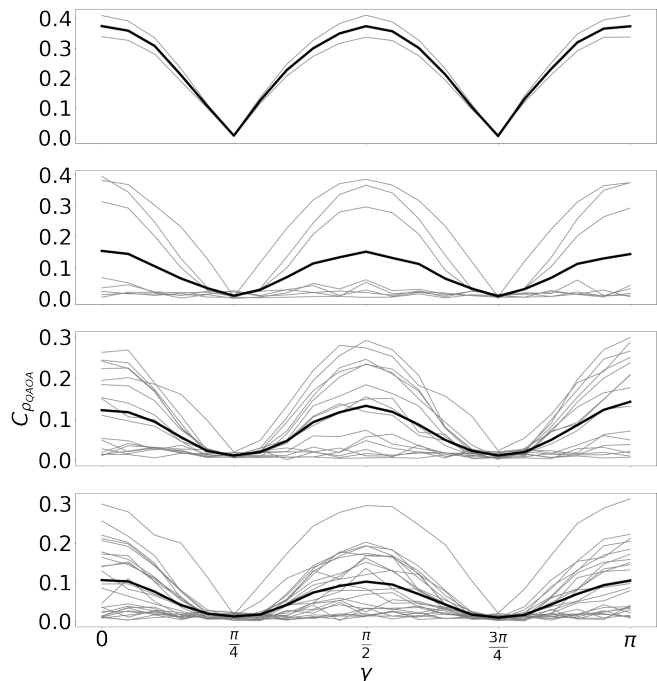


FIG. 2. Coherences $C_{\rho_{QAOA}}$ as a function of γ for a fixed value of $\beta = \frac{\pi}{8}$ for qubits (thin, gray lines) in a linear chain QAOA-MaxCut circuit of (top to bottom) 2, 8, 16 and 24 qubits. The thick (black) line represents the mean coherence across all the qubits.

the circuit execution for a fixed $\beta = \frac{\pi}{8}$ angle [28] and across different γ angles. Fig. 2 shows the coherences $C_{\rho_{QAOA}}$ for each SQRDM in 2, 8, 16 and 24-qubit linear chain QAOA-MaxCut circuits. In the ideal noiseless case, these should follow precisely the form $\frac{1}{2} |\cos^d 2\gamma|$ where d is the degree of the qubit whose coherence is being plotted. Experimentally, we observe a qualitatively similar pattern, except scaled down from its maximum value of $\frac{1}{2}$.

In Fig. 3, we compute the fidelity of the experimental SQRDM of a qubit (No. 0 - see Appendix A for the chip’s layout) after having ran a 24-qubit linear chain QAOA-MaxCut circuit with the ideal SQRDM for that qubit, calculated using Eq. (6). As a comparison, we also plot the maximum fidelity achievable by a classical probabilistic bit. The difference between these two plots provides a measure of the non-classical coherent superposition effects that this qubit can experience during the course of a subsequent quantum computation. If this difference is zero, then at the time of measurement, the qubit offered no more computational power than that offered by a probabilistic classical bit. Fig. 12 in Appendix A provides more information on the coherences in each of the qubits in the executed circuits, as well as their physical locations on the Aspen-9 chip.

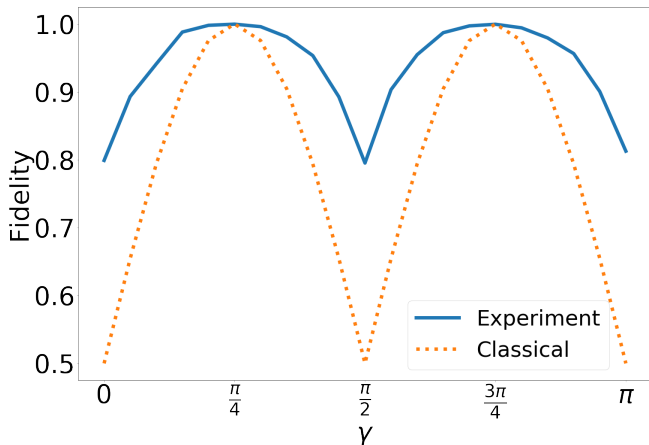


FIG. 3. Experimental fidelity with the ideal SQRDM for a qubit (No. 0 - see Appendix A for chip’s layout) in a 24 linear chain QAOA-MaxCut circuit, compared with the maximum achievable fidelity by a classical probabilistic bit, or a fully decohered qubit.

III. ENTANGLEMENT

Quantum entanglement [25] is a property of a physical system that relates to the intrinsic non-decomposable character of its state. It is famously manifested as a type of correlation that distinguishes classical physics, which is local and deterministic, from quantum mechanics. While it is still debated to what extent non-local correlations are necessary for quantum advantage [29, 30], it is certainly a signature feature of the QAOA protocol and of its relationship with adiabatic quantum computation.

Methods to estimate, quantify, and detect entanglement are highly valuable assets for benchmarking quantum hardware. Measures such as bipartite and multipartite concurrence [31–33], entanglement of formation [32], von Neumann entropy and many more exist [25, 34, 35]. However they are typically difficult to compute and experimentally costly to estimate.

In this work, we construct a family of observables that act as *entanglement witnesses* [36, 37] for some quantum state(s). These are observables whose expectation value is bounded for separable states, so that a violation of such a bound provides a signature of entanglement. Although we focus our attention on QAOA-MaxCut states, the observables we introduce below may also serve as witnesses for other types of states, such as parametric families of states found in variational quantum algorithms, e.g. variational quantum eigensolver (VQE) [38, 39], quantum machine learning [40, 41], or variational Hamiltonian ansatz [3, 42]. From the experimentalist’s perspective, it is a simple matter to run a circuit and obtain measurements in the X , Y and Z bases, and then to simply check whether an observable constructed as described below violates a bound (see Theorem 1). If it does, the experimentalist has verified the presence of entanglement in the circuit, regardless of whether that circuit runs QAOA

or some completely different algorithm.[43] In this sense, the methods we describe below can be perceived as algorithm agnostic, and can be considered to be a general purpose benchmarking tool that can be easily and efficiently deployed for an arbitrary algorithm and system size. Additionally, this method can also be deployed to noisy circuits that are prevalent in the NISQ era, and is not just restricted to detecting entanglement in pure states.

In this section, we first describe general properties of a family of observables and show that they can serve as entanglement witnesses [36] for some quantum states. We then show that a subset of such observables are capable of certifying entanglement in QAOA-MaxCut states. Additionally, we discuss the separability properties of these witnesses. Lastly, we introduce an *entanglement potency* metric on a witness that quantifies the relevance of a witness to a family of states. We conclude the section with experimental results from the Aspen-9 quantum chip.

A. Generalized Bell-type Observables

Here, we introduce observables that are sums of k -local Pauli operators and establish bounds on the expectation values of these observables achievable by fully separable and general quantum states. Our aim is to demonstrate that a large family of such observables can serve as entanglement witnesses up to a spectral shift [36, 37]. Inspired by both Bell-type inequalities [44–46], we introduce a class of observables W with the property

$$\forall_{\rho_{\text{sep}}} \quad \Lambda_{\min} \leq \text{Tr}(W\rho_{\text{sep}}) \leq \Lambda_{\max}, \quad (9)$$

i.e. bounded by separable thresholds $(\Lambda_{\min}, \Lambda_{\max})$. [47] The key ingredient of our construction is the shared structure composed of k -local Pauli operators, and which for an exponentially large subset requires only three types of measurements (X , Y and Z) in order to infer the expectation value and determine if a state violates any of the separable thresholds. In the remainder of this paper, we will mostly focus on violations of the upper bound in Eq. (9), though one could also similarly analyze violations of the lower bound and the resulting signatures of entanglement.

To motivate the construction of our observables, we recall the usual Bell violation. In that scenario, given

$$Q = Z \otimes I, \quad S = \frac{1}{\sqrt{2}}(I \otimes Z + I \otimes X), \\ R = X \otimes I, \quad T = \frac{1}{\sqrt{2}}(-I \otimes Z + I \otimes X),$$

the observable $W_B = (R + Q)S + (R - Q)T = \sqrt{2}(X \otimes X + Z \otimes Z)$ serves as a witness for the Bell state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Below, we describe successive generalizations of this observable.

A particularly simple and immediate one over N qubits is defined as

$$W_{XZ}^{(2,N,G)} = \sum_{\langle i,j \rangle} (X_i \otimes X_j + Z_i \otimes Z_j), \quad (10)$$

where the sum over $\langle i, j \rangle$ runs over some subset E_2 of all possible $\binom{N}{2}$ edges in some graph \mathcal{G} of N nodes. We can generalize this further to similarly structured k -local observables of the following form

$$W_{PQ}^{(k,N,\mathcal{G})} = \sum_{\langle i_1, \dots, i_k \rangle} \left(\bigotimes_{j=i_1}^{i_k} P_j + \bigotimes_{j=i_1}^{i_k} Q_j \right), \quad (11)$$

where $P, Q \in \{X, Y, Z\}$ and $P \neq Q$, the sum over $\langle i_1, \dots, i_k \rangle$ similarly runs over some subset of all possible $\binom{N}{k}$ tuples E_k (generalized edges) of some (generalized) graph \mathcal{G} of N nodes.[48]

We can generalize this construction to observables of the form

$$W_{XYZ}^{(k,N,\mathcal{G})} = \sum_{\langle i_1, \dots, i_k \rangle} \left(\bigotimes_{j=i_1}^{i_k} X_j + \bigotimes_{j=i_1}^{i_k} Y_j + \bigotimes_{j=i_1}^{i_k} Z_j \right) \quad (12)$$

where we have used the same notation as in Eq. (11). Both these types of observables themselves belong to a much larger family, with variable coefficients as well as Pauli operators, as described in Appendix B. This entire family of observables admits an upper and a lower bound in the expectation value achievable by any fully separable state, as shown in Theorem 1. In order to prove this theorem, we first introduce the following lemma.

Lemma 1. Generalized Cauchy-Schwarz inequality: *Given a collection of vectors $\vec{x}^{(1)}, \dots, \vec{x}^{(k)}$, where $\vec{x}^{(i)} \in \mathbb{R}^n$ for $i \in \{1, \dots, k\}$, we have the following inequality*

$$\left| \sum_{i=1}^n \left(\odot_{j=1}^k \vec{x}^{(j)} \right)_i \right| \leq \prod_{j=1}^k \left\| \vec{x}^{(j)} \right\| \quad (13)$$

where $|\cdot|$ denotes the absolute value, $\|\cdot\|$ denotes the Euclidean norm, and \odot denotes the Hadamard product.

Note that Lemma 1 reduces to the usual Cauchy-Schwarz inequality [49] for the Euclidean dot product in the base case of $k = 2$. Based on this lemma, we show that the expectation value with respect to any fully separable state of any observable of the form Eq. (11) or (12) or in fact a much larger family of observables described in Appendix B, satisfies an upper and a lower bound given by (up to a sign factor) the number of generalized edges.

Theorem 1. *The absolute value of the expectation value of any N -qubit k -local ($k \leq N$) observable $W_M^{(k,N,\mathcal{G})}$ of the form Eq. (B1) in any fully separable quantum state is upper bounded as*

$$\left| \langle W_M^{(k,N,\mathcal{G})} \rangle_{sep} \right| \leq |E_k| \quad (14)$$

where $|E_k|$ denotes the number of k -tuples $\langle i_1, \dots, i_k \rangle$ (or generalized edges of the generalized graph \mathcal{G}) being summed over, and $M \in \{1, 2, 3\}$ denotes the number of Pauli operators defined on a single k -tuple (see Appendix B for notational convention).

Theorem 1 applies to all the observables that we have introduced so far, as well as a much larger set. It allows us to fulfill one of the criteria, Eq. (9), for detecting entanglement in the system. However, in order to establish these observables as genuine entanglement witnesses, one needs to show that the absolute expectation value can exceed the separable threshold, i.e. $\left| \langle W_M^{(k,N,\mathcal{G})} \rangle \right| > |E_k|$ for any $M \in \{2, 3\}$ (see Appendix B for notational conventions). In general, it is difficult to analytically establish lower bounds for an arbitrary observable in this family. Indeed, some may not even serve as entanglement witnesses.

A straightforward way to demonstrate that the bound from Theorem 1 can be violated is to provide an explicit construction of a state that does it. We consider an observable $W_{XYZ}^{(N,N,\mathcal{G})}$ of the form in Eq. (12) where the sum runs over a single tuple consisting of all N qubits. Its expectation value in GHZ states $|\psi\rangle = \frac{1}{\sqrt{2}} (|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$ produces the maximum possible value of 3 at $N = 4n$, and the minimum possible value of -3 at $N = 4n - 2$ for $n \in \mathbb{Z}^+$. Numerically, we observe that for $N = 2n + 1$ the largest and smallest eigenvalues are $\pm\sqrt{3}$.

We can similarly consider an observable of the form $W_{XZ}^{(N,N,\mathcal{G})}$ which for $N = 2$ is the standard Bell observable W_B up to a global factor. Its expectation in GHZ states gives the maximum possible value of 2 for $N = 2n$, and whose expectation in (locally equivalent to GHZ) states of the form $\frac{1}{\sqrt{2}} (|01\rangle^{\otimes N/2} - |10\rangle^{\otimes N/2})$ gives the minimum possible value of -2 for $N = 2n$ where $n \in \mathbb{Z}^+$. Similar statements could be constructed via local unitary operations about other observables of the form $W_{PQ}^{(N,N)}$, which are therefore seen to belong to the class of observables in Eq. (9), which are equivalent (up to a spectrum shift) to the entanglement witness 2.0 notion introduced in [37]. See Fig. 4 and Appendix D for more details.

In general for observables of the form $W_{PQ}^{(k,N,\mathcal{G})}$, we can demonstrate a lower bound for the maximal eigenvalue that is larger than the number of edges.

Theorem 2. *The maximum eigenvalue λ_{max} of any observable of the form $W_{PQ}^{(k,N,\mathcal{G})}$ (see Eq. (11) and Appendix B) is bounded below by $\lambda_{max} > |E_k| + O\left(\frac{1}{|E_k|^4}\right)$.*

For proofs of Lemma 1 and the previous two theorems, see Appendix C. Theorem 2 establishes that all observables of the form (11) can serve as entanglement witnesses, at least in the weak sense that they can distinguish fully separable states from (some) quantum states that have some entanglement, perhaps only over a subset of the entire N qubits over which the observable is defined. Note however that this theorem does not necessarily apply to the larger family of operators consisting of two Pauli operators on each edge, $W_2^{(k,N,\mathcal{G})}$ (see Appendix B for notational convention), with possibly different operators acting on each of the qubits in a k -local term, nor those consisting of three Pauli operators on each edge, $W_{XYZ}^{(k,N,\mathcal{G})}$. Indeed, we numerically observe that

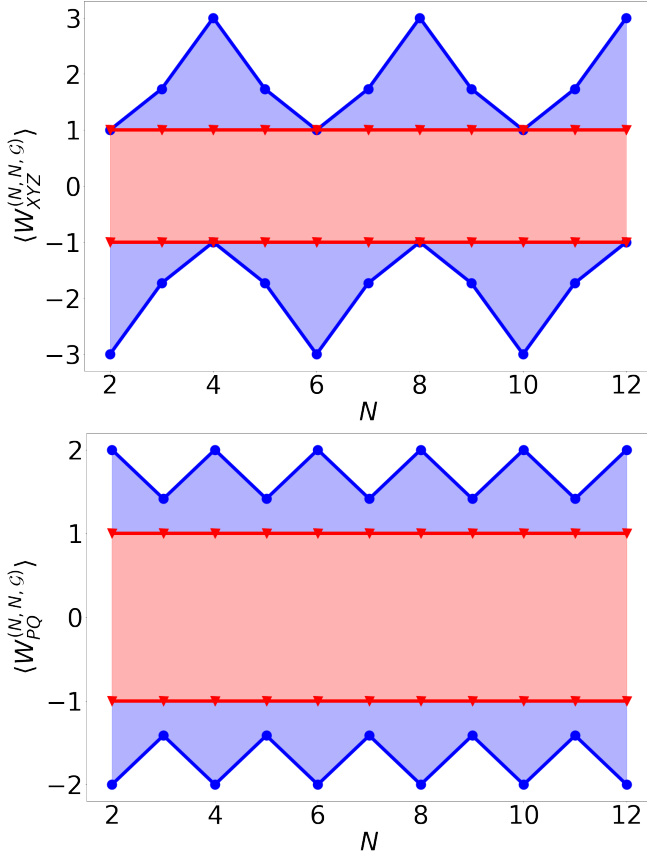


FIG. 4. The maximum eigenvalue compared to the threshold for separable states for observables of the form $W_{XYZ}^{(N,N,G)}$ (top) and $W_{PQ}^{(N,N,G)}$ (bottom), where $P \neq Q$. The blue region is where a violation of the separable threshold is possible, while the red denotes a region where these observables cannot certify, or witness, entanglement since the same expectation value could be produced by a separable state.

the largest eigenvalues of operators of the form $W_{XYZ}^{(2,N,G)}$ equal the number of edges, $|E_k|$, precisely the upper bound for separable states established in Theorem 1, and therefore such observables cannot serve as witnesses.

