

Inferring Quantum Network Topology Using Local Measurements

Daniel T. Chen^{1,*}, Brian Doolittle², Jeffrey Larson¹, Zain H. Saleem¹ and Eric Chitambar³

¹Mathematics and Computer Science Division, Argonne National Laboratory, USA

²Department of Physics, University of Illinois at Urbana-Champaign, USA

³Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, USA



(Received 22 July 2023; revised 25 October 2023; accepted 28 November 2023; published 22 December 2023)

Statistical correlations that can be generated across the nodes in a quantum network depend crucially on its topology. However, this topological information might not be known *a priori*, or it may need to be verified. In this paper, we propose an efficient protocol for distinguishing and inferring the topology of a quantum network. We leverage entropic quantities—namely, the von Neumann entropy and the measured mutual information—as well as measurement covariance to uniquely characterize the topology. We show that the entropic quantities are sufficient to distinguish two networks that prepare GHZ states. Moreover, if qubit measurements are available, both entropic quantities and covariance can be used to infer the network topology without state-preparation assumptions. We show that the protocol can be entirely robust to noise and can be implemented via quantum variational optimization. Numerical experiments on both classical simulators and quantum hardware show that covariance is generally more reliable for accurately and efficiently inferring the topology, whereas entropy-based methods are often better at identifying the absence of entanglement in the low-shot regime.

DOI: [10.1103/PRXQuantum.4.040347](https://doi.org/10.1103/PRXQuantum.4.040347)

I. INTRODUCTION

Quantum entanglement [1] is a static resource that can be shared between parties and used to generate correlations. In nature, entanglement leads to interesting physical phenomena in quantum many-body systems [2], while from an engineering perspective, quantum entanglement offers operational advantages in communication, cryptography, computation, and sensing technologies [3].

We focus on quantum communication networks in which entanglement can be generated and distributed to different parties [4–6]. The entanglement can then be used to assist in a variety of networking applications [6]. The first quantum networks have been developed and will continue to scale [7–10]; therefore, to harness the power of quantum networks, it is important to characterize different types of entanglement structures.

The functionality of a quantum network will depend crucially on its *topology*, which is the particular connectivity structure between the sources and the measurement devices. While significant effort has been dedicated to

detecting multipartite entanglement [11–17], less is known about inferring a network’s topology. In practice, we need an efficient procedure for inferring network topology from the experimental data. Furthermore, such a procedure must be compatible with existing quantum systems, meaning it does not rely upon quantum memory or complex multi-qubit measurements.

Recently, several techniques have been developed to assist in this task. In most approaches, classical data is sampled from a network and tested for compatibility with a given network topology. Numerous network compatibility tests have been developed, including violations of entropic bounds [18–20], network Bell inequalities [21–24], and quantum Finner inequalities [25,26], as well as semidefinite tests on covariance matrices [27–29] and inflation techniques [30,31].

In a separate approach, the von Neumann entropy of independent measurement devices was used to infer whether two networks of GHZ states are equivalent under local unitary transformations [32]. Since each qubit in a GHZ state has a maximally mixed reduced density matrix, an analysis of the local von Neumann entropy alone will fail to distinguish between two networks whose measurement nodes receive the same number of qubits [see Fig. 3(b)]. As a solution, Yang *et al.* [32] proposed using the multipartite Shannon mutual information in addition to the von Neumann entropy, however, this solution might be challenging to scale to networks with $n \gg 0$ nodes since

*daniel_t_chen@brown.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

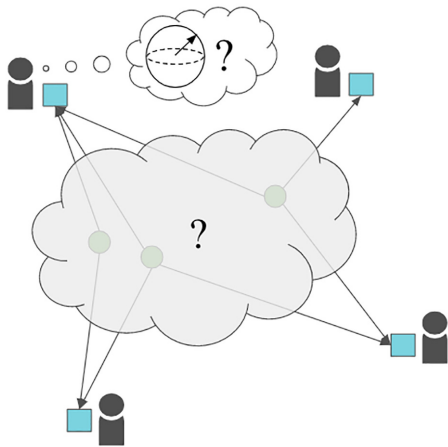


FIG. 1. We consider the problem of inferring the topology (covered by the gray cloud) of a quantum network using measurement data on qubits received from the sources. In addition to not knowing the exact connectivity or whether all links are functioning properly, the senders and receivers may have misaligned Bloch spheres used for qubit encoding and decoding. Lastly, we also consider noise corruption when sending qubits via connecting links.

it requires computing the joint entropy of all 2^n subsets of parties.

The main problem we consider is depicted in Fig. 1. The network is assumed to have a two-layered structure, the first layer consisting of quantum sources, and the second layer consisting of measurement nodes. However, the particular connectivity between these two layers is unknown and our goal is to identify this structure by studying the measurement statistics at the nodes. Furthermore, we do not assume that all nodes on the network share the same reference frame for the encoded information, which means that arbitrary local unitary rotations are permitted at either the source or measurement layers. As a final complication, we also consider the scenario of noisy connecting links.

We demonstrate two approaches to solving the above problem. First, we show that networks of different topologies can be distinguished using only local measurements, albeit requiring trustworthy sources to prepare GHZ states. Building upon Yang *et al.* [32], we use the von Neumann entropy to count the number of sources linked to each node, while we use the measured mutual information—the maximum mutual information observed between two parties using measurements local to each—between two nodes to count the number of sources they share. We extend our analysis to the case where qubitwise uniform depolarizing noise is applied to the sources. We show that the protocol can be useful for comparing network topologies when the noise strength is known, but limited in utility otherwise. On the other hand, when each node measures one qubit, we show that a network’s topology can be fully characterized also with measurements local to each node, and

in a manner robust to noise (we place no assumptions on the noise model). Furthermore, our approach is practical because it scales quadratically with the number of qubits, uses only qubit measurements and does not require a quantum memory. We also demonstrate empirically a topology inference algorithm that can be implemented on quantum hardware using variational quantum optimization methods [33]. The variational optimization approach improves inference capabilities by maximizing the observed correlations while allowing numerical estimations of the von Neumann entropy and measured mutual information, which are entropic quantities that are otherwise expensive to compute. We conduct numerical experiments on simulators and quantum hardware to contrast the choice of correlation measures, namely, mutual information and covariance. In general, covariance-based protocols are better at detecting entanglement structures while optimizing more efficiently. However, we also find that entropic methods tend to more reliably identify the absence of entanglement, particularly in the low-shot regime.

The organization of the paper is as follows. In Sec. II, we formally introduce quantum networks and their relevant entropic quantities, while establishing the notation that we use throughout the paper. Section III introduces protocols for distinguishing the topology of two quantum networks. With state-preparation assumptions, we show that two networks can be distinguished by observing entropic quantities admitted by individual or pairs of measurement nodes. We also discuss instances where the protocol breaks down under noisy channels. Then, Sec. IV briefly discusses variational optimization algorithms for estimating entropic quantities. Lastly, Sec. V extends our previous protocol to measurements on individual qubits, which gives a polynomial-time algorithm for inferring the topology of quantum networks. This algorithm does not depend on *a priori* knowledge of the prepared states and is robust to noise. We implement and test the algorithm on both simulator and quantum hardware to study the practical performance with respect to channel and statistical shot noise.

II. PRELIMINARY ON QUANTUM NETWORKS

This section introduces formally n -local quantum networks. These networks contain several components (sources, nodes, and links) as well as important measurable quantities that can be used to infer the topology of the network.

A. n -local quantum networks

An n -local quantum network, the object of interest in this paper, is an N_q -qubit system, where N_q denotes the number of qubits. Each qubit is indexed by an integer $k \in \{1, 2, \dots, N_q\} \equiv [N_q]$. A quantum network is characterized by sources Λ_i , measurement nodes A_j , and links

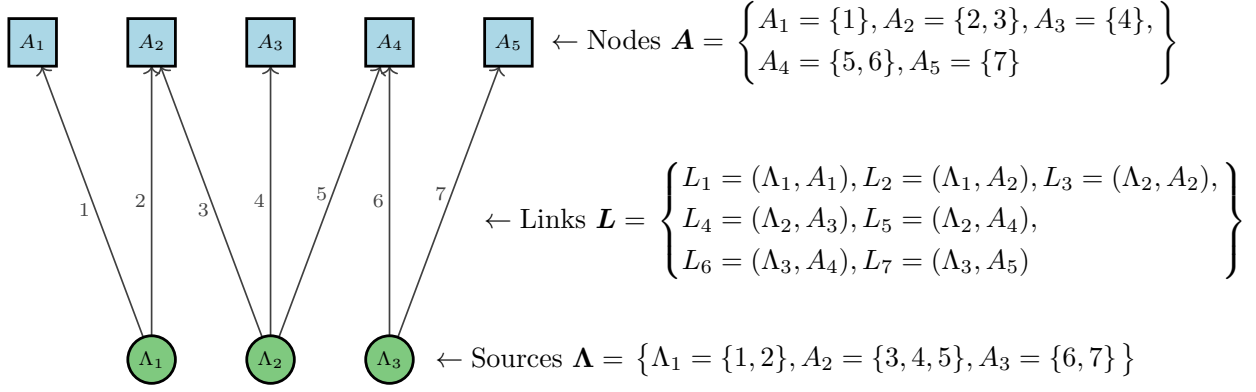


FIG. 2. A quantum network is composed of sources (green circles), links (edges), and measurement nodes (blue squares). Each link sends one qubit from a source to a node. Viewing the nodes and sources jointly as the vertex set, a quantum network can be interpreted as a bipartite graph.