Nevertheless, the family of observables to which Theorem 2 applies contains exponentially many elements, consisting of all non-trivial graph types, with different choices of Pauli operators, with each of these constituting a different entanglement witness. For a fixed N , the total number of such elements is $3 \sum_{k=2}^N (2^{\binom{N}{k}} - 1)$, where the exponent $\binom{N}{k}$ is the total possible number of k -tuples, the base 2 counts whether a given k -tuple/edge belongs to a (generalized) graph or not, the -1 subtracts the trivial graph with no edges (and which does not yield a witness), and 3 counts the possibilities XZ , XY and YZ .

It is also instructive to look at the upper bound of the largest eigenvalue of these observables. Such an upper bound would tell us the extent of violation of the separable threshold provided by Theorem 1 that is possible even in principle. The following lemma establishes a

fairly generic upper bound applicable to all observables to which Theorem 1 applies.

Lemma 2. *The maximum eigenvalue λ_{max} of any N -qubit k -local ($k \leq N$) observable of the form $W_M^{(k,N,G)}$ defined in Eq. (B1) (see Appendix B) is bounded above by $\lambda_{max} \leq M|E_k|$.*

Proof. Using successive applications of Weyl's inequality for the largest eigenvalue of the sum of two Hermitian matrices A and B

$$\lambda_{max}(A+B) \leq \lambda_{max}(A) + \lambda_{max}(B) \quad (15)$$

and the fact that $\lambda_{max}(P) = 1$ for any Pauli operator P , we readily obtain

$$\begin{aligned} \lambda_{max}\left(W_M^{(k,N,G)}\right) &\leq \sum_{\langle i_1, \dots, i_k \rangle} \sum_{m=1}^M \lambda_{max}\left(\alpha_m \bigotimes_{j=1}^k \sigma_{i_j}^{(a_{j,m})}\right) \\ &\leq \sum_{\langle i_1, \dots, i_k \rangle} \sum_{m=1}^M \lambda_{max}\left(\bigotimes_{j=1}^k \sigma_{i_j}^{(a_{j,m})}\right) \\ &= \sum_{\langle i_1, \dots, i_k \rangle} \sum_{m=1}^M (1) = M|E_k| \quad (16) \end{aligned}$$

□

The upper bound of $M|E_k|$ is saturated, though not uniquely and not always, by graphs with edges on disjoint pairs. This can be seen for $M = 2$ by choosing the state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ on each pair to give an expectation value of 2 to the observable $ZZ+XX$ defined on this pair, 2 being the maximum value of this observable, since both $\langle ZZ \rangle \leq 1$ and $\langle XX \rangle \leq 1$. For such graphs, the difference $\Delta W = \lambda_{max} - \max_{\rho_{sep}} \text{Tr}(W \rho_{sep}) = |E_2| = \lfloor \frac{N}{2} \rfloor$.

Numerically, we find that this difference is the maximum value achievable by graphs of any connectivity for $W_{XZ}^{(2,N,G)}$ [50]. In the next sections, we see that observables of the form $W_{XZ}^{(2,N,G)}$ have a special significance for QAOA-Maxcut states.

B. Entanglement Witnesses for QAOA-MaxCut

In QAOA [10], we seek to find the (approximate) ground state of diagonal Hamiltonians. For the particular case of MaxCut over a graph of equal weights, we seek to maximize the cost Hamiltonian $C' = \frac{1}{2} \sum_{\langle i,j \rangle} (I - Z_i Z_j)$, or equivalently, minimize the cost Hamiltonian $C = \sum_{\langle i,j \rangle} Z_i Z_j$. In the basic formulation of QAOA, we start with an initial state of $H^{\otimes N} |0\rangle^{\otimes N}$, and then apply the unitary $e^{-i\beta_j B} e^{-i\gamma_j C}$ a total of p times, where $B = \sum_{i=1}^N X_i$ and the parameters $\vec{\theta} = (\beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_p)$ are to be optimized to minimize the expectation value $\langle C \rangle$.

For QAOA states built out of the 2-local MaxCut cost Hamiltonian, the 2-local observable $W_{XZ}^{(2,N,G)}$ defined

over the same set of edges as the cost Hamiltonian is a natural candidate for an entanglement witness. Indeed, we can view this observable as being constructed directly out of the cost Hamiltonian C as $W_{XZ}^{2,N,\mathcal{G}} = C + H^{\otimes N} C H^{\otimes N}$. Note that if $\rho(\vec{\theta}_p)$ represents the parametric QAOA state after p rounds, and A is any observable, then $\max_{\theta} \text{Tr} \left(A \rho(\vec{\theta}_{p+k}) \right) \geq \max_{\theta} \text{Tr} \left(A \rho(\vec{\theta}_p) \right)$ for all $k \in \mathbb{Z}^+$. In other words, if W is a witness for QAOA $_p$ states, then it is also a witness for QAOA $_{p+k}$ states. We now show that observables from the $W_{XZ}^{(2,N,\mathcal{G})}$ family can serve as entanglement witnesses for QAOA states, noting a lemma before we do so.

Lemma 3. *Given $C = \sum_{\langle ij \rangle \in E_2} Z_i Z_j$ where E_2 is some edge set, and $B = \sum_i X_i$, the expectation values of the operators $X_u X_v$, $Y_u Y_v$ and $Z_u Z_v$ over some edge $\langle uv \rangle \in E_2$ in the $p = 1$ QAOA-MaxCut state $|\psi(\gamma, \beta)\rangle = e^{-i\beta B} e^{-i\gamma C} H^{\otimes N} |0\rangle^{\otimes N}$ are given respectively as*

$$\begin{aligned} \langle X_u X_v \rangle &= \frac{1}{2} \cos^{d_u+d_v-2f} 2\gamma (1 + \cos^f 4\gamma) \\ \langle Y_u Y_v \rangle &= -\frac{1}{2} \sin 4\beta \sin 2\gamma (\cos^{d_u} 2\gamma + \cos^{d_v} 2\gamma) \\ &\quad + \frac{1}{2} \cos^2 2\beta \cos^{d_u+d_v-2f} 2\gamma (1 - \cos^f 4\gamma) \\ \langle Z_u Z_v \rangle &= \frac{1}{2} \sin 4\beta \sin 2\gamma (\cos^{d_u} 2\gamma + \cos^{d_v} 2\gamma) \\ &\quad + \frac{1}{2} \sin^2 2\beta \cos^{d_u+d_v-2f} 2\gamma (1 - \cos^f 4\gamma) \end{aligned} \quad (17)$$

where d_u (d_v) is the number of neighbors of u (v) excluding v (u), and f is the number of triangles in the graph that include the edge $\langle uv \rangle$.

Proof. See Appendix C. \square

Note that for triangle free graphs $f = 0$, we have that $\langle Z_u Z_v \rangle(\beta, \gamma) = \langle Y_u Y_v \rangle(-\beta, \gamma) = \langle Y_u Y_v \rangle(\beta, -\gamma)$, so that if an observable of the form $W_{XZ}^{(2,N,\mathcal{G})}$ is a witness for some QAOA state, then so is $W_{XY}^{(2,N,\mathcal{G})}$ defined over the same graph. We can now show that the constructed observables can serve as entanglement witnesses for QAOA states prepared through MaxCut cost Hamiltonians on different graph topologies \mathcal{G} .

Ring graph

Theorem 3. *Given a ring Hamiltonian $C = \sum_i Z_i Z_{i+1}$ and the standard mixer $B = \sum_i X_i$, the observable defined on the ring graph $\mathcal{G}_{\text{ring}}$*

$$W \equiv W_{XZ}^{(2,N,\mathcal{G}_{\text{ring}})} = \sum_i X_i X_{i+1} + Z_i Z_{i+1} \quad (18)$$

serves as an entanglement witness for QAOA states $\prod_{k=1}^p (e^{-i\beta_k B} e^{-i\gamma_k C}) |+\rangle^{\otimes N}$.

In particular, the gap $\Delta W = \max\langle W \rangle_{\text{QAOA}} - \max\langle W \rangle_{\text{sep}}$ between the maximum expectation value for W achievable by QAOA states and that achievable by separable states is lower bounded as

$$\Delta W \geq N \frac{\sqrt{2}-1}{2} \quad (19)$$

Proof. For ring graphs of N nodes, we have $d_u = d_v = 1$, $f = 0$ for each of N edges. Therefore, the observable has expectation value $\langle W \rangle = N W_e$, where W_e denotes the expectation value $\langle XX + ZZ \rangle$ over a single edge in the ring graph. From Lemma 3, we have

$$W_e = \cos^2 2\gamma + \frac{\sin 4\gamma \sin 4\beta}{2} \quad (20)$$

At all local optima, we have $\partial_{\gamma} \langle W_e \rangle = \partial_{\beta} \langle W_e \rangle = 0$. Simultaneously solving these equalities gives us either $\gamma_a = \frac{n\pi}{4}$ or $\gamma_b = \frac{\pi}{8} (n - \frac{1}{2})$, and correspondingly either $\sin 4\beta_a = 0$ or $\sin 4\beta_b = \pm 1$ (+1 for n odd, and -1 for n even) respectively.

Let $D(\gamma, \beta) = \partial_{\gamma\gamma} W_e \cdot \partial_{\beta\beta} W_e - (\partial_{\beta\gamma} W_e)^2$. Then, $D(\gamma_a, \beta_a) < 0$, indicating a saddle point. For $n = 3, 7, 11, \dots$ or $n = 2, 6, 10, \dots$, we have $D(\gamma_b, \beta_b) > 0$ and $\partial_{\gamma\gamma} W_e > 0$, $\partial_{\beta\beta} W_e > 0$ indicating a local minimum. For $n = 1, 5, 9, \dots$ or $n = 0, 4, 8, \dots$, we have $D(\gamma_b, \beta_b) > 0$ and $\partial_{\gamma\gamma} W_e < 0$, $\partial_{\beta\beta} W_e < 0$ indicating a local maximum. At all these local maxima, $W_e(\gamma_b, \beta_b) = \frac{1+\sqrt{2}}{2}$. Since this is true of all local maxima, this also provides the global maximum.

For separable states, the maximum achievable value for W_e is one. Since all edges contribute the same expectation value in a ring graph, the gap ΔW at $p = 1$ is $N \frac{\sqrt{2}-1}{2}$. For larger p values, this gap can increase in value and the theorem follows. \square

Regular triangle-free graphs

More generally, we can consider regular triangle-free graphs $\mathcal{G}_{\text{triangle-free}}$. This family includes the ring graph as a special case, but also other types such as bipartite graphs. For all such graphs, we have $d_u = d_v = d$ and $f = 0$, so that we have $\langle W_{XZ}^{(2,N,\mathcal{G}_{\text{triangle-free}})} \rangle = N_{\text{edges}} W_e$ where

$$W_e = \cos^{2d} 2\gamma + \sin 4\beta \sin 2\gamma \cos^d 2\gamma \quad (21)$$

This expression can be optimized over γ and β to find its maximum value. We find numerically that $\max_{\gamma,\beta} W_e > 1$ and therefore serves as a witness for arbitrary large d , as seen in Fig. 5. However, it is greatest at $d = 1$ and smoothly decays to 1 for large values of d . This asymptotic decay can be seen by considering Eq. (21) in the limit $d \rightarrow \infty$. In this limit,

$$\lim_{d \rightarrow \infty} \cos^d 2\gamma \rightarrow \begin{cases} +1, & \gamma = n\pi, \\ (-1)^d, & \gamma = (n + \frac{1}{2})\pi, \\ 0, & \text{otherwise} \end{cases} \quad (22)$$

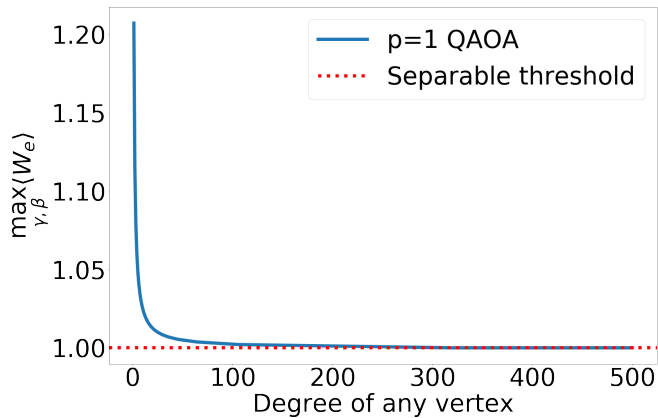


FIG. 5. The maximum achievable expectation value, divided by the number of edges, of observables of the form $W_{XZ}^{(2,N,\mathcal{G}_{\Delta\text{-free}})}$ for regular, triangle-free graphs in $p = 1$ QAOA. At asymptotically large values of the degree of any vertex in the regular graph, this observable fails to serve as an entanglement witness for $p = 1$ QAOA-MaxCut states.

while $\sin 2\gamma \rightarrow 0$ at $\gamma = n\pi$ and $\gamma = (n + \frac{1}{2})\pi$, for $n \in \mathbb{Z}$. Therefore, one term will asymptotically converges to 0 while the other converges to 1, which is the value saturated by separable states.

Linear chain graph

A related graph is that of a linear chain, $\mathcal{G}_{\text{line}}$ which unlike the ring, has open boundary conditions. This is the graph we use in our experiments (see sub-section III G). Here, we can use the fact that $d_u = 0$, $d_v = 1$ (or vice versa) for the two edges at either end, $d_u = d_v = 1$ for all other edges, $f = 0$ for every edge, and the fact that there are $N - 1$ total edges, where N is the number of nodes in the linear chain graph. After some algebraic simplification, the expectation value of $W_{XZ}^{(2,N,\mathcal{G}_{\text{line}})}$ becomes for $N \geq 3$

$$\langle W_{XZ}^{(2,N,\mathcal{G}_{\text{line}})} \rangle = \frac{\sin(4\beta)}{2} [(N-2)\sin(4\gamma) + 2\sin(2\gamma)] + (N-3)\cos^2(2\gamma) + 2\cos(2\gamma) \quad (23)$$

while for $N = 2$, it is $\langle W_{XZ}^{(2,N,\mathcal{G}_{\text{line}})} \rangle = 1 + \sin 4\beta \sin 2\gamma$. It is found numerically that the quantity $\langle W_{\text{line}} \rangle_{\text{max}} \equiv \max_{\beta,\gamma} \langle W_{XZ}^{(2,N,\mathcal{G}_{\text{line}})} \rangle$ grows linearly with the number of qubits for $N \geq 3$, roughly as

$$\begin{aligned} \langle W_{\text{line}} \rangle_{\text{max}} &\approx 1.207N - 1.019 \\ &> N - 1 \end{aligned} \quad (24)$$

while for $N = 2$, $\langle W_{\text{line}} \rangle_{\text{max}} = 2$. Therefore, this observable serves as an entanglement witness for $p = 1$ QAOA.