L_k . Furthermore, we let N denote the number of a quantity, for example, N_s denotes the number of sources and N_m the number of measurement nodes.

More specifically, we can concisely interpret quantum networks as a directed bipartite graph $G = (\{\Lambda, \mathbf{A}\}, \mathbf{L})$. The vertices are partitioned into the sources $\mathbf{\Lambda} = \{\Lambda_i\}_{i=1}^{N_s}$ and measurement nodes $\mathbf{A} = \{A_j\}_{j=1}^{N_m}$. The edges connect sources to nodes, $\mathbf{L} = \{(\Lambda_i, A_j)\}$, and represent the movement of qubits. See Fig. 2 for an example quantum network and an enumeration of its respective parts.

1. Sources

A source, indexed by an integer $i \in [N_s]$, is characterized by the subset of qubits it acts on, $\Lambda_i \subseteq [N_q]$. In a quantum network, N_s sources collectively prepare a state $|\psi\rangle = \bigotimes_{i=1}^{N_s} |\psi^{\Lambda_i}\rangle$ where \mathcal{H} is a 2^{N_q} -dimensional Hilbert space and $|\psi\rangle \in \mathcal{H}$. There are two frequently used states in the remainder of the paper (particularly in Sec. III). The first is the Greenberger-Horne-Zeilinger (GHZ) state, which takes the form

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

where n denotes the number of qubits. When $n = 2$, GHZ states are equivalent to Bell states, denoted by $|\Phi\rangle$. Moreover, partial traces of the GHZ state results in a shared classical random bit, denoted by σ_n , defined as

$$\sigma_n = (|00\dots 0\rangle\langle 00\dots 0| + |11\dots 1\rangle\langle 11\dots 1|)/2. \quad (2)$$

This shared random bit is the other frequently used state. The density matrices of these two states differ by their off-diagonal entries.

2. Links

Links L_k , for $k \in [N_q]$, can be represented graph theoretically as an edge that connects a source node to a measurement node, $L_k = (\Lambda_i, A_j)$. Each link transmits exactly one qubit; hence, there are as many links as qubits. Physically, they are modeled as a quantum, possibly noisy, channel. A quantum channel is mathematically defined as a completely positive trace-preserving (CPTP) map [34] $\mathcal{E} : D(\mathcal{H}) \rightarrow D(\mathcal{H})$, where $D(\mathcal{H})$ denotes the space of density matrices of states in \mathcal{H} . Alternatively, the channel can be expressed in the operator-sum representation [34,35]

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger, \quad \text{where } \sum_i K_i^\dagger K_i = \mathbb{I}, \quad (3)$$

where $\{K_i\}$ are Kraus operators [35]. For example, the depolarizing channel for a one-qubit system ρ with strength γ has the following Kraus operators:

$$\begin{aligned} K_0 &= \sqrt{1 - \frac{3\gamma}{4}} \mathbb{I}_2, \quad K_1 = \sqrt{\frac{\gamma}{4}} \sigma_x, \\ K_2 &= \sqrt{\frac{\gamma}{4}} \sigma_y, \quad K_3 = \sqrt{\frac{\gamma}{4}} \sigma_z, \end{aligned} \quad (4)$$

where \mathbb{I}_2 is a 2×2 identity matrix and $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices. Note, here we are assuming that the channel noise in each link acts independently of each other.

3. Measurement nodes

Measurement nodes receive the incoming qubits and output the corresponding measurement outcomes. For a node $A_j \subseteq [N_q]$, $j \in [N_m]$, we consider a projection-valued measure (PVM) $\{\Pi_{a_j}^{A_j}\}$ that forms a set of orthogonal projectors satisfying $\sum_{a_j} \Pi_{a_j}^{A_j} = \mathbb{I}_{2^{|A_j|}}$. The node measures its local qubits $\rho^{A_j} \in D(\mathcal{H}^{A_j})$ that were received from its

linked sources. We assume measurement nodes are independent of one another, and the network applies the projector $\Pi_{\vec{a}} = \bigotimes_{j=1}^{N_m} \Pi_{a_j}^{A_j}$. Upon measurement, the classical output \vec{a} is obtained with probability,

$$\mathbb{P}(\vec{a}) = \text{tr}(\Pi_{\vec{a}} \mathcal{E}_{\text{tot}}(\rho_{\text{tot}})), \quad (5)$$

where ρ_{tot} designates the total state generated by all the sources and \mathcal{E}_{tot} is the joint channel across all edges. It is worth noting that any permutations needed to map the joint Hilbert space of sources to that of the measurement nodes are included implicitly.

B. Entropic quantities on quantum networks

The paper focuses on two entropic quantities observed on networks: the von Neumann entropy and the measured mutual information. Both quantities convey important information about the topology of the network and will be discussed in more detail in the next section.

1. von Neumann entropy

The von Neumann entropy for a quantum state ρ is defined as

$$S(\rho) = -\text{tr}(\rho \log \rho), \quad (6)$$

where the $\log(\cdot)$ above refers to the matrix logarithm and we use the convention that $\log 0 = 0$. So, any pure state $\rho = |\psi\rangle\langle\psi|$ has a von Neumann entropy of zero. Recall that the Shannon entropy of a probability distribution μ on support \mathcal{X} is defined as

$$H(\mu) = -\sum_{x \in \mathcal{X}} \mu(x) \log \mu(x). \quad (7)$$

When measured in the eigenbasis of ρ , the von Neumann entropy coincides with the Shannon entropy of the distribution over measurement outcomes [34], with all randomness coming from the superposition of pure states in ρ [34]. When measured in any other basis, the Shannon entropy calculated from measurement results is strictly greater than the von Neumann entropy because measurements only add noise. Thus, the von Neumann entropy can be calculated by minimizing the Shannon entropy over the measurement basis, i.e.,

$$S(\rho) = \min_{\{\Pi_{\vec{a}}\}} H(\mathbb{P}(\vec{a})), \quad (8)$$

where $\{\Pi_{\vec{a}}\}$ is a complete set of projections and $\mathbb{P}(\vec{a})$ is the probability distribution upon measuring the quantum state in basis $\{\Pi_{\vec{a}}\}$.

2. Measured mutual information

Intuitively, the mutual information between two random variables quantifies the amount of correlation between them. However, the conventional mutual information defined for quantum systems involves joint measurement between the two parties. Let A_i and A_j be two measurement devices. We introduce the *measured mutual information* as the maximal mutual information between local measurement distributions generated by A_i and A_j ,

$$\mathcal{I}_m(A_i, A_j) = \max_{\{\Pi_{\vec{a}_i}^{A_i} \otimes \Pi_{\vec{a}_j}^{A_j}\}} [H(\mathbb{P}(\vec{a}_i)) + H(\mathbb{P}(\vec{a}_j)) - H(\mathbb{P}(\vec{a}_i, \vec{a}_j))]. \quad (9)$$

If the two measurement nodes are not correlated, then we can decompose the joint distribution into products. Furthermore, since the Shannon entropy of independent random variables is additive, the measured mutual information will go to zero if no correlation—quantum or classical—is shared.

III. DISTINGUISHING NETWORK TOPOLOGY

We first are interested in protocols that can distinguish two network topologies. We define two networks to be the same if they are related by a graph isomorphism, formally defined below.

Definition 1 (Network topology).—Two quantum networks, $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$, have the same topology if there exists bijections $\phi : [N_s] \rightarrow [N_s], \varphi : [N_m] \rightarrow [N_m]$ such that for any edge $L_k^{(1)} = (\Lambda_i^{(1)}, A_j^{(1)})$, there is a corresponding $L_k^{(2)} = (\Lambda_i^{(2)}, A_j^{(2)}) = (\Lambda_{\phi(i)}^{(1)}, A_{\varphi(j)}^{(1)})$.

Note that in the above definition, we gave two bijections ϕ and φ separately for sources and measurement nodes. Conventionally, one bijective map is sufficient to describe the relabeling of vertices. In the context of quantum networks, the two maps are necessary to ensure sources and measurement nodes remain distinct.

Yang *et al.* gave a protocol for distinguishing the topology of quantum networks using GHZ states [32]. More specifically, they proved that the von Neumann entropies measured at each node are the same between two networks, up to a permutation of node indices, if and only if the topologies of the two networks are the same. However, this theorem only holds if no two measurement nodes share more than one source. We find this class of network restricting. In this section, we will introduce an alternative protocol that removes this restriction and distinguishes the topology of two networks where any pairs of nodes can share any number of sources.

A. Topology classification using von Neumann entropy

We begin by reviewing one of the results from Yang *et al.* [32] that we will be extending. Consider an n -local

quantum network \mathcal{N} where no two nodes share more than one source. They constructed the *characteristic vector*

$$V_{\mathcal{N}} = (S(A_1) \ S(A_2) \ \dots \ S(A_{N_m})) \quad (10)$$

to store the von Neumann entropy measured on each node. Then, a quantum network can be uniquely characterized by its characteristic vector.

Lemma 1 (Theorem 6 of Ref. [32]).—Let $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$ be two quantum networks preparing GHZ states and for any two parties A_i and A_j , they share no more than one source, that is,

$$\left| \{ \Lambda_k \in \mathbf{\Lambda} : (\Lambda_k, A_i), (\Lambda_k, A_j) \in \mathbf{L}, i \neq j \} \right| \leq 1. \quad (11)$$

Then, $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$ have the same topology if and only if their characteristic vectors are equal to each other.

It would be helpful to establish a graphical interpretation of the von Neumann entropy. We want to first establish assumptions on the class of network to be considered for the remainder of the section, enumerated below.

Assumption 1.—We assume the following:

- (A) Each source prepares maximally entangled states (GHZ states) up to local unitary transformations. Without loss of generality, assume states are prepared in the form specified in Eq. (1) (see remark below).
- (B) Each source sends at most one qubit to any given measurement device.
- (C) Only measurements local to each measurement device can be performed.

Remark 1.—Item B of Assumption 1 was primarily to avoid “parallel edges.” Note that the statistics obtained from multiple maximally entangled qubits are identical to those obtained with just one qubit. Therefore, along with the assumption that each link represents the movement of a single qubit, having parallel edges will generate the same statistics at the measurement node and multiple networks can reproduce the same measurement outcome. However, if we allow multiple qubits from the same source to be transmitted in one link, we can relax Item B of Assumption 1 at no cost to the protocol’s correctness.

Remark 2.—As mentioned previously, the basis of choice at the sources can be different than the choice at measurement nodes. However, since calculating the von Neumann entropy and the measured mutual information requires optimization over basis sets at the end of each measurement node, the optimal basis will match the reference frame of the sources. Implementation of such a procedure can be achieved via differential programming (cf. Sec. IV or Ref. [33]). Thus, for the remainder of the

paper, the term “computational basis” would be used synonymously with “the source’s reference frame” without loss of generality.

Following Assumption 1B, we can interpret the von Neumann entropy as a graph-theoretic quantity.

Lemma 2.—Let a quantum network satisfy Assumption 1. Then, for any node A_i , $S(A_i) = N_s^{A_i}$, where $N_s^{A_i}$ denotes the number of sources A_i is connected to.

Proof.—Since only GHZ states are prepared and each source can send at most one qubit, the qubits received at node A_i are all maximally entangled with another qubit that is not present in A_i . Thus, the state at node A_i is maximally mixed and has a von Neumann entropy (which, in this case, is equivalent to the Shannon entropy) of A_i is the number of qubits received, $N_s^{A_i}$. ■

Thus, in light of the graph-theoretic interpretation, Lemma 1 states that knowing the number of sources connected to each node is sufficient for knowing the topology of a quantum network. However, the assumption that nodes share no more than one entanglement is crucial and restrictive, as emphasized in the following example.

1. Example: triangle networks

Consider the two networks as shown in Fig. 3. The first network satisfies the assumption Yang *et al.* [32] made. Each node receives two qubits, each from two different sources. Since the subsystem of any maximally entangled state is a maximally mixed one, the von Neumann entropy at each node is 2. On the other hand, the second network consists of only two preparation nodes, each preparing a three-qubit GHZ state. Each node in network 2 also receives two qubits, one from each source. Again by property of maximally entangled states, the von Neumann entropy at each node is 2. One could take a step further and study the von Neumann entropy of the joint state of two measurement nodes only to find out that the two networks yield the same statistics. Thus, observing the von Neumann entropy alone cannot distinguish networks.

As a solution, Yang *et al.* [32] show that the Shannon mutual information can distinguish between the two networks in Fig. 3. Although the Shannon mutual information is evaluated from classical data, this entropic quantity must

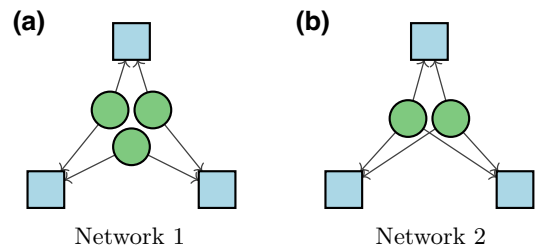


FIG. 3. Example triangle networks that are indistinguishable solely from von Neumann entropy. (a) Network 1. (b) Network 2.

be evaluated for all groupings of parties where the number of groupings scales exponentially with the number of parties. Thus, this approach is not practical for large networks.

We propose the addition of the *measured mutual information*. We claim that the basis that maximizes the Shannon entropy is the computational basis, which will be formally proven later. Take any two nodes in network 1. The Shannon entropy at each node will be 2 since a maximally mixed state is information theoretically equivalent to a fair classical coin flip. The joint state of the nodes can be written as

$$\frac{1}{4} (\mathbb{I}_2 \otimes |\Phi\rangle\langle\Phi| \otimes \mathbb{I}_2), \quad (12)$$

which acts equivalently as three independent fair coin flips. This yields a joint Shannon entropy of 3 with measurements local to each node. Thus, the measured mutual information will be 1 for all pairs of nodes in network 1.

The same does not hold for network 2! The joint state of any pair of nodes in network 2

$$\frac{1}{4} (|00\rangle\langle 00| + |11\rangle\langle 11|)^{\otimes 2} \quad (13)$$

has a Shannon entropy of 2 upon measuring separately in the respective nodes. Thus, the measured mutual information in network 2 is 2.

B. Protocol for distinguishing network topology

The example above gave evidence for a graph-theoretic interpretation of both entropic quantities—the von Neumann entropy of a node gives the number of sources the node is connected to, whereas the measured mutual information gives the number of sources the two nodes share. We formally present this in the lemma below, whose proof is deferred to Appendix A.

Lemma 3.—Let a quantum network satisfy Assumption 1. Then, for any two measurement nodes A_i and A_j , $\mathcal{I}_m(A_i; A_j) = N_s^{A_i, A_j}$, where $N_s^{A_i, A_j}$ denotes the number of sources they share.

The interpretation presented in Lemmas 2 and 3 will be useful for proving the correctness of our protocol. Furthermore, in spirit the characteristic vector in Ref. [32], we define the *characteristic matrix* of a quantum network to be

$$M_{\mathcal{N}} = \begin{pmatrix} S(A_1) & \mathcal{I}_m(A_1; A_2) & \dots & \mathcal{I}_m(A_1; A_{N_m}) \\ \mathcal{I}_m(A_2; A_1) & S(A_2) & \mathcal{I}_m(A_2; A_3) & \dots \\ \vdots & \dots & \dots & \vdots \\ \mathcal{I}_m(A_{N_m}; A_1) & \dots & \mathcal{I}_m(A_{N_m}; A_{N_m-1}) & S(A_{N_m}) \end{pmatrix} \quad (14)$$

where the diagonal is the characteristic vector $V_{\mathcal{N}}$ containing the von Neumann entropy and the off diagonals are the measured mutual information. Note that the matrix is symmetric, $M_{\mathcal{N}} = M_{\mathcal{N}}^T$. By introducing the off-diagonal terms, we can quantify the number of sources two nodes share. This addition allows us to extend the previous network classification protocol to include cases where more than one entanglement is shared between nodes.

We note that the characteristic matrix bears resemblance to the covariance matrix used in the semidefinite tests for network compatibility [27–29] where the off diagonals of the covariance matrix are nonzero if and only if a source correlates two measurements. A key distinction is that the covariance matrix is evaluated from the classical data sampled from the network, whereas the characteristic matrix is evaluated by optimizing the measurements with respect to von Neumann entropy and measured mutual information. Indeed, it is significantly more efficient to evaluate

the covariance matrix, however, having control over the measurement apparatus improves our ability to probe the strength of the correlation between measurement devices. Thus, the characteristic matrix might give a more detailed view of the network’s topology. We study the differences between entropic and covariance methods later in the paper (cf. Sec. V) for inferring the topology of quantum networks using qubit measurements.