Fully connected graph

Lastly, we may also consider the fully connected graph $\mathcal{G}_{\text{full}}$. In this case, we use the fact that $d_u = d_v = f = N - 2$ for every edge, and the fact that there are a total of $|E_2| = \binom{N}{2}$ edges. The expectation value of $W_{\text{full}} \equiv W_{XZ}^{(2,N,\mathcal{G}_{\text{full}})}$ becomes

$$\langle W_{\text{full}} \rangle = \binom{N}{2} \left(\frac{1 + \cos^{N-2} 4\gamma}{2} + \sin 4\beta \sin 2\gamma \cos^{N-2} 2\gamma + \frac{\sin^2 2\beta (1 - \cos^{N-2} 4\gamma)}{2} \right) \quad (25)$$

Let $W_e(\gamma, \beta, N) \equiv \langle W_{\text{full}} \rangle / |E_2|$ denote the term in the paranthesis above. Requiring that $W_e > 1$ and that therefore W_{full} serves as an entanglement witness for the $p = 1$ QAOA state is equivalent to requiring that

$$1 - \cos^{N-2} 4\gamma < 2 \tan 2\beta \sin 2\gamma \cos^{N-2} 2\gamma \quad (26)$$

Numerically, we find that this condition is satisfied by some (γ, β) (e.g. by $\gamma = \pi/499$, $\beta = 122\pi/499$) for up to $N = 5000$. In the large N limit however, the observable $W_{XZ}^{(2,N,\mathcal{G}_{\text{full}})}$ asymptotically fails to be a witness as the maximum value it attains is no more than that achieved by separable states. We can see this by noting the asymptotic behavior of $\cos^n x$ in the limit of large n as in Eq. (22), and the fact that $\sin x = 0$ whenever $\cos x = 1$ to find

$$\lim_{N \rightarrow \infty} \max_{\gamma, \beta} W_e(\gamma, \beta, N) \rightarrow 1 \quad (27)$$

which is precisely the bound for separable states.

In addition to the types of graphs analyzed above, we numerically checked that the observables $W_{XZ}^{(2,5,\mathcal{G})}$ serve as entanglement witnesses for $p = 1$, $N = 5$ QAOA-MaxCut states defined over any of the $2^{\binom{5}{2}} - 1$ non-trivial graphs, where the observable is defined over the same set of edges as the cost Hamiltonian. We conjecture that this holds true for any finite N .

The asymptotic decay we observe at large N for complete graphs and large d for regular triangle-free graphs for the (normalized by the number of edges) maximum expectation values of $W_{XZ}^{(2,N,\mathcal{G})}$ in $p = 1$ QAOA states may reflect the fact that it becomes increasingly difficult to solve the corresponding MaxCut problem for larger values of N and d respectively at $p = 1$, and that a larger number of QAOA rounds (p) may be required to solve it. It is generally difficult to study the performance of QAOA at large values of p , but for reasonably small values of fixed N and d , we numerically find that $\max_{\gamma, \beta} \langle W_{XZ}^{(2,N,\mathcal{G})} \rangle$ saturates around the same p as when the cost Hamiltonian saturates.

C. XYZ witnesses

In addition to the $W_{XZ}^{(2,N,\mathcal{G})}$ family of observables, we may also consider observables of the form $W_{XYZ}^{(N,N,\mathcal{G})}$. In this case, we can calculate the expectation values of $\langle X^{\otimes N} \rangle$, $\langle Y^{\otimes N} \rangle$, $\langle Z^{\otimes N} \rangle$ rather easily. We note that QAOA-MaxCut states are \mathbb{Z}_2 symmetric (see for example [13, 14] for additional reference), and we can express all such states as

$$|\Psi\rangle = \sum_{x \in \mathcal{X}_0} c_x (|x\rangle + |\bar{x}\rangle) \quad (28)$$

Where \mathcal{X}_0 is the set of bitstrings that start with 0, and \bar{x} represents the ones' complement of x . We then have

$$\begin{aligned} X^{\otimes N} |\Psi\rangle &= \sum_{x \in \mathcal{X}_0} c_x X^{\otimes N} (|x\rangle + |\bar{x}\rangle) \\ &= \sum_{x \in \mathcal{X}_0} c_x X^{\otimes N} (|\bar{x}\rangle + |x\rangle) = |\Psi\rangle \end{aligned} \quad (29)$$

so that $\langle X^{\otimes N} \rangle = \langle \Psi | \Psi \rangle = 1$. Moreover,

$$Z^{\otimes N} |\Psi\rangle = \sum_{x \in \mathcal{X}_0} c_x ((-1)^{|x|} |x\rangle + (-1)^{|\bar{x}|} |\bar{x}\rangle) \quad (30)$$

where $|x|$ denotes the hamming weight of x . Note that for even N , $(-1)^{|x|} = (-1)^{|\bar{x}|}$ and for odd N , $(-1)^{|x|} = (-1)^{1+|\bar{x}|}$. Therefore,

$$\langle Z^{\otimes N} \rangle = \begin{cases} 0 & \text{odd } N, \\ \sum_{x \in \mathcal{X}_0} 2(-1)^{|x|} |c_x|^2 & \text{even } N \end{cases} \quad (31)$$

Furthermore,

$$\begin{aligned} \langle Y^{\otimes N} \rangle &= \langle (-i)^N Z^{\otimes N} X^{\otimes N} \rangle \\ &= (-i)^N \langle Z^{\otimes N} \rangle \end{aligned} \quad (32)$$

for even N . Based on the above observations, we see that necessary conditions for a violation of the separable threshold of 1 are $N = 2n$ for $W_{XZ}^{(N,N,\mathcal{G})}$ and $W_{XY}^{(N,N,\mathcal{G})}$, and $N = 4n$ for $W_{YZ}^{(N,N,\mathcal{G})}$ and $W_{XYZ}^{(N,N,\mathcal{G})}$, where $n \in \mathbb{Z}^+$. However, these constraints alone do not guarantee that the separable threshold is violated, which may not always be possible for certain types of graphs. However, we do observe numerically for a few small allowed values of N that such violations are indeed possible at least for some graphs.

D. Separability

Having established an upper bound for the expectation value of a large family of observables in fully separable states, it is a natural to ask next whether the type of entanglement detected by certain witnesses is genuine N -partite entanglement, or could be achieved with entanglement over only a subset of those N qubits. In other words, one can look at the k -separability properties of the observables described above. An N -partite

pure quantum state $|\Psi_{ksep}\rangle$ is k -separable iff it can be expressed in the form

$$|\Psi_{ksep}\rangle = |\Psi_1\rangle \otimes |\Psi_2\rangle \otimes \cdots \otimes |\Psi_k\rangle \quad (33)$$

and likewise, a mixed state ρ_{ksep} is k -separable iff it can be expressed in the form

$$\rho_{ksep} = \sum_i p_i |\Psi_{ksep}^{(i)}\rangle \langle \Psi_{ksep}^{(i)}| \quad (34)$$

We know from theorem 1 that the maximum expectation value achievable by a fully separable state for any observable of the form $W_M^{(k,N,\mathcal{G})}$ is $|E_k|$. On the other hand, we have shown that this bound is violated by N -qubit GHZ states for observables of the form $W_{XYZ}^{(N,N,\mathcal{G})}$ whenever $N = 4n$ and for those of the form $W_{XZ}^{(N,N,\mathcal{G})}$ whenever $N = 2n$ for some $n \in \mathbb{Z}^+$. However, such violations do not necessarily certify genuine N -partite entanglement.

To illustrate, consider an M -separable product state of the form $|\Phi_n^M\rangle = \otimes_{m=1}^M |\psi_n^{(m)}\rangle$, where $|\psi_n\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$ is the n -qubit GHZ state. Then, the maximal expectation value of 3 for the observable $W_{XYZ}^{(N,N,\mathcal{G})}$ is achieved by both an N -qubit GHZ state $|\psi_N\rangle$, but also any M -separable state $|\Phi_n^M\rangle$ whenever $N = 4nM$. Likewise, $\langle \psi_N | W_{XZ}^{(N,N,\mathcal{G})} | \psi_N \rangle = \langle \Phi_n^M | W_{XZ}^{(N,N,\mathcal{G})} | \Phi_n^M \rangle = 2$, the maximal value, whenever $N = 2nM$.

For $W_{XZ}^{(2,N,\mathcal{G})}$, the upper bound in Lemma 2 implies that if the edge set is given by $E_2 = \cup_{j=1}^3 E_2^{(j)}$ where $E_2^{(1)}$ and $E_2^{(2)}$ are respectively the edges contained within each halves of some partition of the graph \mathcal{G} , and $E_2^{(3)}$ the edges that cross that partition, then any bi-separable state could achieve an expectation value of at most $2(|E_2^{(1)}| + |E_2^{(2)}| + |E_2^{(3)}|)$, which can certainly be exceeded if the upper bound of $2(|E_2^{(1)}| + |E_2^{(2)}| + |E_2^{(3)}|)$ is saturated, in which case the observable would certify genuine N -partite entanglement[51]. However, such a saturation of the upper bound is not guaranteed. In principle, the generalized Cauchy-Schwarz inequality (Lemma 1) could be employed in conjunction with the inequalities proved earlier, as well as the physicality requirement $\sum_{\alpha=1}^{4^k} \langle P_\alpha \rangle^2 \leq 2^k$, where the sum runs over all k -qubit Pauli operators in any k -qubit state ρ , to obtain some upper bounds on the expectation value of such observables in k -separable states. However, in practice we can find tighter bounds using numerical techniques. For each $k \leq N$, we compute the largest expectation value achievable by any k -separable pure state, setting an upper bound to the largest expectation value achievable by any k -separable mixed state.

Concretely, we adopt a trigonometric parametrization of a classical probability vector $\vec{p} = (p_1, \dots, p_{2^N})$ as

$$p_i = \sin^2 \theta_{i-1} \prod_{j=i}^{N-1} \cos^2 \theta_j \quad (35)$$

with $\theta_0 = \pi/2$, which naturally enforces the normalization $\sum_i p_i = 1$. A pure N -qubit quantum state can then be parametrized in terms of this probability vector and $2^N - 1$ relative phases

$$|\psi\rangle_i = \sqrt{p_i} e^{i\theta_i} \quad (36)$$

where $\theta_0 = 0$.

As an example, numerically optimizing for the expectation value of $W_{XZ}^{(2,5,\mathcal{G})}$ in various k -separable states ($k \leq 5$), we observe in Fig. 6 that the witness $W_{XZ}^{(2,5,\mathcal{G})}$ serves to certify genuine multi-partite entanglement in the linear chain QAOA-MaxCut state over $N = 5$ qubits at $p \geq 2$, while at $p = 1$ the witness can certify that the produced QAOA state is not 3-(or higher) separable, but cannot certify that it is not 2-separable. On the other hand, we also observe in Fig. 6 that the corresponding witness $W_{XZ}^{(2,5,\mathcal{G})}$ serves to certify genuine multi-partite entanglement in the complete (all-to-all connected) graph QAOA-MaxCut state over $N = 5$ qubits at all p values.

In both these examples, there is a strict hierarchy of the maximum achievable expectation value $\langle W_{XZ}^{(2,N,\mathcal{G})} \rangle$ according to the separability of the class of states in which that expectation is computed, i.e. $\max_{\rho_{k\text{-sep}}} \langle W \rangle > \max_{\rho_{k'\text{-sep}}} \langle W \rangle$ whenever $k < k'$. This is not always the case, and the existence of such a strictly ordered hierarchy depends, at least in the case of the $W_{XZ}^{(2,N,\mathcal{G})}$, as well as their cousins in the $W_{PQ}^{(2,N,\mathcal{G})}$ family, on the structure of the graph whose edges are being summed over. In particular, if the graph \mathcal{G} in $W_{PQ}^{(2,N,\mathcal{G})}$ is only non-trivially defined on $m \leq N$ qubits, then the maximum expectation value is the same for all k -separable states whenever $k \leq N - m + 1$.

E. Entanglement Potency

Since the entanglement structure of multi-qubit states is very rich [25, 52], it is relevant to ask how many entangled states can the discussed Bell-type observables detect? In order to answer this question, we introduce a metric on witnesses, called *entanglement potency*, and compute its value on the space of QAOA-MaxCut states (according to a particular distribution) as well as Haar random states. We see that these observables have non-negligible entanglement potency for QAOA-MaxCut states, but close to zero potency for Haar random states, making them a suitable choice for detecting entanglement generated by this particular ansatz.

Definition 1. *The entanglement potency of an entanglement witness W with respect to some measurable set of states ξ is given by*

$$P_\xi(W) = \frac{V_d^\xi}{V_T^\xi}, \quad (37)$$

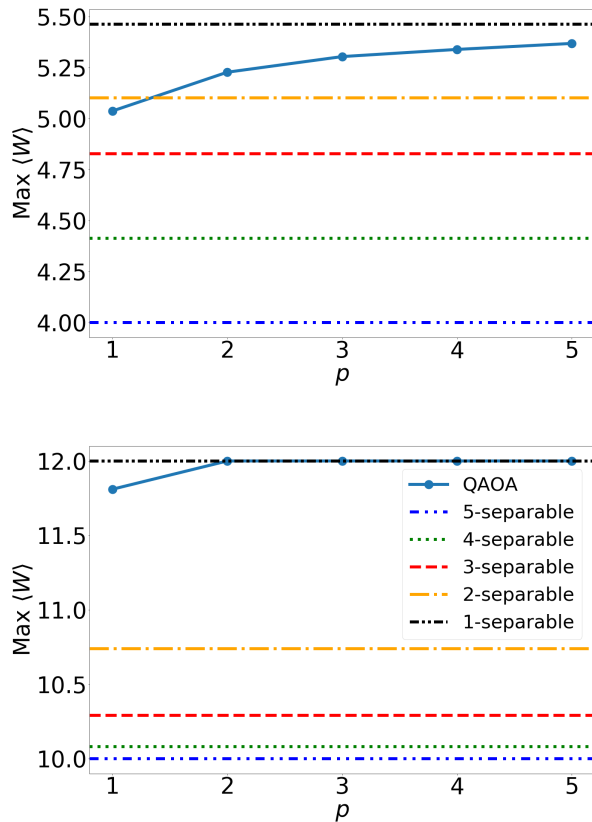


FIG. 6. Separability properties of the entanglement witness $W = W_{XZ}^{(2,5,\mathcal{G})}$ for a linear chain graph (top) and complete graph (bottom) MaxCut problem. We numerically compute and report the maximum achievable expectation value of this observable in k -separable states for various values of $k \leq N = 5$, and compare it to the maximum expectation value that QAOA states can achieve for various values of p .

where V_T^ξ is the total volume of quantum states in the set ξ , and V_d^ξ is the volume of such states whose entanglement is detected by W .

The set ξ could be a parametric family of states, such as QAOA or VQE, or it could also be the set of all N -qubit states (e.g. with Hurwitz parametrization [53, 54]). Since the set is measurable, the definition allows us the freedom to draw from arbitrary distributions over the chosen set of states. In practice, we would draw states according to some fixed distribution over some collection, and measure the fraction of drawn states that violate the separable threshold for the observable W . This provides an unbiased estimate of the entanglement potency defined above, and approaches its ideal value asymptotically as the number of samples becomes large.

One may expect that the potency of any N -qubit observable in Haar random states decreases with N and perhaps vanishes at large N , since each observable can detect only a small subset of entangled states, while the volume of the entire set of states grows exponentially.

This is similar to the vanishingly small volume of separable states at large N [55]. We numerically observe that for $W_{XZ}^{(2,N,\mathcal{G})}$ on fully connected and ring graphs, the potency is negligible while for $W_{XZ}^{(N,N,\mathcal{G})}$ it decreases with N (see Table I). For QAOA-MaxCut, we observe a non-negligible amount of entangled states detected by the investigated observables (see Fig. 7, 8 and Table I).

Applying this metric to QAOA-MaxCut, we consider states prepared via the ring graph Hamiltonian together with observables of the type $W_{XZ}^{(2,N,\mathcal{G}_{\text{ring}})}$ defined on the same ring graph. The potency is then given by the volume of (γ, β) for which the normalized expectation value of this witness given in Eq. (20) is greater than one. This can be expressed as the following integral

$$P_{\text{QAOA}_1}(W_{XZ}^{(2,N,\mathcal{G}_{\text{ring}})}) = \frac{1}{(2\pi)^2} \int_0^{2\pi} d\beta \int_0^{2\pi} d\gamma \Theta \left(\cos^2(2\gamma) + \frac{1}{2} \sin(4\gamma) \sin(4\beta) - 1 \right) \quad (38)$$

where Θ is the Heaviside step function. This integral evaluates to 1 only for angles that violate the separable bound, and is normalized by the total volume of $(2\pi)^2$ (i.e. the area of $[0, 2\pi) \times [0, 2\pi)$ of β, γ intervals).

Although we could try to evaluate this integral numerically, in practice we use the Monte Carlo (MC) method to sample QAOA-MaxCut states with uniformly random values of γ and β from $[0, 2\pi)$ for $p = 1$ and $p = 5$ and measure the fraction of states that violate the separable threshold for the witness, i.e. satisfy $\langle W_{XZ}^{(2,N,\mathcal{G}_{\text{ring}})} \rangle > N$. These results are depicted in the bottom figure of Fig. 7. We repeat the MC analysis for fully connected MaxCut problem with the witness sharing the fully connected graph structure (see top figure in Fig. 7). One notices that the potency of the witness decreases for $p = 5$ QAOA-MaxCut states compared to $p = 1$ states. This behavior is unsurprising, since more layers of QAOA increases the algorithm's expressibility, allowing a larger fraction of non-detectable states to be explored, so that the fraction of detectable entangled states should decrease.

We also numerically estimate the entanglement potency of $W_{XZ}^{(N,N,\mathcal{G})}$ with respect to $p = 1$ QAOA states prepared via fully connected graphs with Hamiltonians

$$H = \sum_{\langle i,j \rangle \in E_f} J_{i,j} Z_i Z_j, \quad (39)$$

with $J_{i,j} \in \{-1, +1\}$. In particular, we investigate the case for $N = 4$, and employ the MC technique to estimate the potency of $W_{XZ}^{(4,4,\mathcal{G})}$ with respect to the set of $p = 1$ QAOA states prepared according to each of 64 possible Hamiltonians (from which only 6 are non-isomorphic), with uniformly randomly drawn values of γ and β (see Fig. 8). Additionally, we check the scaling properties of randomly selected Hamiltonians that

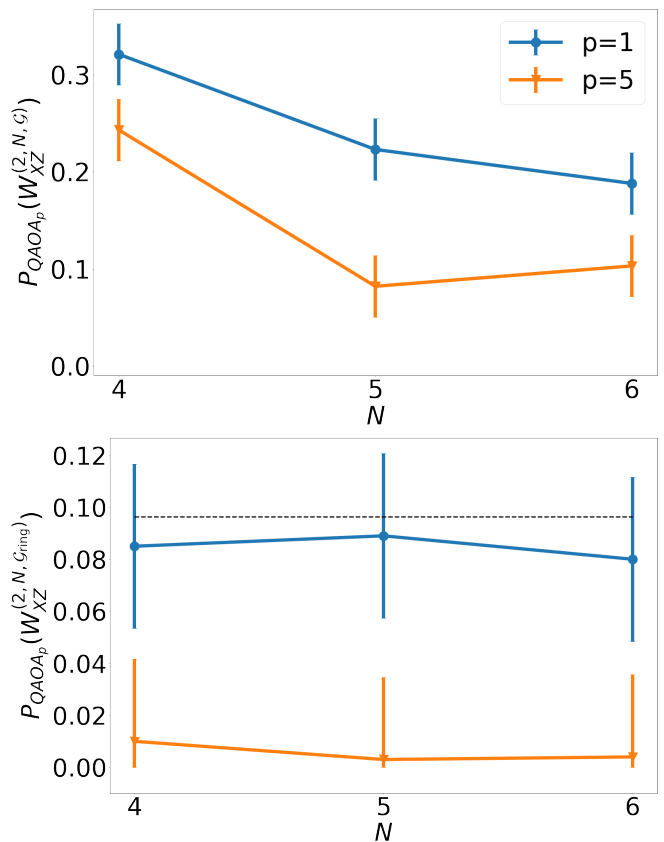


FIG. 7. Entanglement potencies for $W_{XZ}^{(2,N,\mathcal{G})}$ with both the witness observable and problem Hamiltonian corresponding to (top) fully connected N -qubit graph, for QAOA MaxCut problem, and (bottom) N -qubit ring graph, with $p = 1$ (blue) and $p = 5$ (orange) sampled numerically with 1000 uniform random (γ, β) angles. Dashed line (bottom) shows numerical integration of Eq. (38) at ≈ 0.096 . Although we do not depict this in the figure, we similarly sampled 10,000 Haar random states (as opposed to only QAOA-MaxCut states), and found no violation of the separable threshold, so we consider the entanglement potency of both these observables for Haar random states to be zero for each of the investigated values of N .

display \mathbb{Z}_2 symmetry (i.e. randomly sampling $J_{i,j}$ coefficients from $\{-1, 1\}$) which constitute ξ . We report these entanglement potencies in Fig. 8 and in Table I.

F. Relationship to other algorithms

The preceding discussion has identified an exponentially large class of observables that each serve as entanglement witnesses for *some* respective state(s). A priori, these states may be arbitrary. However, we have shown that these are especially well tailored for QAOA-MaxCut states. Indeed, a subset of the observables we construct are in general relevant for combinatorial optimization problems with an overall \mathbb{Z}_2 symmetry, i.e. for diagonal cost Hamiltonians C such that $[C, X^{\otimes N}] = 0$

N	QAOA $p = 1$			QAOA $p = 5$			Random
	mean	max	min	mean	max	min	mean
4	0.48	0.96	0	0.48	0.85	0.16	0.12 ± 0.01
5	0	0	0	0	0	0	0.09 ± 0.01
6	0.47	0.52	0.42	0.50	0.53	0.46	0.05 ± 0.01

TABLE I. Entanglement potency for a set of problem Hamiltonians with randomly chosen $J_{ij} \in \{-1, 1\}$ (i.e. with \mathbb{Z}_2 symmetry), along with the minimum and maximum potency found in the set of problem Hamiltonians. Each potency was calculated using 1000 random angle samples. The data at $N = 4, 5, 6$ were calculated using 60, 80, and 100 random problem Hamiltonians, respectively - note that based on the histogram Fig. 8 it is likely that for $N = 6$ there can exist Hamiltonians for which $P_\xi(W_{XZ}^{N,N,\mathcal{G}}) = 0$, where ξ represents the set of $p = 1$ QAOA states prepared according to the corresponding Hamiltonian. The random state data represents the potency of the observable over 10,000 Haar random states.

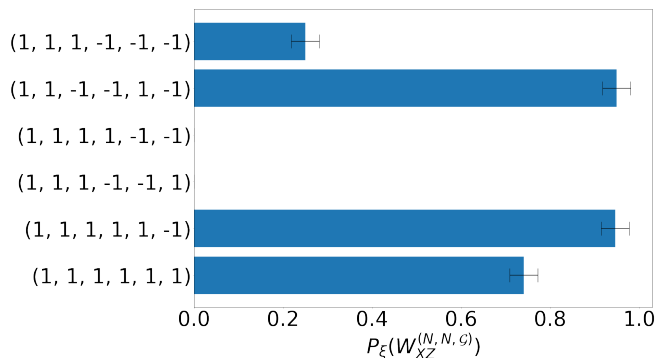


FIG. 8. Entanglement potency of witness observables $W_{XZ}^{(N,N,\mathcal{G})}$ for random fully connected problem Hamiltonians of the form $\sum_{\langle i,j \rangle} J_{ij} Z_i \otimes Z_j$ with $J_{ij} \in \{-1, 1\}$. Here, ξ represents the set of $p = 1$ QAOA states prepared according to the corresponding Hamiltonian. We depict the potency for all possible non-isomorphic fully connected Hamiltonians on 4 nodes (specified by J_{ij} terms along the y-axis), sampled with 1000 random angles. Error bars represent sampling uncertainty $\sigma = \frac{1}{\sqrt{N}} \approx 0.03$.

holds. For such problems, the solution state is of the form

$$|\psi\rangle = \otimes_{i=0}^{N-1} X_i^{n_i} |GHZ\rangle \quad (40)$$

where $n_i \in \{0, 1\}$, or a superposition of such states for different assignments of n_i , and $|GHZ\rangle = \frac{1}{\sqrt{2}} (|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$. We have already identified entanglement witnesses for GHZ states from the family of observables we constructed above, such as $W_{XYZ}^{N,N,\mathcal{G}} = X^{\otimes N} + Y^{\otimes N} + Z^{\otimes N}$ (see Appendix D for more details). Therefore, an entanglement witness for a state of the form Eq. (40) is given by

$$W'_{XZ} = X^{\otimes N} + (-1)^{\sum n_i} (Y^{\otimes N} + Z^{\otimes N}) \quad (41)$$

while those from the W_{XYZ} family can be similarly constructed. Thus, we see that the entanglement witnesses we have constructed can be used to certify entanglement in a quantum state output by an algorithm for exact optimization as well as approximate optimization, such as QAOA.

Note however that not all quantum algorithms may solve combinatorial optimization problems by producing states of the form Eq. (40), and may instead output classical bitstrings, or separable states. Moreover, the witnesses we describe only certify the entanglement in final output states of the form Eq. (40), or approximately this form, and do not reveal how entanglement may assist in the computation between the initial and final states. Nevertheless, we note that since structured ansätze such as QAOA are by nature layered, one may probe entanglement generation for intermediate states by limiting the number of layers to some desired number.

For other VQEs that may involve the optimization of non-diagonal Hamiltonians, the Hamiltonian itself may serve as an entanglement witness. This may be the case for Hamiltonians describing physical systems where the ground state is highly entangled, and no optimal solution can be described as a separable state or a classical bitstring. This is in stark contrast to the case of combinatorial optimization problems, where the Hamiltonians are diagonal in the computational basis, an optimal solution can be described by a single bitstring, and the cost Hamiltonian itself alone cannot certify entanglement in a state. Therefore, the construction of entanglement witnesses for combinatorial optimization problems is more non-trivial than others, but here we have shown how to construct such witnesses for QAOA-MaxCut states in particular, and for quantum algorithms solving optimization problems with \mathbb{Z}_2 symmetry more generally, as argued above.

One may construct an entanglement witness for some generic parametric state $|\psi(\vec{\theta})\rangle$ as

$$W = \max_{\vec{\theta}} \alpha \mathbb{I} - |\psi(\vec{\theta})\rangle \langle \psi(\vec{\theta})| \quad (42)$$

where α is a parameter tuned such that $\langle W \rangle$ lies on either side of an inequality for the parametric state and some other states against which we wish to certify. However, such a construction may involve exponentially many observables, estimating which may require exponentially many measurement settings.

On the other hand, Theorems 1 and 2 establish that a large class of observables serve as entanglement witnesses for *some* states, without necessarily specifying what

those state(s) might be. Measuring these observables requires measurements in only three different bases, and therefore requires trivial resources. After running any generic quantum algorithm that may produce an entangled state in the final output, one may measure the final entangled state in the X, Y and Z bases, sample a fixed number, e.g. poly(N) many, observables from the family $W_{XYZ}^{k,N,\mathcal{G}}$. If the bound in Theorem 1 is then violated, one has detected entanglement in the final state. This protocol therefore provides an algorithm agnostic method to detect entanglement, though without any guarantees that it would work for generic entangled states. However, we have shown above that a subset of the constructed observables are especially well suited to detect entanglement for QAOA-MaxCut states, and more generically for quantum states that encode exact or approximate solutions to combinatorial optimization problems with \mathbb{Z}_2 symmetry.

G. Results - Entanglement

For various numbers of qubits, we report the experimental results of measuring the value of the entanglement witness $W_{XZ}^{(2,N,\mathcal{G}_{\text{line}})}$ for linear chain $p = 1$ QAOA-MaxCut states defined over the same set of edges as the respective cost Hamiltonian. We use randomized compiling [56, 57] to twirl errors on the physical gates into stochastic errors. We fit a simple depolarizing model onto the results [58, 59], and obtain a reasonably good fit, as shown in Fig. 9. More details are provided in Appendix A.

In general, noise will reduce the chances of detecting entanglement. In the case of a global depolarizing noise channel

$$\rho \rightarrow (1 - p_{\text{noise}})\rho + p_{\text{noise}} \frac{\mathbb{I}}{2^N} \quad (43)$$

with noise parameter p_{noise} , the expectation value in the noisy mixed state scales down as $(1 - p_{\text{noise}})\langle W \rangle$, where $\langle W \rangle$ represents the expectation value in the ideal pure QAOA state. Experimentally, we find a violation of the separable threshold at $N = 2$ for the observables $W_{XZ}^{(2,2,\mathcal{G})}$ and $W_{XY}^{(2,2,\mathcal{G})}$ as depicted in Fig. 10 (a similar violation was recently reported experimentally in [60]), we failed to find any such violations for $N \geq 3$. In Appendix A, we provide the values of the noise parameters that were obtained from fitting a global depolarizing channel to the data. It is found that for $N \geq 3$, these are above the critical threshold of the depolarizing noise parameter that one obtains from Eq. (24)

$$p_{\text{noise}} \lesssim 1 - \frac{N - 1}{1.207N - 1.019} \quad (44)$$

If the noise level on the hardware is below the threshold defined by Eq. (44) for $N \geq 3$, then one would be able

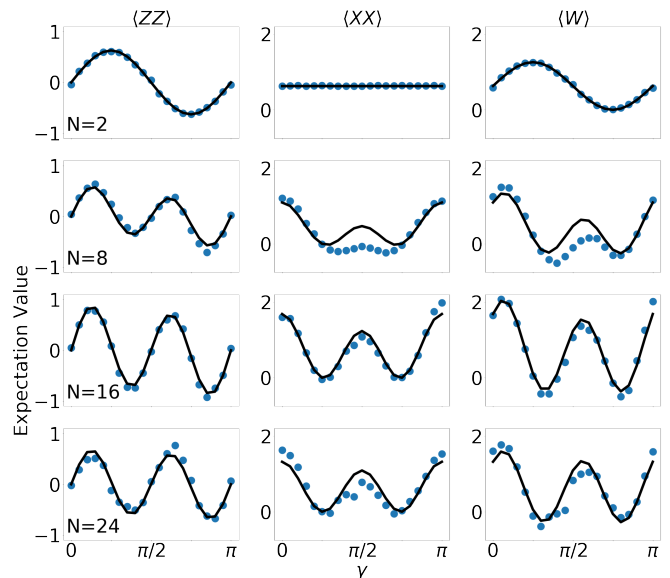


FIG. 9. Experimental results for a linear chain $W = W_{XZ}^{(2,N,\mathcal{G}_{\text{linear}})} = \sum_i^{N-1} X_i X_{i+1} + Z_i Z_{i+1}$ witness observable from the Rigetti Aspen-9 QPU. For $N = 2, 8, 16, 24$ qubits (top-bottom), we plot the expectation values of the sum of all ZZ terms (left-most column), XX terms (middle column), and the witness observable W (right-most column). We overlay the experimental data points with a fitted model that is given by some scale factor $(1 - p_{\text{noise}})$ times the analytical expression Eq. (23).

to certify entanglement in the linear chain $p = 1$ QAOA-MaxCut state. For $N = 2$, the noise on the hardware was sufficiently below the critical threshold $p_{\text{noise}} < \frac{1}{2}$ to enable the witness to detect entanglement in the QAOA state.