We now show that the topology of an n -local quantum network can be fully characterized by the characteristic matrix $M_{\mathcal{N}}$, which specifies the von Neumann entropy at each node and the measured mutual information of each pair of nodes.

Theorem 1.—Let two quantum networks, $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$, satisfy Assumption 1. $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$ have the same topology (cf. Definition 1) if and only if $S(A_i^{(1)}) = S(A_i^{(2)})$ for all nodes A_i and $\mathcal{I}_m(A_i^{(1)}; A_j^{(1)}) = \mathcal{I}_m(A_i^{(2)}; A_j^{(2)})$ for all pairs of nodes A_i, A_j .

We defer the proof of the theorem to Appendix B.

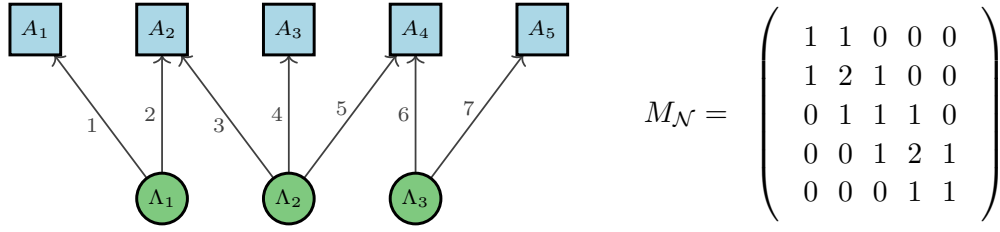


FIG. 4. Example quantum network and its respective characteristic matrix $M_{\mathcal{N}}$. Theorem 1 shows that $M_{\mathcal{N}}$ uniquely characterizes a quantum network. However, inferring the network from $M_{\mathcal{N}}$ is nontrivial.

The theorem states that one can uniquely find the topology of a network from its characteristic matrix. Suppose entropic quantities can be reliably calculated, verifying if two networks have the same topology requires only that the number of queries grows polynomially with respect to the number of nodes.

Inferring the topology from the characteristic matrix remains a difficult task. See Fig. 4 for an example network and its corresponding characteristic matrix $M_{\mathcal{N}}$. Knowing the topology, $M_{\mathcal{N}}$ can be straightforwardly obtained using Lemmas 2 and 3. However, we encourage the reader to try the other direction. Even though $M_{\mathcal{N}}$ indicates whether A_i and A_j share a source(s), finding the correct number of sources N_s and assigning nodes to the respective sources appear to be highly nontrivial. Naively, one could search through all possible quantum networks with N_s sources, for all possible N_s . The search space grows exponentially and would not be tractable for large networks. Thus, we defer the existence of a polynomial-time algorithm for decoding the characteristic matrix $M_{\mathcal{N}}$ as a future direction.

C. Distinguishing the topology of noisy networks

Maximally entangled states such as GHZ states are fragile and easily corrupted by noise. In this section, we hope to establish robustness for the classification protocol introduced subjected to depolarizing noise, that is, for a quantum state ρ , a depolarizing channel \mathcal{E}_γ performs the map

$$\mathcal{E}_\gamma(\rho) = (1 - \gamma)\rho + \frac{\gamma}{2^n}\mathbb{I}_{2^n}, \quad (15)$$

where $\gamma \in [0, 1]$ is the parameter of the channel, and n is the number of qubits involved in state ρ . In the quantum network setting, depolarizing noise acts jointly on qubits prepared by a source, and sources are affected independently.

1. Example: triangle network revisited

Consider the triangle network 1 shown in Fig. 3(a). Like in the noiseless case, the Shannon entropy at each measurement device is independent of the choice of measurement basis, and $H(A_i) = S(A_i) = 1$ for all measurement nodes

in the network. In an attempt to characterize the topology, we look at the measured mutual information between devices A_i and A_j , where the joint state is

$$\rho_{A_i \cup A_j} = \frac{\mathbb{I}_2}{2} \otimes |\Phi\rangle\langle\Phi| \otimes \frac{\mathbb{I}_2}{2}. \quad (16)$$

If each qubit is sent through a depolarizing channel of the same noise parameter, then the joint state received $\mathcal{E}_\gamma(\rho_{A_i \cup A_j})$ becomes

$$\mathcal{E}_\gamma(\rho_{A_i \cup A_j}) = \frac{\mathbb{I}_2}{2} \otimes \left((1 - \gamma) |\Phi\rangle\langle\Phi| + \frac{\gamma}{4} \mathbb{I}_4 \right) \otimes \frac{\mathbb{I}_2}{2}, \quad (17)$$

which yields the Shannon entropy of

$$-H(\mathcal{E}_\gamma(\rho_{A_i \cup A_j})) = \frac{2 - \gamma}{2} \log(2 - \gamma) + \frac{\gamma}{2} \log \gamma - 4 \quad (18)$$

when measured using the computational basis for both devices. When the channel is noiseless, i.e., $\gamma = 0$, then $\mathcal{I}_m(A_i; A_j) = 1$ and we recover the results shown in Sec. III A. However, if the channel is completely noisy, meaning that $\gamma = 1$, then $H(\mathcal{E}_\gamma(\rho_{A_i \cup A_j})) = 4$ and the measured mutual information is zero, which can lead one at the receiver end to think that qubits received at A_i and A_j are independent of one another.

The same calculation can be applied to network 2 in Fig. 3. Let $\sigma_2 = (|00\rangle\langle 00| + |11\rangle\langle 11|)/2$. Knowing that the joint system $\rho_{A_i \cup A_j} = \sigma_2 \otimes \sigma_2$, we focus on the behavior of one σ_2 knowing that the remaining system behaves identically and independently. We can find the noisy joint system of σ_2 to be

$$\mathcal{E}_\gamma(\sigma_2) = \frac{\gamma}{4} \mathbb{I}_4 + (1 - \gamma)\sigma_2. \quad (19)$$

The above state yields the following Shannon entropy when measured in the computational basis.

$$-H(\mathcal{E}_\gamma(\sigma_2)) = \frac{2 - \gamma}{2} \log(2 - \gamma) + \frac{\gamma}{2} \log \gamma - 2. \quad (20)$$

Since the joint system is separable, the entropy of the joint system simply adds.

$$-H(\mathcal{E}_\gamma(\rho_{A_i \cup A_j})) = (2 - \gamma) \log(2 - \gamma) + \gamma \log \gamma - 4. \quad (21)$$

Again, we can cover the noiseless and completely random behavior when $\gamma = 0$ or $\gamma = 1$, respectively. However, we are interested in determining if the two systems are ever indistinguishable due to noise.

Claiming the computational basis is again the basis of choice that maximizes the mutual information and from Eqs. (18) and (21), the measured mutual information is, respectively,

$$\mathcal{I}_m^{(1)}(A_i; A_j) = \frac{2 - \gamma}{2} \log(2 - \gamma) + \frac{\gamma}{2} \log \gamma, \quad (22)$$

$$\mathcal{I}_m^{(2)}(A_i; A_j) = (2 - \gamma) \log(2 - \gamma) + \gamma \log \gamma. \quad (23)$$

We can see that the measured mutual information of the first network is half of the second one. Consequently, this means that for depolarizing of strength $\gamma \in [0, 1)$, the measured mutual information will always have a nonzero gap and we will be able to distinguish the two networks through local measurements.

We extend Theorem 1 to the case of applying uniform global noise on sources. To do so, we first want to show that the same strategy of measuring in the computational basis in the noiseless case is still valid in the presence of noise (which is an extension of Lemmas 5 and 6 in the Appendix to the noisy case).

Lemma 4.—Consider a quantum network satisfying Assumption 1, measurement nodes A_i and A_j . Moreover, the network is made up of depolarizing channels that act on each qubit with strength γ . Then, the local measurement basis that maximizes the Shannon mutual information, that is

$$\operatorname{argmax}_{\{\Pi_{\vec{a}_i}^{A_i}, \Pi_{\vec{a}_j}^{A_j}\}} H(\mathbb{P}(\vec{a}_i)) + H(\mathbb{P}(\vec{a}_j)) - H(\mathbb{P}(\vec{a}_i, \vec{a}_j)) \quad (24)$$

is the computational basis.