IV. CONCLUSIONS

In this paper, we have introduced a practical method to verify key non-classical properties of a quantum algorithm's implementation on a physical device. Using measurements in only the three Pauli bases, we reconstruct the experimental single-qubit reduced density matrices (SQRDMs), and interpret their coherence as a basic measure of non-classicality, or quantumness. We identified a large family of observables, that can serve as entanglement witnesses and that could be measured using the same measurement data. Although our work focuses on QAOA-MaxCut, the same procedure could in principle be used to test the entanglement of any other state prepared on a (noisy) quantum device – since Theorem 1 provides both upper and lower bounds for separable states, the experimentalist simply has to collect bitstrings in the X, Y and Z bases, compute the expectation value for a suitable observable and check if either the upper or lower bound is violated. Given the measurement data, estimating the expectation value of these observables is at

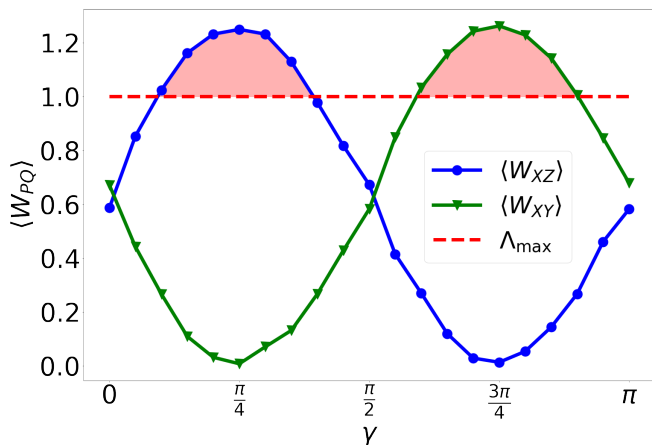


FIG. 10. Expectation value of two observables: (green, wedges) $W_{XY} = W_{XY}^{(2,2,\mathcal{G}_{\text{linear}})}$ and (blue dots) $W_{XZ} = W_{XZ}^{(2,2,\mathcal{G}_{\text{linear}})}$ on Rigetti Aspen-9 chip for QAOA $p = 1$ Max-Cut problem (see Appendix A for details of the experimental setup) as a function of γ angle ($\beta = \frac{\pi}{8}$). Red area indicates region of violation of the separable bound $\Lambda_{\max} = 1$ (red dashed line).

most polynomial in the number of qubits, and therefore efficient.

Our work has also proposed a generalization of Bell-type observables, for which we established a non-trivial inequality. The generalized Cauchy-Schwarz inequality we used to prove one of our main results may have independent interest in other areas of quantum information, computer science, and related fields. We showed that entanglement witnesses for variational circuits can be constructed out of the cost Hamiltonian itself, typically via additional terms that are equivalent to the cost Hamiltonian up to local transformations, without resorting to decomposing the projector of the entangled state into a possibly exponentially large set of measurable operators. This may inspire the construction of yet more witnesses in future work. In particular, we imagine that the techniques we have outlined in this paper could provide a foundation to the identification of similar witnesses in other parametric families of circuits, such as those found in Variational Quantum Eigensolvers (VQE) or Quantum Machine Learning (QML), prevalent in the NISQ era.

For QAOA-MaxCut in particular, we noted that the cost Hamiltonian and the corresponding 2-local witness defined on the same graph saturate in maximum expectation value at around the same p . Thus, even though further rounds of QAOA might generate more entanglement, the witness does not capture this extra entanglement. Instead, the observation that it tends to saturate around the same p hints that it may provide a measure of the amount of useful entanglement that the algorithm employs in the optimization problem. In general, it is desirable to obtain measures of and verification procedures for the amount of entanglement that is algorithmically relevant, and our work provides a step in that direction.

As gate fidelities and qubit counts and connectivities on near-term devices improve, it will become increasingly important to verify that the hardware resources are employing genuinely non-classical resources in the execution of some algorithm, without which a quantum advantage would be impossible. Additionally, identifying the entanglement properties in quantum circuits can help us in understanding the role of entanglement as a resource for quantum computing.

ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under agreement No. HR00112090058 and IAA 8839. F.W., J.S., Z.W., P.A.L. and D.V. also acknowledge USRA NASA Academic Mission Services (contract NNA16BD14C). The authors wish to thank Stuart Hadfield and Jeffrey Marshall for useful comments and feedback, as well as the Quantum Benchmark Inc. team for support on error modeling and mitigation. The experimental results were made possible by contributions across the Rigetti enterprise, and we especially acknowledge effort from Alex Hill, Nicolas Didier, Joseph Valery, and Greg Stiehl.

Appendix A: Experimental details

Experimental results were obtained on the 32 qubit Rigetti Aspen-9 system using a QAOA $p = 1$ circuit ansatz shown in Fig. 11. Our implementation of the linear chain topology achieved constant circuit-depth by construction with parallel gates. Moreover, this approach enabled a systematic investigation, since adding new elements to the chain did not disrupt the prior circuit design. Median average gate fidelity for the CZ gate across the 24 qubit linear array under study was $F_{CZ} = 94\%$ at the time of experiments, predominantly limited by decoherence. To mitigate residual coherent error, all circuits were randomly compiled 100 times under Pauli twirling [56, 57]. Each circuit instance compiled to a random bit-flip pattern over the qubit register before measurement (later undone in post-processing) to remove readout bias. Each unique circuit was executed 300 times.

After fitting a global depolarizing model as described in Section III G, we find that to 3 decimal places, $p_{\text{noise}} \approx 0.370$ for $N = 2$, which is less than the critical threshold of $p_{\text{noise}} < \frac{1}{2}$, so that we observe a violation of the separable threshold at $N = 2$. At $N = 8, 16$ and 24 , to 3 decimal places we obtain $p_{\text{noise}} \approx 0.844, 0.888$ and 0.943 respectively. These are all above the critical thresholds of $p_{\text{critical}} \approx 0.190, 0.180$ and 0.177 (to 3 decimal places, for $N = 8, 16$ and 24 respectively) defined by Eq. (44) so that for these values of N , we do not observe a violation of the separable threshold.

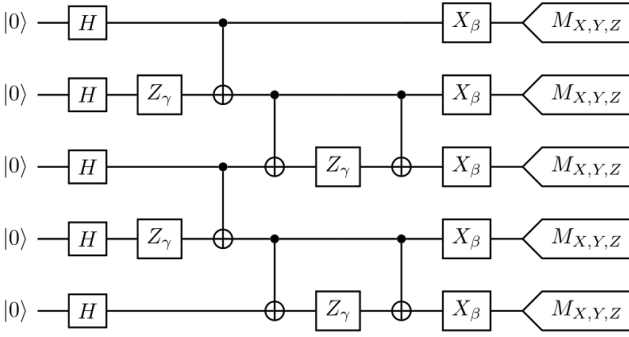


FIG. 11. Example circuit for QAOA, $p = 1$ circuit ansatz ($N=5$), optimized for parallel execution on a linear chain topology at arbitrary size. CNOT gates are compiled to native CZ gates. Single-qubit rotations are compiled to continuous-angle Z_ϕ and fixed-angle X_{90} gates. Measurements ($M_{X,Y,Z}$) are collected for three axes (X, Y, Z) using tomographic pre-rotations and Z -basis measurements.

In Fig. 12, we plot the coherences for some fixed value of $\beta = \frac{\pi}{8}$ vs. γ ($\gamma = k\frac{\pi}{20}$ for $k = 0, 1, \dots, 20$). In the same figure, we additionally report average (middle semicircle), minimal (inner semicircle), and maximal (outer semicircle) coherences over all active qubits for each γ . One can notice, that Rigetti Aspen-9 chip can build non-zero signatures of quantumness over all qubits across various problems of different size, subjected to non-negligible noise.

Appendix B: General observables

Here, we describe generalizations of the observables $W_{PQ}^{(k,N,\mathcal{G})}$ and $W_{XYZ}^{(k,N,\mathcal{G})}$ defined in Eqs. (11) and (12) respectively, to which Theorem 1 applies. We consider observables of the form

$$W_M^{(k,N,\mathcal{G})} = \sum_{\langle i_1, \dots, i_k \rangle} \sum_{m=1}^M \alpha_m \bigotimes_{j=1}^k \sigma_{i_j}^{(a_{j,m})} \quad (\text{B1})$$

where we restrict to $M \leq 3$, and

1. $\alpha_m \in \{0, 1\}$ for all $1 \leq m \leq M = 3$,
2. the indices $a_{j,m} \in \{1, 2, 3\}$ specify which single-qubit (non-identity) Pauli operator acts on the j -th qubit i_j in the m -th term,
3. $a_{j,1} \neq a_{j,2} \neq a_{j,3}$ for each j ,

and the superscript \mathcal{G} refers to the generalized graph, i.e. the set of k -tuples $\langle i_1, \dots, i_k \rangle$ being summed over. All observables of the form Eqs. (11) and (12) belong to this family. Note that in those expressions, we use the subscript to denote a Pauli string, not a numerical value for M . We choose this convention to simplify our expressions whenever the indices $a_{j,m}$ are fixed for all values of j once m is specified. The number of characters in

the subscript string then specify the number of non-zero coefficients α_m . Thus, $W_{XZ}^{(k,N,\mathcal{G})}$ is really an observable of the form $W_2^{(k,N,\mathcal{G})}$ where $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = 0$ and we have fixed $a_{j,1} = 1$ and $a_{j,2} = 3$. Similarly, $W_{XYZ}^{(k,N,\mathcal{G})}$ is an observable of the form $W_3^{(k,N,\mathcal{G})}$ where $\alpha_m = 1$ and we have fixed $a_{j,m} = m$.

Using the same convention, we can also build other observables such as

$$W_{XY}^{(k,N,\mathcal{G})} = \sum_{\langle i_1, \dots, i_k \rangle} \left(\bigotimes_{j=i_1}^{i_k} X_j + \bigotimes_{j=i_1}^{i_k} Y_j \right), \quad (\text{B2})$$

which is an observable of the form $W_2^{(k,N,\mathcal{G})}$ with $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = 0$ and $a_{j,1} = 1$, $a_{j,2} = 2$ for all j . Another such example is

$$W_{YZ}^{(k,N,\mathcal{G})} = \sum_{\langle i_1, \dots, i_k \rangle} \left(\bigotimes_{j=i_1}^{i_k} Y_j + \bigotimes_{j=i_1}^{i_k} Z_j \right) \quad (\text{B3})$$

which is again an observable of the form $W_2^{(k,N,\mathcal{G})}$ with $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = 0$ and $a_{j,1} = 2$, $a_{j,2} = 3$ for all j . We refer to any observable of the form $W_{XZ}^{(k,N,\mathcal{G})}$, $W_{XY}^{(k,N,\mathcal{G})}$ or $W_{YZ}^{(k,N,\mathcal{G})}$ as that of the form $W_{PQ}^{(k,N,\mathcal{G})}$. Note that $W_{ZYX}^{(k,N,\mathcal{G})}$ or $W_{YZX}^{(k,N,\mathcal{G})}$ etc. are operationally the same as $W_{XYZ}^{(k,N,\mathcal{G})}$.

However, none of these observables exploit the additional freedom allowed by Theorem 1 to choose different $a_{j,m}$ for different values of j as well as m . We can also build more complicated observables by allowing the choice of the indices $a_{j,m}$ to vary with the choice of j as well as m , as long as condition 3 above is met. For example, with $M = 2$, $k = 6$ we could construct $a_{1,1} = 1$, $a_{2,1} = 2$, $a_{3,1} = 3$, $a_{4,1} = 3$, $a_{5,1} = 2$, $a_{6,1} = 1$ and $a_{1,2} = 3$, $a_{2,2} = 1$, $a_{3,2} = 2$, $a_{4,2} = 1$, $a_{5,2} = 3$, $a_{6,2} = 2$ to give

$$W = \sum_{\langle i,j,k,l,m,n \rangle} (X_i Y_j Z_k Z_l Y_m X_n + Z_i X_j Y_k X_l Z_m Y_n) \quad (\text{B4})$$

As another example, with $M = 3$, $k = 5$ we could construct

$$W = \sum_{\langle i,j,k,l,m \rangle} (X_i Y_j Z_k X_l Y_m + Z_i X_j Y_k Y_l X_m + Y_i Z_j X_k Z_l Z_m) \quad (\text{B5})$$

and so on. Theorem 1 applies to this entire family of observables, which is much larger than those of the form in Eqs (11) and (12) alone.

Appendix C: Proofs

Proof of Lemma 1

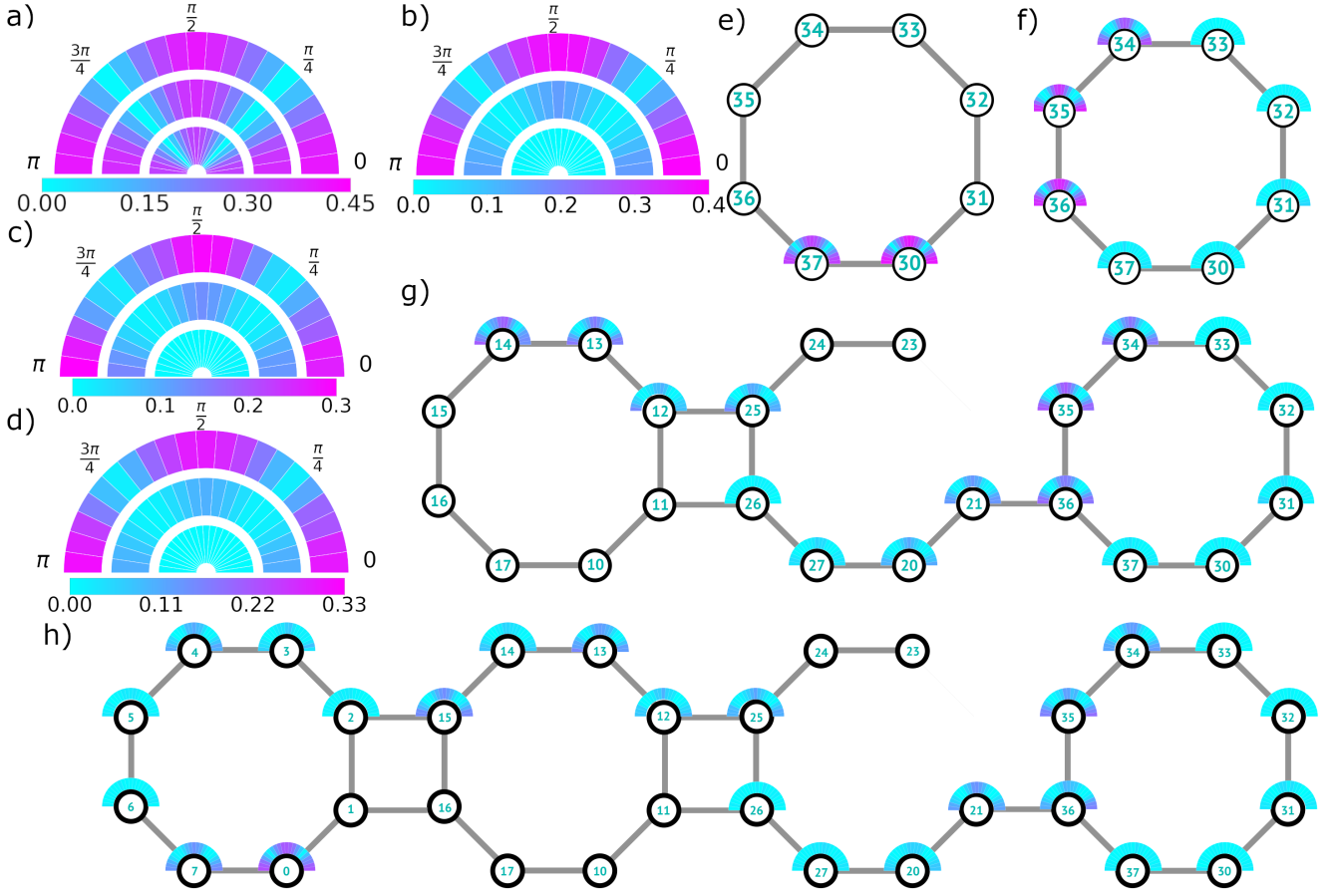


FIG. 12. Coherences of SQRDMs for MaxCut $p = 1$ QAOA run on Rigetti Aspen-9 chip (see text in Appendix A for more details on the experiment) - a) and e): 2-qubit problem; b) and f): 8-qubit; c) and g): 16-qubit, d) and h): 24-qubit . Semicircles are divided into segments, each represents γ angle ($\gamma = k\frac{\pi}{20}$ for $k = 0, 1, 2, \dots, 20$) with the coherence values color-coded (below a)-d) charts) for that QAOA setup (for all figures we used $\beta = \frac{\pi}{8}$). Figures a)-d) report maximal (outer semicircle), mean (middle semicircle) and minimal (inner semicircle) values of coherences across all used qubits. e)-h) specify active qubit's coherences and their position in Aspen-9 chip.