Proof.—Let ρ be either a Bell pair $|\Phi\rangle\langle\Phi|$ or shared random bits σ_2 . Under depolarizing noise on the source, the state becomes

$$\mathcal{E}_\gamma(\rho) = (1 - \gamma)\rho + \frac{1}{4}\mathbb{I}_4. \quad (25)$$

Note that for any unitary U applied onto the noisy state, the effect of the noise stays unchanged, that is

$$U\mathcal{E}_\gamma(\rho)U^\dagger = (1 - \gamma)U\rho U^\dagger + \frac{1}{4}\mathbb{I}_4. \quad (26)$$

Thus, the measurement basis maximizing the Shannon entropy of ρ remains the same for all $\gamma > 0$. ■

We are interested in whether there exists a $\gamma \in [0, 1)$ such that the measured mutual information of two networks is the same. If such γ exists, denote γ^* , then the two networks are identical according to the protocol within some small neighborhood of γ^* . On the other hand, if the measured mutual information remains distinct for all γ and pairs across two networks, then the protocol will be applicable for any noise level that is not completely depolarizing while assuming an infinite precision. Below, we provide one condition that guarantees such robustness to noise.

Theorem 2.—Consider two quantum networks $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$ satisfying Assumption 1. For depolarizing channels with known strength $\gamma \in [0, 1)$, we can distinguish the topology of $\mathcal{N}^{(1)}$ from $\mathcal{N}^{(2)}$.

Proof.—Under the assumption that each preparation node can send at most one qubit, for any two measurement nodes A_i and A_j , the joint state $\rho_{A_i A_j}$ will be the tensor product of \mathbb{I}_2 , σ_2 , and Bell pairs. In particular, the von Neumann entropy of each device is invariant with respect to noise. Moreover, the measured mutual information of σ_2 and $|\Phi\rangle\langle\Phi|$ is the same for all $\gamma \in [0, 1]$. Let $I(\gamma) = \mathcal{I}_m(\mathcal{E}_\gamma(|\Phi\rangle\langle\Phi|)) = \mathcal{I}(\mathcal{E}_\gamma(\sigma_2))$, the measured mutual information as a function of γ will be $\mathcal{I}_m(A_i; A_j) = N_s^{A_i, A_j} I(\gamma)$.

If $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(2)}$ have different topologies, it means that there exists at least one pair of nodes A_i and A_j such that (without loss of generality) $N_s^{A_i^{(1)}, A_j^{(1)}} > N_s^{A_i^{(2)}, A_j^{(2)}}$. Thus, for any depolarizing channel of strength γ , $\mathcal{I}_m(A_i^{(1)}; A_j^{(1)}) > \mathcal{I}_m(A_i^{(2)}; A_j^{(2)})$. Thus, by Theorem 1, the two networks can be distinguished given sufficient shots taken to estimate each entropic quantity. ■

Note that Theorem 2 actually provides a limited sense of noise robustness, that is, we need *a priori* knowledge of γ . The theorem can remain useful for comparing two quantum networks with unknown topologies but are under the influence of depolarizing noise of the same strength or verifying the topology of one noisy network. Recall that there are no straightforward algorithms for deriving the topology from the characteristic matrix even in the noiseless case. Moreover, inferring topology only becomes more difficult when the two networks are exposed to noises of different strengths. Furthermore, the assumption of depolarizing noise is largely due to simplicity, and we leave formal analysis of the protocol using other noise models as future work.

IV. VARIATIONAL QUANTUM NETWORK TOPOLOGY INFERENCE

In this section we introduce our variational quantum optimization framework for inferring the topology of sources in noisy uncharacterized quantum networks. We implement our approach as Python software

called `qNetTI`: The Quantum Network Topology Inferer, which is publicly available on GitHub [36]. Our software applies the PennyLane framework for cross-platform differential programming [37], and builds upon `qNetVO`: the Quantum Network Variational Optimizer software [33,38]. Our variational optimization techniques are hardware agnostic and can easily be extended to many general quantum network characterization, validation, and verification tasks. We apply our variational quantum optimization methods in Sec. V, in which we compare entropy-based and covariance-based topology inference when local qubit measurements are applied.

In general, when inferring the topology of sources in a quantum network, the respective covariances or entropic quantities must be sufficiently large such that correlations between measurements, or lack thereof, can be witnessed. However, since the measurement nodes do not have a standard reference frame for each of their measurement bases, a naive measurement choice will not necessarily reflect the correlations between measurement devices with accuracy. Thus, by optimizing the covariances or entropic quantities used to infer network topology, the network inference protocol is made more robust to the errors that result from nonoptimal measurement choices.

To further motivate our hardware-compatible optimization scheme, we note that the entropic quantities such as von Neumann entropy or measured mutual information contain an optimization implicitly in their operational definition [see Eqs. (8) and (9), respectively]. Indeed, the von Neumann entropy at a given measurement device cannot be ascertained by performing a single measurement. That is, either the reduced density matrix measured at the device must be known, or the measurements must be optimized such that the Shannon entropy is minimized as in Eq. (8). A similar remark can also be made for optimization used to obtain the measured mutual information in Eq. (9). To obtain these quantities in practice a quantum-hardware-compatible optimization technique is needed, hence we apply variational quantum-optimization methods.

Variational quantum optimization is a type of hybrid quantum-classical algorithm in which a classical computer tunes a parameterized quantum circuit such that a cost function is minimized [39]. The quantum circuit is evaluated on quantum hardware leading to the hardware being optimized or trained for the task encoded by the cost function. Generally, variational quantum algorithms can be applied to a wide range of optimization and simulation problems and show promise of providing practical advantages [40]. While their seminal applications were mainly in quantum computing, variational quantum algorithms have recently been proposed as a technique for optimizing noisy and uncharacterized quantum network hardware for various tasks [33].

Within our hybrid optimization framework for quantum network topology inference, we construct a variational

ansatz as follows. Let the collection of sources prepare the state $|\psi\rangle = \bigotimes_{i=1}^{N_s} |\psi^{\Lambda_i}\rangle$ while each measurement node performs the PVM measurement $\{\Pi_{a_j}^{A_j}(\vec{\theta}^{A_j})\}_{a_j \in A_j}$ where

$$\Pi_{a_j}^{A_j}(\vec{\theta}^{A_j}) = U^{A_j}(\vec{\theta}^{A_j})^\dagger |a_j\rangle\langle a_j| U^{A_j}(\vec{\theta}^{A_j}) \quad (27)$$

and $U^{A_j}(\vec{\theta}^{A_j})$ is a unitary operator parameterized by $\vec{\theta}^{A_j} \in \mathbb{R}^m$. In total the measurement $\Pi_{\vec{a}}(\Theta) = \bigotimes_{j=1}^{N_m} \Pi_{a_j}^{A_j}(\vec{\theta}^{A_j})$ is applied while the probability of measuring outcome $\vec{a} = (a_j)_{j=1}^{N_m}$ is

$$\mathbb{P}(\vec{a}|\Theta) = |\langle \psi | \Pi_{\vec{a}}(\Theta) | \psi \rangle|^2. \quad (28)$$

Then, for measurement node A_j , we may rewrite the von Neumann entropy in Eq. (8) as the cost function

$$S(\rho_{A_j}) = \min_{\vec{\theta}^{A_j} \in \mathbb{R}^m} H(\mathbb{P}(a_j | \vec{\theta}^{A_j})). \quad (29)$$

Likewise, measured mutual information in Eq. (9) becomes

$$\begin{aligned} -\mathcal{I}_m(A_i; A_j) &= \min_{\vec{\theta}^{A_i} \in \mathbb{R}^{m_i}, \vec{\theta}^{A_j} \in \mathbb{R}^{m_j}} H\left(\mathbb{P}(a_i, a_j | \vec{\theta}^{A_i}, \vec{\theta}^{A_j})\right) \\ &\quad - H\left(\mathbb{P}(a_i | \vec{\theta}^{A_i})\right) - H\left(\mathbb{P}(a_j | \vec{\theta}^{A_j})\right), \end{aligned} \quad (30)$$

in which the minus sign on the measured mutual information results from the convention of minimizing the cost function in variational optimization.

To optimize the quantities in Eqs. (29) and (30), a gradient-descent algorithm is used. Consider a generic cost function $f(\Theta)$ that we aim to minimize. Then, we can incrementally approach the optimal settings $\Theta^* = \arg \min_{\Theta} f(\Theta)$ by updating the settings as

$$\Theta' = \Theta - \eta \nabla_{\Theta} f(\Theta), \quad (31)$$

where $\eta > 0$ is a small stepsize and the gradient of $f(\Theta)$, $\nabla_{\Theta} f(\Theta)$, is a vector pointing in the direction of steepest ascent. Therefore, in many small steps the algorithm navigates its way to a local minimum in the landscape of the cost function.

To apply these variational-optimization methods to infer the topology of an unknown quantum state $|\psi\rangle$, it must be run on the quantum network's hardware. The reason is that $|\psi\rangle$ is not known and therefore cannot be reconstructed or characterized. To evaluate gradients on quantum hardware we make use of the parameter-shift rule [41], which evaluates gradients on quantum hardware with a computational complexity that is linear in the number of parameters. Hence, our variational methods for quantum network topology inference can conceivably be applied with efficiency in practical quantum networking systems.

As a final remark, we note that our variational quantum network topology inference scheme provides several improvements to existing techniques. First, previous methods do not typically consider the measurements to be tunable. By optimizing measurements, our variational methods could improve the effectiveness of existing topology inference approaches in which the violation of entropic bounds or Bell inequalities are used to test for network topology [18–26,30,31]. Second, quantities such as the von Neumann entropy or measured mutual information cannot be obtained for unknown quantum states. Typically, calculating these quantities would require estimating the entire quantum state via state tomography. Alternatively, our quantum-hardware-compatible variational methods can approximate these quantities through optimization. Thus, entropy-based inference methods can be applied in practice. Finally, our methods are hardware agnostic. The parameter-shift rule requires only a parameterized description of the unitary applied by the measurement device, how the unitary is physically implemented is not relevant.

V. INFERRING NETWORK TOPOLOGY FROM QUBIT MEASUREMENTS

Previously, we considered the qubits in a measurement node to be measured jointly. While there exists a unique mapping from network topology to the characteristic matrix (cf. Theorem 1), constructing the map in a reasonable amount of time, particularly for large networks, is

$$Q_{\mathcal{N}} = \begin{pmatrix} S(q_1) & \mathcal{I}_m(q_1; q_2) & \dots & \mathcal{I}_m(q_1; q_{N_q}) \\ \mathcal{I}_m(q_2; q_1) & S(q_2) & \mathcal{I}_m(q_2; q_3) & \dots \\ \vdots & \dots & \dots & \vdots \\ \mathcal{I}_m(q_{N_q}; q_1) & \dots & \mathcal{I}_m(q_{N_q}; q_{N_q-1}) & S(q_{N_q}) \end{pmatrix}. \quad (32)$$

Similar to $M_{\mathcal{N}}$, the diagonal entries stores the von Neumann entropy of each single-qubit state and the off-diagonal entries stores the correlation quantified by the measured mutual information. By treating each qubit as its own measurement node, the qubit characteristic matrix inherits all the properties of $M_{\mathcal{N}}$, and gains additional structure that can help with decoding the entanglement structure.

Theorem 3.—Consider an n -local network \mathcal{N} measured using local qubit projectors $\Pi_a^{\mathcal{N}} = \bigotimes_{j=1}^m \Pi_{a_j}^{q_j}$ where $a_j \in \{0, 1\}$ and $q_j \in [N_q]$ index the measured qubit. Moreover, suppose Assumption 1 holds. Then, the network’s

difficult. Moreover, qubits undergoing depolarizing channels with unknown noise parameters can render the protocol useless. Thus, the classification protocol introduced in Sec. III would be of practical use only in very limited situations.

These issues can be resolved if measurements are available on the qubit level, which is often the case for real-life situations. The increased granularity allows for a protocol that can infer network topology from measurements local to each node in polynomial time. Moreover, the procedure no longer requires GHZ state preparations, is entirely robust to noise induced by quantum channels, and can maintain performance using low numbers of circuit evaluations.

The protocol is also not restricted to measured mutual information to estimate classical correlation. In particular, we show that covariance and classical mutual information are both suitable choices for identifying the entanglement structure of quantum networks. While the covariance-based methods generally exhibit more stable convergence and are more robust to noise, entropic alternatives excel under the low-shot regime.

A. Entropy-based protocol for inferring network topology

We can derive a protocol for inferring network topology by straightforwardly extending the characteristic matrix $M_{\mathcal{N}}$ [Eq. (14)] to qubits. Let $\{q_i, i \in [N_q]\}$, be the set of qubits in the network. Then, we can define the *qubit characteristic matrix*:

topology is completely characterized by the qubit characteristic matrix $Q_{\mathcal{N}}$ [Eq. (32)], where the i th row lists the qubits entangled with qubit q_i and the number of sources N_s is equivalent to the number of unique rows (or columns) of $Q_{\mathcal{N}}$.

Proof.—Consider each qubit to be its own measurement node. Then, the matrix $Q_{\mathcal{N}}$ completely characterizes the network by invoking Theorem 1.

To see the relationship between the number of unique rows (or columns) by the symmetry of $Q_{\mathcal{N}}$, fix a particular source Λ_k . Then, for any row i where $q_i \in \Lambda_k$, the (i, j) th entry is 1 if and only if $q_j \in \Lambda_k$. And since each qubit

can only come from one source, all qubits whose corresponding rows are equal to one another are from the same source. Therefore, partitioning the set of qubits into sets whose respective rows are equal recovers the qubit-source mapping, and the number of partitions is the number of sources. ■

It is important to note that we do not need Item A of Assumption 1 to hold to infer the topology. If we construct another matrix B defined componentwise by

$$B_{i,j} = \begin{cases} 1 & \text{if } Q_{N,i,j} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (33)$$

and treat each qubit as individual measurement nodes, then B is equivalent to Q_N with the same entanglement structure assuming GHZ state preparation. To infer the topology, we can simply apply Theorem 3 to the matrix B . Thus, for the remainder of the section, we use Q_N to refer to both the qubit characteristic matrix in Eq. (32) while assuming GHZ preparation and this binary matrix B without the state-preparation assumption as they are functionality equivalent.

The above theorem gives an algorithm for reconstructing the network topology given the qubitwise characteristic matrix Q_N . First, assume knowledge of the measurement node that each qubit is sent to. The columns of Q_N , each representing a qubit, can then be grouped into their respective measurement nodes. On the other hand, the rows of Q_N can be partitioned into sets of indices with identical rows, that is, let Λ_i be a set such that for all $j, k \in \Lambda_i$, $Q_{N,j,*} = Q_{N,k,*}$. As the notation suggests, this set of indices is the set of qubits from the source Λ_i . Lastly, if $Q_{N,r,s} = 1$, qubits r (in node A_j) and s (in node A_k) share a source Λ_i and the triplet (A_j, Λ_i, A_k) exists in the network. In Fig. 5, we give an elementary demonstration of the algorithm on the triangle network presented in Fig. 3(b).

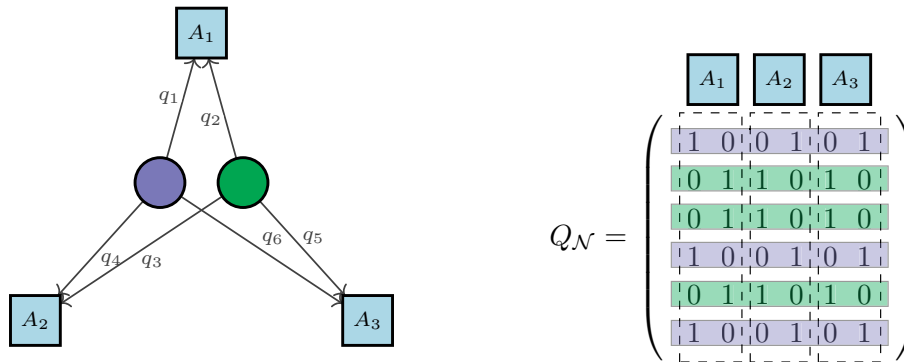


FIG. 5. Application of Theorem 3 on the network in Fig. 3(b). The columns of Q_N are organized by the nodes each qubit is from, and unique rows are grouped together. Each group of unique rows corresponds to a preparation node (purple or green) and the connectivity can be found by observing the nonzero entries in each row.

Performing joint measurements, even for qubits received in the same measurement node, can be experimentally demanding. Thus, limiting ourselves to qubitwise measurement actually enhances the practicality of the protocol. The fine-grained measurements provide a simple algorithm for determining the topology in time quadratic to the number of qubits. Moreover, sources are not restricted to preparing only GHZ states; any entangled states will do. Given infinite precision and noiseless channels, each entry of Q_N can store whether there exist correlations between qubits. As correlations can only arise from quantum entanglement, two qubits are from the same source if and only if a nonzero correlation is observed.

Since the protocol aims at identifying zero or nonzero correlations, as long as the quantum channel is not completely destructive, one can always infer the topology given enough shots to suppress statistical noise. This statement is formalized below.

Theorem 4.—Consider a noisy network \mathcal{N} that is measured using local qubit measurements and $\rho^{\text{Net}} = \bigotimes_{i=1}^n \rho^{\Lambda_i}$. Its topology is completely characterized by the matrix Q_N as described in Theorem 3.