Proof. We will first prove the following identity involving the Hadamard product using induction

$$\left\| \odot_{j=1}^k \vec{x}^{(j)} \right\| \leq \prod_{j=1}^k \left\| \vec{x}^{(j)} \right\| \quad (\text{C1})$$

where the Hadamard product is defined as

$$\left(\odot_{j=1}^k \vec{x}^{(j)} \right)_i = \prod_{j=1}^k x_i^{(j)}. \quad (\text{C2})$$

First, we prove the base case involving two vectors \vec{a} , $\vec{b} \in \mathbb{R}^n$ and their Hadamard product $\vec{v} = \vec{a} \odot \vec{b}$. Then,

$$\begin{aligned} \|\vec{v}\| &= \sqrt{(a_1 b_1)^2 + \dots + (a_n b_n)^2} \\ &\leq \sqrt{a_1^2 \|\vec{b}\|^2 + \dots + a_n^2 \|\vec{b}\|^2} \\ &= \|\vec{a}\| \|\vec{b}\| \end{aligned} \quad (\text{C3})$$

where the inequality follows from noting that the absolute value of any component of a vector is bounded above by the (Euclidean) norm of that vector by definition, $|b_i| \leq \|\vec{b}\| = \sqrt{\sum_{j=1}^n b_j^2}$ for any $\vec{b} \in \mathbb{R}^n$ and $i \in [n]$. Next, we assume that Eq. (C1) holds true for some collection of $k-1$ vectors $\vec{x}^{(1)}, \dots, \vec{x}^{(k-1)} \in \mathbb{R}^n$, i.e. defining $\vec{v} = \odot_{j=1}^{k-1} \vec{x}^{(j)}$, we have $\left\| \odot_{j=1}^{(k-1)} \vec{x}^{(j)} \right\| = \|\vec{v}\| \leq \prod_{j=1}^{k-1} \|\vec{x}^{(j)}\|$. Then,

$$\begin{aligned} \left\| \odot_{j=1}^k \vec{x}^{(j)} \right\| &= \left\| \vec{v} \odot \vec{x}^{(k)} \right\| \\ &= \sqrt{v_1^2 (x_1^{(k)})^2 + \dots + v_n^2 (x_n^{(k)})^2} \\ &\leq \sqrt{\|\vec{v}\|^2 (x_1^{(k)})^2 + \dots + \|\vec{v}\|^2 (x_n^{(k)})^2} \\ &= \|\vec{v}\| \|\vec{x}^{(k)}\| = \prod_{j=1}^k \|\vec{x}^{(j)}\| \end{aligned} \quad (\text{C4})$$

and so Eq. (C1) follows by induction. We can now prove the main statement of our Lemma. Defining $\vec{w} = \odot_{j=1}^{m-1} \vec{x}^{(j)}$, we now have

$$\begin{aligned} \left| \sum_{i=1}^n \left(\odot_{j=1}^m \vec{x}^{(j)} \right)_i \right| &= \left| \sum_{i=1}^n \left(\vec{w} \odot \vec{x}^{(m)} \right)_i \right| \\ &= |\vec{w} \cdot \vec{x}^{(m)}| \\ &\leq \|\vec{w}\| \|\vec{x}^{(m)}\| \\ &\leq \prod_{j=1}^{m-1} \|\vec{x}^{(j)}\| \|\vec{x}^{(m)}\| \\ &= \prod_{j=1}^m \|\vec{x}^{(j)}\| \end{aligned} \quad (\text{C5})$$

which is the statement of the Lemma. In the proof above, the first inequality follows from the standard Cauchy-Schwarz inequality, while the second inequality follows from Eq. (C1). \square

Proof of Theorem 1

Proof. Given a pure state $|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle$, we have

$$\begin{aligned} \langle X \rangle_\psi &= \sin \theta \cos \phi \\ \langle Y \rangle_\psi &= \sin \theta \sin \phi \\ \langle Z \rangle_\psi &= \cos \theta \end{aligned} \quad (\text{C6})$$

A fully separable pure state is given by $|\Psi\rangle = \otimes_{j=1}^N \left(\cos \frac{\theta_j}{2} |0\rangle + e^{i\phi_j} \sin \frac{\theta_j}{2} |1\rangle \right)$. For $M = 1$ alone, it is easy to prove the the upper bound of the absolute expectation values in such states, as follows

$$\begin{aligned} \left| \langle W_1^{(k,N,\mathcal{G})} \rangle \right| &= \left| \sum_{\langle i_1, \dots, i_k \rangle} \prod_{j=1}^k \langle \sigma_{i_j}^{(a_{j,1})} \rangle \right| \\ &\leq \left| \sum_{\langle i_1, \dots, i_k \rangle} \prod_{j=1}^k (1) \right| = |E_k| \end{aligned} \quad (\text{C7})$$

For $1 \leq M \leq 3$, we note that $\langle W_M^{(k,N,\mathcal{G})} \rangle$ can be written as a Hadamard product of k vectors

$$\begin{aligned} \langle W_M^{(k,N,\mathcal{G})} \rangle &= \sum_{\langle i_1, \dots, i_k \rangle} \sum_{m=1}^M \prod_{j=1}^k \langle \sigma_{i_j}^{(a_{j,m})} \rangle \\ &= \sum_{\langle i_1, \dots, i_k \rangle} \sum_{m=1}^M \left(\odot_{j=1}^k \vec{v}_{i_j}^{(M)} \right)_m \end{aligned} \quad (\text{C8})$$

where $\vec{v}_{i_j}^{(M)} = \left(\langle \sigma_{i_j}^{(a_{j,1})} \rangle, \dots, \langle \sigma_{i_j}^{(a_{j,M})} \rangle \right) \in \mathbb{R}^M$ (since expectations of Pauli operators are real) where j labels some index in a k -tuple. Note that $\|\vec{v}_{i_j}^{(1)}\| \leq 1$ since the maximum (absolute) expectation value of any Pauli operator is 1. Next, since $\forall j, a_{j,1} \neq a_{j,2} \neq a_{j,3}$, we have

that $\|\vec{v}_{i_j}^{(2)}\|$ is the square root of the sum of squares of $\langle X \rangle$ and $\langle Y \rangle$, $\langle X \rangle$ and $\langle Z \rangle$, or $\langle Y \rangle$ and $\langle Z \rangle$. All of these cases are bounded above by $\|\vec{v}_{i_j}^{(3)}\|$. Furthermore,

$$\begin{aligned} \|\vec{v}_{i_j}^{(3)}\| &= \sqrt{\sin^2 \theta_{i_j} \cos^2 \phi_{i_j} + \sin^2 \theta_{i_j} \sin^2 \phi_{i_j} + \cos^2 \theta_{i_j}} \\ &= 1 \end{aligned} \quad (\text{C9})$$

so that we have $\|\vec{v}_{i_j}^{(M)}\| \leq 1$ for any $M \in \{1, 2, 3\}$. Therefore, we obtain

$$\begin{aligned} \left| \langle W_M^{(k,N,\mathcal{G})} \rangle \right| &= \left| \sum_{\langle i_1, \dots, i_k \rangle} \sum_{m=1}^M \left(\odot_{j=1}^k \vec{v}_{i_j}^{(M)} \right)_m \right| \\ &\leq \sum_{\langle i_1, \dots, i_k \rangle} \left| \sum_{m=1}^M \left(\odot_{j=1}^k \vec{v}_{i_j}^{(M)} \right)_m \right| \\ &\leq \sum_{\langle i_1, \dots, i_k \rangle} \prod_{j=1}^k \|\vec{v}_{i_j}^{(M)}\| \\ &\leq \sum_{\langle i_1, \dots, i_k \rangle} \prod_{j=1}^k (1) = |E_k| \end{aligned} \quad (\text{C10})$$

where the first inequality follows from repeated use of the triangle inequality $|a+b| \leq |a| + |b|$ for any $a, b \in \mathbb{R}$, the second inequality follows from Lemma 1 and the last inequality follows from the bound $\|\vec{v}_{i_j}^{(M)}\| \leq 1$ as noted above. Next, suppose we are given some fully separable mixed state

$$\rho = \sum_j \alpha_j \rho_{j_1} \otimes \dots \otimes \rho_{j_N} \quad (\text{C11})$$

where each $\alpha_j > 0$ and $\sum_j \alpha_j = 1$. Each of the density operators in the tensor product can be expressed as

$$\rho_{j_k} = \sum_l c_{j_k,l} |\psi_{j_k,l}\rangle \langle \psi_{j_k,l}| \quad (\text{C12})$$

with $c_{j_k,l} > 0$ and $\sum_l c_{j_k,l} = 1$. Therefore, by absorbing these constants into the definition of the α_j 's, we have

$$\rho = \sum_j \alpha_j |\psi_{j_1}\rangle \langle \psi_{j_1}| \otimes \dots \otimes |\psi_{j_N}\rangle \langle \psi_{j_N}| \quad (\text{C13})$$

and defining the separable pure states $|\Psi_j\rangle \langle \Psi_j| = \otimes_{m=1}^N |\psi_m\rangle \langle \psi_m|$, we have

$$\begin{aligned} \left| \text{Tr} \left(\rho W_M^{(k,N,\mathcal{G})} \right) \right| &= \left| \text{Tr} \left(\sum_j \alpha_j |\Psi_j\rangle \langle \Psi_j| \cdot W_M^{(k,N,\mathcal{G})} \right) \right| \\ &= \left| \sum_j \alpha_j \text{Tr} \left(|\Psi_j\rangle \langle \Psi_j| \cdot W_M^{(k,N,\mathcal{G})} \right) \right| \\ &\leq \sum_j \alpha_j \left| \text{Tr} \left(|\Psi_j\rangle \langle \Psi_j| \cdot W_M^{(k,N,\mathcal{G})} \right) \right| \\ &\leq \sum_j \alpha_j \cdot |E_k| = |E_k| \end{aligned} \quad (\text{C14})$$

where the first inequality follows from repeated use of the triangle inequality, and the last inequality follows from Eq. (C10) for pure states. \square

Proof of Theorem 2

Proof. We will first prove the theorem for the case of $W_{XZ}^{(k,N,\mathcal{G})}$. One way to lower bound the largest eigenvalue of $W = \sum_{\langle i_1, \dots, i_k \rangle} \bigotimes_{j=i_1}^{i_k} X_j + \bigotimes_{j=i_1}^{i_k} Z_j$ is to simply compute its expectation value in some state. A convenient choice is $|+\rangle^{\otimes N}$. In this state, we find $\langle W \rangle = |E_k|$, therefore $\lambda_{max} \geq |E_k|$. However, we would like to increase this lower bound, since $|E_k|$ is met by separable states. In [61], a sharper lower bound for the largest eigenvalue of any $m \times m$ symmetric matrix is provided as

$$\lambda_{max} \geq \frac{N_1}{m} + 2 \left(\frac{N_3}{2m} - \frac{N_1 N_2}{m^2} + \frac{N_1^3}{2m^3} \right) \lambda_0^{-2} + O(t^{-4}) \quad (\text{C15})$$

where $u = [1 \ 1 \ \dots \ 1]^T$ is the $m \times 1$ all-ones vector, $N_k = u^T A^k u$, $B = \frac{1}{\sqrt{m}} \max_{1 \leq j \leq m} (W_{jj} + \sum_{i=1; i \neq j}^m |W_{ij}|)$, $\lambda_0 = t\sqrt{m}$ and t is any real number satisfying $t \geq B$. In our case, we have $m = 2^N$.

We already noted that $N_1/m = |E_k|$. Now,

$$\begin{aligned} W^2 &= \left(\sum_{\langle i_1, \dots, i_k \rangle} \left(\bigotimes_{m=i_1}^{i_k} X_m + \bigotimes_{m=i_1}^{i_k} Z_m \right) \right)^2 \\ &= \sum_{\langle i_1, \dots, i_k \rangle} \sum_{\langle j_1, \dots, j_k \rangle} (X_{i_1} \dots X_{i_k} X_{j_1} \dots X_{j_k} + \\ &\quad X_{i_1} \dots X_{i_k} Z_{j_1} \dots Z_{j_k} + Z_{i_1} \dots Z_{i_k} X_{j_1} \dots X_{j_k} \\ &\quad + Z_{i_1} \dots Z_{i_k} Z_{j_1} \dots Z_{j_k}) \end{aligned} \quad (\text{C16})$$

Under expectation in the $|+\rangle^{\otimes N}$ state, the first term becomes $\left(\sum_{\langle i_1, \dots, i_k \rangle} \langle X_{i_1} \rangle \dots \langle X_{i_k} \rangle \right)^2 = |E_k|^2$. Both terms involving a product of $X \dots X$ and $Z \dots Z$ terms are zero under expectation, since either the Z_i terms will survive under the product, or multiply with an X_i to produce a Y_i , and since $\langle +|Z|+ \rangle = \langle +|Y|+ \rangle = 0$, the entire term vanishes in expectation. The last term $Z_{i_1} \dots Z_{i_k} Z_{j_1} \dots Z_{j_k}$ can only contribute whenever $\langle i_1, \dots, i_k \rangle$ and $\langle j_1, \dots, j_k \rangle$ specify the same k -tuple, so that $\left\langle \sum_{\langle i_1, \dots, i_k \rangle} \sum_{\langle j_1, \dots, j_k \rangle} Z_{i_1} \dots Z_{i_k} Z_{j_1} \dots Z_{j_k} \right\rangle = |E_k|$. Altogether, this gives us

$$\frac{N_2}{m} = \langle W^2 \rangle = |E_k|^2 + |E_k| \quad (\text{C17})$$

To compute the N_3 term, we observe that W^3 consists of four types of terms:

1. One term of the form $X_{i_1} \dots X_{i_k} X_{j_1} \dots X_{j_k} X_{l_1} \dots X_{l_k}$, which under expectation give a total contribution of $|E_k|^2$, following the same reasoning as in the computation of N_2/m above,

2. Three terms of the form $(X_{i_1} \dots X_{i_k})^2 (Z_{i_1} \dots Z_{i_k})$, each of which give a contribution of 0 under expectation, since the Z_m 's have no choice but to either persist or multiply with an X_m to give Y_m , and $\langle Z_m \rangle = \langle Y_m \rangle = 0$,

3. Three terms of the form $(X_{i_1} \dots X_{i_k}) (Z_{i_1} \dots Z_{i_k})^2$, each of which give a contribution of $|E_k|^2$ since $\sum_{i_1, \dots, i_k} \langle X_{i_1} \dots X_{i_k} \rangle = |E_k|$, and $\left\langle \left(\sum_{\langle i_1, \dots, i_k \rangle} Z_{i_1} \dots Z_{i_k} \right)^2 \right\rangle = |E_k|$ for the same reason we noted earlier in the computation of N_2/m ,

4. One term of the form $Z_{i_1} \dots Z_{i_k} Z_{j_1} \dots Z_{j_k} Z_{l_1} \dots Z_{l_k}$, which has non-trivial contributions from triangles within a graph, and is generally tedious to compute.