Proof.—Let $\rho_i = \text{tr}_{j \neq i}[\rho^{\text{Net}}]$ for all $j \in [N_q]$. Then, if $S(\rho_i) > 0$ a source may exist that correlates the qubit with other qubits. Next, the measured mutual information $\mathcal{I}_m(q_i, q_j) = 0$, if and only if $\rho_i \otimes \rho_j$. This implies that when $\mathcal{I}_m(q_i, q_j) > 0$, a source must be present to correlate the two independent qubit measurements. Therefore, the matrix Q_N only has nonzero elements on its off diagonal if there exist sources to correlate the qubits. In practice, finite samples are taken and the scalar value of $\mathcal{I}_m(q_i, q_j)$ should only be counted as nonzero if it is sufficiently larger than the statistical fluctuations of uncorrelated qubits. ■

Note that in the above theorem, a source can be so noisy that it separates as $\bigotimes_j \rho_j$. We argue that the source no longer qualifies as such precisely because it no longer distributes shared randomness. Otherwise, so long as a sufficient number of measurements are taken and nonseparable

states are prepared at each source, the qubitwise characteristic matrix $Q_{\mathcal{N}}$ is sufficient for determining the network topology. On a different note, for certain choices of noise, such as colored noise and qubit dephasing noise, the characteristic matrix is preserved; thus, both Theorems 2 and 4 will hold under these noise models.

1. Alternative definition for mutual information on networks

Constructing the characteristic matrix $Q_{\mathcal{N}}$ requires the number of independent circuit evaluations to grow at a rate equal to the number of unique qubit pairs. We associate a unique measurement basis that achieves the maximum of Eq. (9) for each pair. In the case of N_q being large, we might be tempted to define a measurement basis common across all qubit pairs that maximizes an analogous quantity that takes the mutual information of all pairs into account. Formalizing this intuition, we seek measurement operators $\{\Pi_{\bar{x}} = \bigotimes_{q_i} \Pi_{x_i}^{q_i}\}$ such that

$$\max_{\{\Pi_{\bar{x}}\}} \sum_{i < j} H(\mathbb{P}(x_i)) + H(\mathbb{P}(x_j)) - H(\mathbb{P}(x_i, x_j)) \quad (34)$$

and we let the mutual information between any two qubits obtained from such basis be the *classical mutual information* (as opposed to the original *bipartite* measured mutual information).

Comparing Eq. (34) with the expression for measured mutual information in Eq. (9), we can see that

$$\begin{aligned} & \max_{\{\Pi_{\bar{x}}\}} \sum_{i < j} H(\mathbb{P}(x_i)) + H(\mathbb{P}(x_j)) - H(\mathbb{P}(x_i, x_j)) \\ & \leq \sum_{i < j} \max_{\{\Pi_{x_i}^{q_i} \otimes \Pi_{x_j}^{q_j}\}} H(\mathbb{P}(x_i)) + H(\mathbb{P}(x_j)) - H(\mathbb{P}(x_i, x_j)). \end{aligned} \quad (35)$$

Thus, for each unique pair of qubits, measured mutual information between the qubit pair is at least that of the one calculated from the classical mutual information of the whole network. We can construct an example where the inequality is, in fact, strict.

Remark 3.—Consider a tripartite network that has the following mixed state:

$$\begin{aligned} \rho = \frac{1}{8} & \left(\mathbb{I} \otimes \mathbb{I} \otimes \mathbb{I} + \frac{1}{2} \sigma_x \otimes \sigma_x \otimes \mathbb{I} \right. \\ & \left. + \frac{1}{2} \sigma_y \otimes \mathbb{I} \otimes \sigma_y + \frac{1}{2} \mathbb{I} \otimes \sigma_z \otimes \sigma_z \right). \end{aligned} \quad (36)$$

For this network, the measurement basis that maximizes the correlation—mutual information and covariance (see next section) alike—between any pair of qubits is orthogonal to that of any other pair. Since the bipartite measured

mutual information consists of a collection of measurement bases, one for each pair, correlations can be fully observed. In the case of classical mutual information, in which we seek to find one optimal basis for all qubit pairs, there will always be correlations that are not fully observed.

On the other hand, the benefit of considering the network mutual information is clear: the number of circuit evaluations needed is constant with respect to the number of qubits.

2. Covariance-based topology inference

The measured mutual information is not the only measure of the correlation between two random variables. Inspired by the literature [27–29], we propose a network topology inference protocol by applying the decoding scheme above on covariance matrices.

We treat each qubit measurement as a random variable taking values in $\{-1, 1\}$. From the definition of covariance, we get

$$\text{Cov}(q_i, q_j) = \sum_{x, y \in \{-1, +1\}} xy \mathbb{P}(q_i = x, q_j = y) - \bar{q}_i \bar{q}_j \quad (37)$$

$$= \sum_{x, y \in \{-1, +1\}} xy \text{tr} \left(\left(\Pi_x^{q_i} \otimes \Pi_y^{q_j} \right) \rho_{q_i q_j} \right) - \bar{q}_i \bar{q}_j \quad (38)$$

where $\rho_{q_i q_j}$ is the state of qubits i and j and

$$\bar{q}_i = \sum_{x \in \{-1, +1\}} x \mathbb{P}(q_i = x) \quad (39)$$

$$= \sum_{x \in \{-1, +1\}} x \text{tr} \left(\Pi_x^{q_i} \text{tr}_j(\rho_{q_i}) \right), \quad (40)$$

and similarly for \bar{q}_j . The qubit covariance matrix, denoted by $C_{\mathcal{N}}$, is defined entrywise where the (i, j) th entry contains the absolute value of the covariance between qubits i and j .

The magnitude of the covariance depends on the measurement basis. If the basis is chosen arbitrarily, there is a chance that the basis of choice yields low covariance, thereby skewing the inference. To reduce the probability of such an event from occurring, we can find the basis that maximizes the distance between the covariance matrix and the origin, for example, for some choice of local measurement basis $\{\Pi_{\bar{x}} = \bigotimes_{q_i} \Pi_{x_i}^{q_i}\}$, we hope to find

$$\arg \max_{\{\Pi_{\bar{x}}\}} \text{tr} \left(C_{\mathcal{N}}^\dagger C_{\mathcal{N}} \right). \quad (41)$$

However, note that Remark 3 still applies and there exist correlations that cannot be fully observed from the covariance matrix alone. However, this does not interfere with inferring the network topology.

From here, the procedure for inferring the network topology is identical to that of the mutual information-based protocol. We provide a binary description for each pair of qubits in a network—1 if the covariance is above zero, and 0 (or within an acceptable statistical margin of error) otherwise. If two qubits are entangled, then their covariance is nonzero as long as the two qubits are not measured in mutually unbiased bases, for example, σ_x and σ_z . Thus, the protocol straightforwardly applies by replacing the qubit characteristic matrix $Q_{\mathcal{N}}$ with the covariance matrix $C_{\mathcal{N}}$.

Remark 4.—The covariance methods presented in Refs. [27–29] aim at constructing a single covariance matrix after collecting classical measurement data. In comparison, our protocol sequentially queries the network and finds the optimal covariance matrix that maximizes the correlation observed among all pairs of qubits. Our method avoids the possibility of choosing a measurement basis that leads to a small observable correlation, though at the cost of requiring more extensive procedures like variational optimization.

B. Comparing entropic and covariance-based protocols

While the mutual information and covariance both quantify the amount of correlation between qubits, each method possesses unique attributes that make one more preferable than the other depending on the context. In the following section, we will compare the two methods in terms of their computational complexity, the ability to accurately infer the topology under noisy quantum channels, and under statistical noise from taking a finite number of measurements. We will also demonstrate the capability of both methods on quantum hardware.

1. Computational complexity

The number of circuit evaluations needed to evaluate the covariance matrix is constant in the number of qubits. Constructing the characteristic matrix using classical mutual information needs twice the number of evaluations as the diagonal and off-diagonal entries must be constructed independently. Lastly, in the case of using the bipartite measured mutual information, the number of circuit evaluations grows quadratically in the number of qubits. One can optimize this to achieve linear scaling by evaluating all nonoverlapping pairs, but is still time consuming compared to the two constant-time alternatives.

The time complexity needed to compute the cost functions ranks similarly. Computing the distance induced by the Schmidt norm of the covariance matrix from the origin

requires scanning through each entry of the matrix, resulting in a quadratic scaling with respect to the number of qubits. Computing mutual information requires specifying the marginal distribution of all qubit pairs. One can think of the measurement outcome as a probability distribution with a domain that grows exponentially with the number of qubits. Obtaining the marginal distribution for a pair of qubits requires summing over exponentially many terms, and therefore is a time-consuming procedure. In the case of measured mutual information, one can choose to query one pair of qubits at a time, sacrificing query complexity for ease of computing the cost function.