Even without explicitly calculating this last term, we have

$$\frac{N_3}{m} = \langle W^3 \rangle \geq |E_k|^3 + 3|E_k|^2 \quad (\text{C18})$$

Combining these results, we have

$$\begin{aligned} &\frac{N_3}{2m} - \frac{N_1 N_2}{m^2} + \frac{N_1^3}{2m^3} \\ &\geq \frac{|E_k|^3 + 3|E_k|^2}{2} - |E_k|^3 - |E_k|^2 + \frac{|E_k|^3}{2} \\ &= \frac{|E_k|^2}{2} > 0 \end{aligned} \quad (\text{C19})$$

Letting $t = \frac{\sqrt{m}}{2} B = |E_k|$, we have $t \geq B$ for all $N \geq 2$, and $\lambda_0 = t\sqrt{m} = |E_k| \sqrt{2^N}$. Plugging these values into Eq. (C15), we find that

$$2 \left(\frac{N_3}{2m} - \frac{N_1 N_2}{m^2} + \frac{N_1^3}{2m^3} \right) \lambda_0^{-2} \geq \frac{1}{2^N} > 0$$

for any finite N , so that

$$\begin{aligned} \lambda_{max} &\geq |E_k| + \frac{1}{2^N} + O\left(\frac{1}{|E_k|^4}\right) \\ &> |E_k| + O\left(\frac{1}{|E_k|^4}\right) \end{aligned} \quad (\text{C20})$$

which shows that the theorem holds for $W_{XZ}^{(k,N,\mathcal{G})}$. Noting that the eigenspectrum of any observable A remains invariant under a unitary transformation $U^\dagger A U$, we can perform single-qubit rotations $U = \bigotimes_{q=1}^N R_X^{(q)}(\pi/2)$ (where $R_X^{(q)}(\alpha) = \exp(-i\alpha X_q)$) at every qubit to have Eq. (C20) apply to observables of the form $W_{XY}^{(k,N,\mathcal{G})}$ as well. Similarly, choosing $U = \bigotimes_{q=1}^N R_Z^{(q)}(\pi/2)$ yields the same bound for observables of the form $W_{YZ}^{(k,N,\mathcal{G})}$. This exhausts all the possibilities for $W_{PQ}^{(k,N,\mathcal{G})}$, and the theorem holds. \square

Proof of Lemma 3

Proof. These expressions can be derived using essentially the procedure described in [62]. In that paper, the authors derive the expression for $\langle C_{uv} \rangle$, where $C_{uv} = \frac{1}{2}(\mathbb{I} - Z_u Z_v) \rightarrow \langle Z_u Z_v \rangle = 1 - 2\langle C_{uv} \rangle$, and where the QAOA state is built from the cost Hamiltonian $C' = \frac{1}{2} \sum_{\langle u,v \rangle} (\mathbb{I} - Z_u Z_v)$. The expression they derive for $\langle Z_u Z_v \rangle$ matches our expression for $\langle Z_u Z_v \rangle$ as long as one adjusts for the difference in convention by replacing $\gamma \rightarrow 2\gamma$ and $\langle Z_u Z_v \rangle \rightarrow 2\langle Z_u Z_v \rangle$ in going from the notation of [62] to our expression.

The expression for $\langle X_u X_v \rangle$ is derived similarly in the Heisenberg picture. We first note that $e^{i\beta B} X_u X_v e^{-i\beta B} = X_u X_v$ for $B = \sum_{i=1}^N X_i$, so that

$$\begin{aligned} \langle X_u X_v \rangle &= \text{Tr}(e^{i\gamma C} e^{i\beta B} X_u X_v e^{-i\beta B} e^{-i\gamma C} \rho_0) \\ &= \text{Tr}(e^{i\gamma C} X_u X_v e^{-i\gamma C} \rho_0) \end{aligned} \quad (\text{C21})$$

where $\rho_0 = \prod_{q=1}^N \frac{1}{2}(\mathbb{I} + X_q)$ is the initial state. Further, since $[Z_u Z_v, X_u X_v] = 0$ and $[Z_i Z_j, X_u X_v] = 0$ whenever $i \neq u, j \neq v$, the only terms that survive are those involving u (v) and its neighbors excluding v (u). Denoting the number of such neighbors as d_u and d_v respectively for u and v , and defining $C_u = \sum_{j=1}^{d_u} Z_u Z_{a_j}$ and $C_v = \sum_{i=1}^{d_v} Z_v Z_{b_i}$, we have

$$\begin{aligned} &\text{Tr}(e^{i\gamma C} X_u X_v e^{-i\gamma C} \rho_0) \\ &= \text{Tr}(e^{i\gamma C_u} e^{i\gamma C_v} X_u X_v e^{-i\gamma C_u} e^{-i\gamma C_v} \rho_0) \\ &= \text{Tr}(e^{i2\gamma C_v} e^{i2\gamma C_u} X_u X_v \rho_0) \end{aligned} \quad (\text{C22})$$

where in the last line we have used the fact that $X_m Z_m = -Z_m X_m$ to move the $e^{-i\gamma C_u}$ and $e^{-i\gamma C_v}$ terms past $X_u X_v$. Now, $\text{Tr}(A_u B_v \rho_0) = 1$ only when $A, B \in \{\mathbb{I}, X\}$, so that the only terms in the product $e^{i2\gamma C_v} e^{i2\gamma C_u}$ that can contribute non-trivially are those that are proportional to the identity. Therefore, using $e^{i\alpha Z_a Z_b} = \cos \alpha \mathbb{I} + i \sin \alpha Z_a Z_b$, we see that for triangle-free edges,

$$\langle X_u X_v \rangle = \cos^{d_u+d_v} 2\gamma \quad (\text{C23})$$

which agrees with the expression in Lemma 3 for the case $f = 0$. In the more general case, only terms in the expansion that are proportional to even powers of Z_u and Z_v (and therefore the identity) contribute non-trivially. We then have

$$\begin{aligned} &\langle X_u X_v \rangle \\ &= \sum_{j=0,2,4,\dots}^f \binom{f}{j} (\cos 2\gamma)^{d_u+d_v-2j} (i \sin 2\gamma)^{2j} \\ &= \cos^{d_u+d_v-2f} 2\gamma \\ &\quad \times \sum_{j=0,2,4,\dots}^f \binom{f}{j} (\cos^2 2\gamma)^{f-j} (-\sin^2 2\gamma)^j \end{aligned} \quad (\text{C24})$$

and using the identity

$$\sum_{j=0,2,4,\dots}^f \binom{f}{j} a^{f-j} b^j = \frac{1}{2} [(a+b)^f + (a-b)^f] \quad (\text{C25})$$

we finally obtain

$$\langle X_u X_v \rangle = \frac{1}{2} \cos^{d_u+d_v-2f} 2\gamma (1 + \cos^f 4\gamma) \quad (\text{C26})$$

as stated in the Lemma.

To derive the expression for $\langle Y_u Y_v \rangle$, we make use of the relation $SWAP_{uv} = \frac{1}{2}(X_u X_v + Y_u Y_v + Z_u Z_v + \mathbb{I})$ and derive the expectation value for $SWAP_{uv}$ first. Notice that $[Y_u Y_v, B] = -[Z_u Z_v, B]$ and $[X_u X_v, B] = [\mathbb{I}, B] = 0$, so that $[SWAP_{uv}, B] = 0$, and therefore the mixing unitaries do not contribute to the expectation value:

$$\begin{aligned} \langle SWAP_{uv} \rangle &= \text{Tr}(e^{i\gamma C} e^{i\beta B} X_u X_v e^{-i\beta B} e^{-i\gamma C} \rho_0) \\ &= \text{Tr}(e^{i\gamma C} SWAP_{uv} e^{-i\gamma C} \rho_0) \end{aligned} \quad (\text{C27})$$

Inside the trace, only terms involving u or v contribute, and by making use of $SWAP_{uv} e^{i\gamma Z_u Z_{u'}} SWAP_{uv} = e^{i\gamma Z_v Z_{v'}}$, we have

$$\begin{aligned} &\langle SWAP_{uv} \rangle \\ &= \text{Tr}\left(e^{i\gamma \sum_{u' \in \mathcal{N}(u)} Z_u Z_{u'}} e^{i\gamma \sum_{v' \in \mathcal{N}(v)} Z_v Z_{v'}}\right. \\ &\quad \left. SWAP_{uv} e^{-i\gamma \sum_{u' \in \mathcal{N}(u)} Z_u Z_{u'}} e^{-i\gamma \sum_{v' \in \mathcal{N}(v)} Z_v Z_{v'}} \rho_0\right) \\ &= \text{Tr}\left(e^{i\gamma(Z_u - Z_v)(\sum_{u' \in \mathcal{N}'(u)} Z_{u'} - \sum_{v' \in \mathcal{N}'(v)} Z_{v'})} SWAP_{uv} \rho_0\right) \\ &= \text{Tr}\left(e^{i\gamma(Z_u - Z_v)(\sum_{u' \in \mathcal{N}'(u)} Z_{u'} - \sum_{v' \in \mathcal{N}'(v)} Z_{v'})} \rho_0\right) \end{aligned} \quad (\text{C28})$$

where $\mathcal{N}(u)$ is the set of neighbors of u , $\mathcal{N}'(u)$ is the set of nodes that are neighbors of u but not v or neighbors of v , with $\mathcal{N}'(v)$ defined similarly, and the last step is due to $SWAP_{uv} \rho_0 = \rho_0$. In the second equality above, we have made use of the fact that $\sum_{u' \in \mathcal{N}(u)} Z_{u'} - \sum_{v' \in \mathcal{N}(v)} Z_{v'} = \sum_{u' \in \mathcal{N}'(u)} Z_{u'} - \sum_{v' \in \mathcal{N}'(v)} Z_{v'}$. Since the exponential contributes either identity or Z 's, terms in ρ_0 that have X trace to zero, and we further have

$$\begin{aligned} &\langle SWAP_{uv} \rangle \\ &= \frac{1}{2^N} \text{Tr}\left(e^{i\gamma(Z_u - Z_v)(\sum_{u' \in \mathcal{N}'(u)} Z_{u'} - \sum_{v' \in \mathcal{N}'(v)} Z_{v'})}\right) \\ &= \frac{1}{2^N} \times \\ &\quad \text{Tr}\left(\prod_{u' \in \mathcal{N}'(u)} (\cos \gamma + i \sin \gamma Z_u Z_{u'}) (\cos \gamma - i \sin \gamma Z_v Z_{u'})\right. \\ &\quad \left.\prod_{v' \in \mathcal{N}'(v)} (\cos \gamma - i \sin \gamma Z_u Z_{v'}) (\cos \gamma + i \sin \gamma Z_v Z_{v'})\right) \end{aligned} \quad (\text{C29})$$

There are two types of products that come from the above expansion. The first kind are those in which terms from

both the $\mathcal{N}'(u)$ and $\mathcal{N}'(v)$ products are proportional to the identity, which give the following contribution

$$\begin{aligned} & \sum_{i=0,2,4,\dots}^{d'_u} \binom{d'_u}{i} (\cos^2 \gamma)^{d'_u-i} (\sin^2 \gamma)^i \\ & \times \sum_{j=0,2,4,\dots}^{d'_v} \binom{d'_v}{j} (\cos^2 \gamma)^{d'_v-j} (\sin^2 \gamma)^j \\ & = \frac{1}{4} \left(1 + \cos^{d'_u} 2\gamma\right) \left(1 + \cos^{d'_v} 2\gamma\right) \end{aligned} \quad (\text{C30})$$

where $d'_u = |\mathcal{N}'(u)| = d_u - f$, and d'_v is defined similarly, and we have used Eq. (C25). The second kind of terms are those in which terms from both the $\mathcal{N}'(u)$ and $\mathcal{N}'(v)$ products are proportional to $Z_u Z_v$, so that the overall product is still proportional to the identity. Terms like these give the following overall contribution

$$\begin{aligned} & \sum_{i=1,3,5,\dots}^{d'_u} \binom{d'_u}{i} (\cos^2 \gamma)^{d'_u-i} (\sin^2 \gamma)^i \\ & \times \sum_{j=1,3,5,\dots}^{d'_v} \binom{d'_v}{j} (\cos^2 \gamma)^{d'_v-j} (\sin^2 \gamma)^j \\ & = \frac{1}{4} \left(1 - \cos^{d'_u} 2\gamma\right) \left(1 - \cos^{d'_v} 2\gamma\right) \end{aligned} \quad (\text{C31})$$

where we have used the identity

$$\sum_{j=1,3,5,\dots}^f \binom{f}{j} a^{f-j} b^j = \frac{1}{2} [(a+b)^f - (a-b)^f] \quad (\text{C32})$$

Adding Eqs. (C30) and (C31), we get the total contribution

$$\langle SWAP_{uv} \rangle = \frac{1}{2} \left(1 + \cos^{d_u+d_v-2f} 2\gamma\right) \quad (\text{C33})$$

Using the earlier expressions for $\langle X_u X_v \rangle$ and $\langle Z_u Z_v \rangle$, it follows that the expression for $\langle Y_u Y_v \rangle$ is as given in the Lemma. \square

Appendix D: Properties of $W_{XYZ}^{(N,N,\mathcal{G})}$ and $W_{PQ}^{(N,N,\mathcal{G})}$

Here we present properties of the $W_{XYZ}^{(N,N,\mathcal{G})}$ observable (which can later be used to infer those of $W_{PQ}^{(N,N,\mathcal{G})}$). In particular, we show that some of the GHZ-type states yield expectation values of ± 3 for an even number of qubits, and that they are eigenstates of the considered observable. Additionally, it follows from Theorem 1 that separable states yield expectation values that are bounded by ± 1 .

First, let's calculate the expectation value with respect to the *standard* incarnation of an N -qubit GHZ state

$$|\psi\rangle = \frac{|00\dots 0\rangle + |11\dots 1\rangle}{\sqrt{2}} = \frac{|\bar{0}\rangle + |\bar{1}\rangle}{\sqrt{2}}, \quad (\text{D1})$$

where we used $|\bar{k}\rangle = |kk\dots k\rangle$ for brevity of notation. The expectation value of $W_{XYZ}^{(N,N,\mathcal{G})}$ with respect to $|\psi\rangle$ is

$$\langle \psi | W_{XYZ}^{(N,N,\mathcal{G})} | \psi \rangle = \frac{3}{2} + \frac{(-1)^N}{2} + \frac{1}{2} ((-i)^N + (i)^N), \quad (\text{D2})$$

where we used

$$\begin{aligned} \langle \bar{0} | Z^{\otimes N} | \bar{0} \rangle &= 1, & \langle \bar{1} | Z^{\otimes N} | \bar{1} \rangle &= (-1)^N, & \langle \bar{0} | Z^{\otimes N} | \bar{1} \rangle &= 0, \\ \langle \bar{1} | Z^{\otimes N} | \bar{0} \rangle &= 0, & \langle \bar{0} | X^{\otimes N} | \bar{0} \rangle &= 0, & \langle \bar{1} | X^{\otimes N} | \bar{1} \rangle &= 0, \\ \langle \bar{0} | X^{\otimes N} | \bar{1} \rangle &= 1, & \langle \bar{1} | X^{\otimes N} | \bar{0} \rangle &= 1, & \langle \bar{0} | Y^{\otimes N} | \bar{0} \rangle &= 0, \\ \langle \bar{1} | Y^{\otimes N} | \bar{1} \rangle &= 0, & \langle \bar{0} | Y^{\otimes N} | \bar{1} \rangle &= (-i)^N, & \langle \bar{1} | \bar{Y} | \bar{0} \rangle &= (i)^N. \end{aligned}$$

For $N = 4k$, Eq. (D2) gives $\langle \psi | W_{XYZ}^{(N,N,\mathcal{G})} | \psi \rangle = +3$. It is easy to generate other N -qubit GHZ states that also give an expectation value of $+3$ by appropriately flipping an even number of qubits with the X operator. Such a transformation leaves the expectation value unchanged, since $XYX = -Y$ and $XZX = -Z$. A further odd number of transformations will cancel the negative sign. There are a total of 2^{N-2} such GHZ-like states that produce this expectation value, which can be obtained from the following counting formula

$$\sum_{i=1}^{k-1} \binom{4k}{2i} + \frac{1}{2} \binom{4k}{2k} + 1 = 2^{N-2}. \quad (\text{D3})$$

The first term counts the number of allowed flips $2, 4, \dots, N/2 - 1$ on all possible combinations of qubits, noting that by the \mathbb{Z}_2 symmetry of the GHZ state, no more than half of the qubits need be flipped. The next term is the number of ways that we can flip exactly half of the qubits, with the factor $\frac{1}{2}$ eliminating \mathbb{Z}_2 symmetric duplicates, and the final addition of 1 just counts the original $|\psi\rangle$ state.

Similarly, one can construct GHZ-type states for systems composed of $N = 4k + 2$ qubits to give the expectation value -3 . Starting with the state

$$|\phi\rangle = \frac{|00\dots 01\rangle - |11\dots 10\rangle}{\sqrt{2}}, \quad (\text{D4})$$

and by applying similar reasoning as for the case above, one can show that $\langle \phi | W_{XYZ}^{(N,N,\mathcal{G})} | \phi \rangle = -3$. The number of states obtainable from $|\phi\rangle$ by an even number of flips, and that therefore produce the same expectation value -3 is again 2^{N-2} as above.

Now, we show that the states identified above are also eigenstates of $W_{XYZ}^{(N,N,\mathcal{G})}$. Knowing that

$$X |0\rangle = |1\rangle \quad X |1\rangle = |0\rangle, \quad (\text{D5})$$

$$Y |0\rangle = i |1\rangle \quad Y |1\rangle = (-i) |0\rangle, \quad (\text{D6})$$

$$Z |0\rangle = |0\rangle \quad Z |1\rangle = -|1\rangle, \quad (\text{D7})$$

we see that

$$X^{\otimes N} (|\bar{0}\rangle + |\bar{1}\rangle) = (|\bar{1}\rangle + |\bar{0}\rangle), \quad (\text{D8})$$

$$Y^{\otimes N} (|\bar{0}\rangle + |\bar{1}\rangle) = ((i)^N |\bar{1}\rangle + (-i)^N |\bar{0}\rangle), \quad (\text{D9})$$

$$Z^{\otimes N} (|\bar{0}\rangle + |\bar{1}\rangle) = (|\bar{0}\rangle + (-1)^N |\bar{1}\rangle), \quad (\text{D10})$$

so that $|\psi\rangle$ is an eigenstate of $W_{XYZ}^{(N,N,\mathcal{G})}$ for $N = 4k$, and the sum of the above terms gives its corresponding eigenvalue of $+3$. Similarly, other GHZ states that are generated by flipping an even number of qubits in $|\psi\rangle$ are also eigenvectors with the same $+3$ eigenvalue. One can incorporate the bit flip (X) operator to all $X^{\otimes N}$, $Y^{\otimes N}$ or $Z^{\otimes N}$ terms and use Pauli algebra, e.g. for flipping the first two qubits we will have terms like: $(XX) \otimes (XX) = \mathbb{I} \otimes \mathbb{I}$, $(YX) \otimes (YX) = -Z \otimes Z$ and $ZX \otimes ZX = -Y \otimes Y$. Thus, for example, the state

$$|\psi_{1,2}\rangle = |1100\dots 0\rangle + |0011\dots 1\rangle = (X_1 X_2)(|\bar{0}\rangle + |\bar{1}\rangle) \quad (\text{D11})$$

is an eigenstate of $W_{XYZ}^{(N,N,\mathcal{G})}$, since $X^{\otimes N}|\psi_{1,2}\rangle = Y^{\otimes N}|\psi_{1,2}\rangle = |\psi_{1,2}\rangle$, and perhaps less obviously

$$\begin{aligned} Y^{\otimes N}|\psi_{1,2}\rangle &= -Z_1 Z_2 Y^{\otimes(N-2)}(|\bar{0}\rangle + |\bar{1}\rangle) \\ &= -(i)^{N-2}|0011\dots 1\rangle - (-1)^{N-2}|1100\dots 0\rangle \\ &= -(i)^{N-2}X_1 X_2|1\rangle - (-i)^{N-2}X_1 X_2|0\rangle = |\psi_{1,2}\rangle, \end{aligned} \quad (\text{D12})$$

and similarly for all other states obtained by flipping an even number of qubits in $|\psi\rangle$. For eigenstates with the -3 eigenvalue, the reasoning follows analogous steps.

Based on Theorem 1, the upper and lower bounds of $W_{XYZ}^{(N,N,\mathcal{G})}$ are $+1$ and -1 respectively, since we have only a single generalized edge. Therefore, this observable serves as a provable entanglement witness [37] for an even number of qubits. For odd numbers of qubits, we numerically found the expectation value to be $\pm\sqrt{3}$. Similarly, if we restrict to only two terms (e.g. XZ , XY and YZ), one can follow essentially the same logic and demonstrate that an even number of qubits give an expectation value ± 2 in GHZ states, while an odd number of qubits are numerically found to give an expectation value of $\pm\sqrt{2}$. Once again, these expectation values violate the separable thresholds of $\pm|E| = \pm 1$ from Theorem 1.

-
- [1] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, Defining and detecting quantum speedup, *Science* **345**, 420 (2014).
- [2] A. J. C. Woitzik, P. K. Barkoutsos, F. Wudarski, A. Buchleitner, and I. Tavernelli, Entanglement production and convergence properties of the variational quantum eigensolver, *Phys. Rev. A* **102**, 042402 (2020).
- [3] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, Exploring entanglement and optimization within the hamiltonian variational ansatz, *PRX Quantum* **1**, 020319 (2020).
- [4] P. Díez-Valle, D. Porrás, and J. J. García-Ripoll, Quantum variational optimization: The role of entanglement and problem hardness, *Phys. Rev. A* **104**, 062426 (2021).
- [5] J. Job and D. Lidar, Test-driving 1000 qubits, *Quantum Science and Technology* **3**, 030501 (2018).
- [6] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Miller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, Quantum optimization using variational algorithms on near-term quantum devices, *Quantum Science and Technology* **3**, 030503 (2018).
- [7] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, *Phys. Rev. A* **100**, 032328 (2019).
- [8] K. Wang, Z. Song, X. Zhao, Z. Wang, and X. Wang, Detecting and quantifying entanglement on near-term quantum devices (2020), arXiv:2012.14311 [quant-ph].
- [9] A. Cornelissen, J. Bausch, and A. Gilyn, Scalable benchmarks for gate-based quantum computers (2021), arXiv:2104.10698 [quant-ph].
- [10] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm (2014), arXiv:1411.4028 [quant-ph].
- [11] S. A. Caldwell, N. Didier, C. A. Ryan, E. A. Sete, A. Hudson, P. Karalekas, R. Manenti, M. P. da Silva, R. Sinclair, E. Acala, N. Alidoust, J. Angeles, A. Bestwick, M. Block, B. Bloom, A. Bradley, C. Bui, L. Capelluto, R. Chilcott, J. Cordova, G. Crossman, M. Curtis, S. Deshpande, T. E. Bouayadi, D. Girshovich, S. Hong, K. Kuang, M. Lenihan, T. Manning, A. Marchenkov, J. Marshall, R. Maydra, Y. Mohan, W. O'Brien, C. Osborn, J. Otterbach, A. Papageorge, J.-P. Paquette, M. Pelstring, A. Polloreno, G. Prawiroatmodjo, V. Rawat, M. Reagor, R. Renzas, N. Rubin, D. Russell, M. Rust, D. Scarabelli, M. Scheer, M. Selvanayagam, R. Smith, A. Staley, M. Suska, N. Tezak, D. C. Thompson, T.-W. To, M. Vahidpour, N. Vodrahalli, T. Whyland, K. Yadav, W. Zeng, and C. Rigetti, Parametrically activated entangling gates using transmon qubits, *Phys. Rev. Applied* **10**, 034050 (2018).
- [12] M. Reagor, C. B. Osborn, N. Tezak, A. Staley, G. Prawiroatmodjo, M. Scheer, N. Alidoust, E. A. Sete, N. Didier, M. P. da Silva, E. Acala, J. Angeles, A. Bestwick, M. Block, B. Bloom, A. Bradley, C. Bui, S. Caldwell, L. Capelluto, R. Chilcott, J. Cordova, G. Crossman, M. Curtis, S. Deshpande, T. El Bouayadi, D. Girshovich, S. Hong, A. Hudson, P. Karalekas, K. Kuang, M. Lenihan, R. Manenti, T. Manning, J. Marshall, Y. Mohan, W. O'Brien, J. Otterbach, A. Papageorge, J.-P. Paquette, M. Pelstring, A. Polloreno, V. Rawat, C. A. Ryan, R. Renzas, N. Rubin, D. Russel, M. Rust, D. Scarabelli, M. Selvanayagam, R. Sinclair, R. Smith, M. Suska, T.-W. To, M. Vahidpour, N. Vodrahalli, T. Whyland, K. Yadav, W. Zeng, and C. T. Rigetti, Demonstration of universal parametric entangling gates on a multi-qubit lattice, *Science Advances* **4**, 10.1126/sciadv.aao3603 (2018), <https://advances.sciencemag.org/content/4/2/eaao3603.full.pdf>.

- [13] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, Obstacles to variational quantum optimization from symmetry protection, *Phys. Rev. Lett.* **125**, 260505 (2020).
- [14] R. Shaydulin, S. Hadfield, T. Hogg, and I. Safro, Classical symmetries and the quantum approximate optimization algorithm, *Quantum Inf. Process.* **20**, 359 (2021).
- [15] O. Gühne and G. Tth, Entanglement detection, *Physics Reports* **474**, 1 (2009).
- [16] L. Chen and Y.-X. Chen, Multiqubit entanglement witness, *Phys. Rev. A* **76**, 022330 (2007).
- [17] O. Gühne, C.-Y. Lu, W.-B. Gao, and J.-W. Pan, Toolbox for entanglement detection and fidelity estimation, *Phys. Rev. A* **76**, 030305(R) (2007).
- [18] D. P. DiVincenzo and D. Loss, Quantum computers and quantum coherence, *Journal of Magnetism and Magnetic Materials* **200**, 202218 (1999).
- [19] C. Cohen-Tannoudji, *Théorie quantique du cycle de pompage optique. Vérification expérimentale des nouveaux effets prévus*, Ph.D. thesis, Université Paris (1962).
- [20] G. M. DAriano, M. D. Laurentis, M. G. A. Paris, A. Porzio, and S. Solimeno, Quantum tomography as a tool for the characterization of optical devices, *Journal of Optics B: Quantum and Semiclassical Optics* **4**, S127S132 (2002).
- [21] R. Blume-Kohout, Optimal, reliable estimation of quantum states, *New Journal of Physics* **12**, 043034 (2010).
- [22] M. G. Paris and M. F. Sacchi, Quantum tomography, *Advances in Imaging and Electron Physics* **128**, 206 (2003).
- [23] F. Vasconcelos, M. Kjaergaard, T. Menke, S. Gustavsson, T. P. Orlando, and W. D. Oliver, Extending quantum state tomography for superconducting quantum processors.
- [24] With the assumption that the reduced density matrices can be reliably estimated via single-qubit MLE.
- [25] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
- [26] A. Bera, T. Das, D. Sadhukhan, S. Singha Roy, A. Sen(De), and U. Sen, Quantum discord and its allies: a review of recent progress, *Reports on Progress in Physics* **81**, 024001 (2017).
- [27] R. Jozsa, Fidelity for mixed quantum states, *Journal of Modern Optics* **41**, 2315 (1994), <https://doi.org/10.1080/09500349414552171>.
- [28] While the coherence for a single qubit does not depend on our choice of β , this particular choice was used to maximize the ideal expectation value of the entanglement witness we describe for the linear chain QAOA-MaxCut circuit, see Sections. IIIB and IIIG.
- [29] R. Jozsa and N. Linden, On the role of entanglement in quantum-computational speed-up, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **459**, 20112032 (2003).
- [30] A. Datta, A. Shaji, and C. M. Caves, Quantum discord and the power of one qubit, *Physical review letters* **100**, 050502 (2008).
- [31] S. Hill and W. K. Wootters, Entanglement of a pair of quantum bits, *Phys. Rev. Lett.* **78**, 5022 (1997).
- [32] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, *Physical Review Letters* **80**, 22452248 (1998).
- [33] A. R. R. Carvalho, F. Mintert, and A. Buchleitner, Decoherence and multipartite entanglement, *Phys. Rev. Lett.* **93**, 230501 (2004).
- [34] M. B. Plenio and S. S. Virmani, An introduction to entanglement theory, *Quantum information and coherence* , 173 (2014).
- [35] M. Horodecki, Entanglement measures., *Quantum Inf. Comput.* **1**, 3 (2001).
- [36] D. Chruściński and G. Sarbicki, Entanglement witnesses: construction, analysis and classification, *Journal of Physics A: Mathematical and Theoretical* **47**, 483001 (2014).
- [37] J. Bae, D. Chruściński, and B. C. Hiesmayr, Mirrored entanglement witnesses, *npj Quantum Information* **6**, 10.1038/s41534-020-0242-z (2020).
- [38] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, *Nature communications* **5**, 1 (2014).
- [39] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New Journal of Physics* **18**, 023023 (2016).
- [40] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195 (2017).
- [41] E. Farhi and H. Neven, Classification with quantum neural networks on near term processors (2018), arXiv:1802.06002 [quant-ph].
- [42] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).
- [43] If it fails, this method cannot say anything about whether the state is entangled or not.
- [44] J. S. Bell, *Speakable and Unsayable in Quantum Mechanics: Collected Papers on Quantum Philosophy* (Cambridge University Press, 2004).
- [45] A. Aspect, J. Dalibard, and G. Roger, Experimental test of bell's inequalities using time-varying analyzers, *Phys. Rev. Lett.* **49**, 1804 (1982).
- [46] A. Aspect, P. Grangier, and G. Roger, Experimental realization of einstein-podolsky-rosen-bohm gedankenexperiment: A new violation of bell's inequalities, *Phys. Rev. Lett.* **49**, 91 (1982).
- [47] Note that both the upper and lower bounds can be easily transformed to form standard entanglement witness bounds through the addition/subtraction of a scaled identity operator.
- [48] It is understood that in the local terms appearing in Eqs. (10) and (11), the identity operator acts on any of the $(N - k)$ qubits that do not appear in the tensor product. Thus, for example, the 2-local 5-qubit observable $X_1 \otimes X_4$ would really denote $X_1 \otimes I_2 \otimes I_3 \otimes X_4 \otimes I_5$.
- [49] H.-H. Wu and S. Wu, Various proofs of the cauchy-schwarz inequality, *Octogon Mathematical Magazine* **17**, 221 (2009).
- [50] If this is true, then we would have the sharper upper bound $\lambda_{max} \leq |E| + \lfloor \frac{N}{2} \rfloor$ for any observable of the form $W_{PQ}^{(2,N,G)}$, since the observables $W_{XY}^{(2,N,G)}$ and $W_{YZ}^{(2,N,G)}$ have the same eigenspectrum as the corresponding observable $W_{XZ}^{(2,N,G)}$ once the graph is fixed.
- [51] We could similarly extend this argument to $W_M^{(2,N,G)}$ (see Appendix B), and also use the same reasoning to $W_M^{(k,N,G)}$ to establish the upper bound of $M \left(\sum_{j=1}^k |E_k^{(j)}| \right) + |E_k^{(k+1)}|$ for k -separable states, where

- the edge sets $E_k^{(1)}, \dots, E_k^{(k)}$ are restricted to each of k -partitions, and $E_k^{(k+1)}$ consists of k -tuples with one index in each of the k partitions.
- [52] I. Bengtsson and K. Życzkowski, *Geometry of quantum states: an introduction to quantum entanglement* (Cambridge university press, 2017).
- [53] A. Hurwitz, ber die erzeugung der invarianten durch integration, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1897**, 71 (1897).
- [54] K. Życzkowski and H.-J. Sommers, Induced measures in the space of mixed quantum states, Journal of Physics A: Mathematical and General **34**, 71117125 (2001).
- [55] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, Volume of the set of separable states, Phys. Rev. A **58**, 883 (1998).
- [56] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).
- [57] S. J. Beale, A. Carignan-Dugas, D. Dahlen, J. Emerson, I. Hincks, P. Iyer, A. Jain, D. Hufnagel, E. Ospanov, J. Saunders, A. Stasiuk, J. J. Wallman, and A. Winick, True-q (2020).
- [58] C. Xue, Z.-Y. Chen, Y.-C. Wu, and G.-P. Guo, Effects of quantum noise on quantum approximate optimization algorithm, Chinese Physics Letters **38**, 030302 (2021).
- [59] J. Marshall, F. Wudarski, S. Hadfield, and T. Hogg, Characterizing local noise in QAOA circuits, IOP SciNotes **1**, 025208 (2020).
- [60] A. Gold, J. P. Paquette, A. Stockklauser, M. J. Reagor, M. S. Alam, A. Bestwick, N. Didier, A. Nersisyan, F. Oruc, A. Razavi, B. Scharmann, E. A. Sete, B. Sur, D. Venturelli, C. J. Winkleblack, F. Wudarski, M. Harburn, and C. Rigetti, Entanglement across separate silicon dies in a modular superconducting qubit device, npj Quantum Information **7**, 142 (2021).
- [61] P. V. Miegheem, A new type of lower bound for the largest eigenvalue of a symmetric matrix, Elsevier: Linear Algebra and its Applications 427 (2007) 119-129.
- [62] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, Quantum approximate optimization algorithm for maxcut: A fermionic view, Phys. Rev. A **97**, 022304 (2018).
- [63] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, Phys. Rev. Lett. **80**, 2245 (1998).