C. Inference under noisy quantum channels

Though conceptually identical, covariance is generally less affected by noisy quantum channels compared to mutual information. Consider a Bell state undergoing depolarizing channels applied qubitwise. We can then write the mutual information and covariance analytically in terms of the noise parameter $\gamma \in [0, 1]$.

$$I_m(q_1; q_2) = \frac{1 + (1 - \gamma)^2}{2} \log(1 + (1 - \gamma)^2) + \frac{\gamma(2 - \gamma)}{2} \log(\gamma(2 - \gamma)), \quad (42)$$

$$\text{Cov}(q_1, q_2) = (1 - \gamma)^2. \quad (43)$$

The same can be done for amplitude-damping noise applied onto each qubit, which has Kraus operators

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1 - \gamma} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} \quad (44)$$

for each qubit, and noise parameter $\gamma \in [0, 1]$. The maximal correlation is observed when measuring both qubits in the σ_x basis, which, respectively, takes the form

$$I_m(q_1; q_2) = \frac{2 - \gamma}{2} \log(2 - \gamma) + \frac{\gamma}{2} \log \gamma, \quad (45)$$

$$\text{Cov}(q_1, q_2) = 1 - \gamma. \quad (46)$$

More detailed calculations are deferred to Appendix C. We compare these theoretic quantities with numerical solutions obtained via variational optimization. We repeated ten independent optimization trials per choice of noise parameter, and the results are shown in Fig. 6.

Both analytical and numerical results show that covariance decays slower with respect to the noise parameter γ in the case of depolarizing and amplitude-damping noise. This suggests a noise regime where mutual-information-based protocols misidentify the existence of entanglement while the covariance matrix can accurately infer the network topology. On a more optimistic note, we observe that

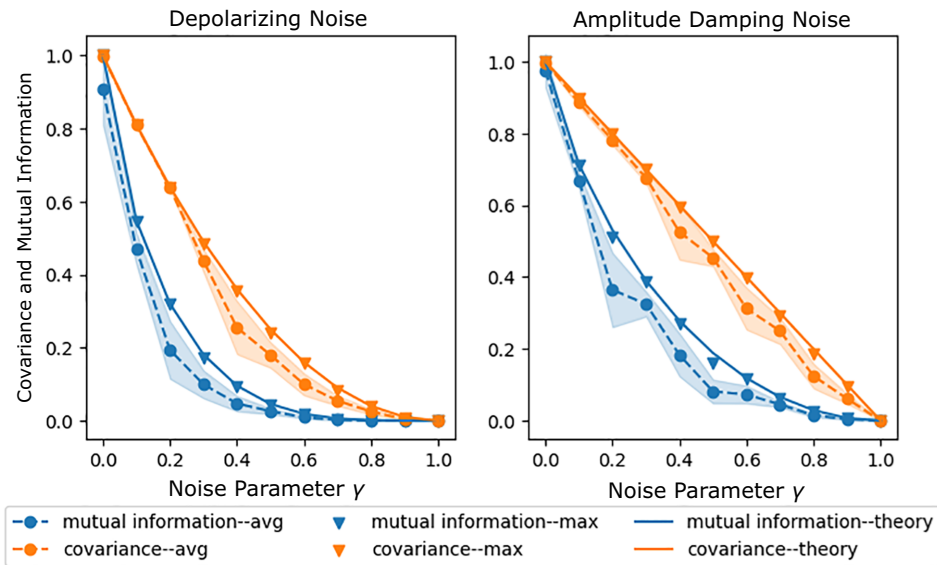


FIG. 6. The effect of noisy quantum channels on the observed correlation. A Bell state with fixed local rotations was prepared, and we applied (left) depolarizing noise to each qubit and (right) amplitude-damping noise to each qubit using PennyLane’s mixed-state simulator. We ran ten independent trials of variational optimization with step sizes of 0.05 for 30 steps, plotting the average (circles with shaded standard error) and maximum (triangle) observed across trials. We also compare the empirical value obtained through simulation with theory (solid lines), cf. Eqs. (42)–(43) and (45)–(46).

the respective theoretical values can be achieved numerically given an ensemble of independent optimization trials, and a typical optimization run produces a result that approximates the theoretical.

1. Inference under finite shot noise

In addition to noise induced by quantum channels, statistical shot noise is another practical source of uncertainty. In particular, having computational methods that can operate in the low-shot regime greatly increases its applicability as quantum resources are often scarce. Shot error can pose an issue for iterative methods such as variational optimization as noisy gradient estimates can limit the ability to effectively navigate the optimization landscape. We put our inference protocol to the test with a five-qubit network where the first three qubits are entangled in a W state, i.e., the W state takes the form

$$|W\rangle = \frac{1}{\sqrt{3}} (|001\rangle + |010\rangle + |100\rangle), \quad (47)$$

and the remaining two in a GHZ state. The optimization result for qubit pairs with and without entanglement is shown in Fig. 7.

Consistent with the results in Fig. 6, covariance methods identify the presence of entanglement more strongly than entropic measures. However, covariance tends to “overshoot” and falsely declares the presence of entanglement for uncorrelated qubits. We can set an arbitrary threshold for detecting correlations—we decide that two qubits

are entangled if and only if the calculated correlation (mutual information or covariance) is above the defined threshold. In the case of figure Fig. 7, the threshold is set to 0.05, indicated by the black line. When qubits are not entangled, mutual information rapidly vanishes while covariance stays above the threshold within the range of the shot counts considered. Thus, in the low-shot regime, mutual information might be more suitable as it is more likely to correctly reconstruct the entanglement structure of the underlying network.

2. Hardware experiments

We now apply on IBM’s quantum hardware our variational scheme for inferring network topology using local qubit measurements. As an example we again consider the five-qubit state preparation $|\psi\rangle = |W\rangle \otimes |\Phi\rangle$ where the $|W\rangle$ is the W state in Eq. (47) and $|\Phi\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle)$ is a two-qubit maximally entangled state. We also apply our variational network inference scheme to five-qubit GHZ states and the five-qubit zero state (see Appendix D). In each example, a known state is prepared while we optimize a variational ansatz that parameterizes arbitrary qubit projective measurements as $\Pi_x = U^\dagger(\vec{\theta}) |x\rangle\langle x| U(\vec{\theta})$ where $\vec{\theta} \in \mathbb{R}^3$. We then compare the performance of our variational inference scheme on both noisy IBM hardware and noiseless classical simulator when 10, 100, 1000, and 10 000 shots are considered.

To quantify the performance of the optimization, we consider the inference error, which we define as the

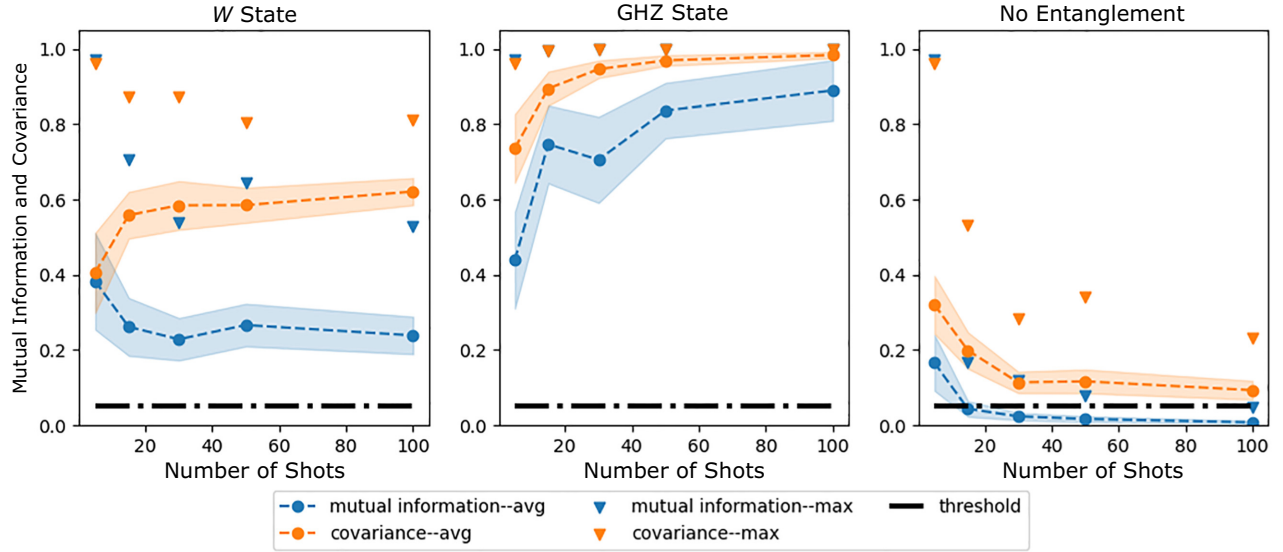


FIG. 7. The effect of statistical shot noise on calculating correlations under various prepared states. Mutual information and covariance are estimated using variational optimization with a step size of 0.05 for 30 steps. Each plot shows the correlation between qubit pairs under different entanglement structures. The threshold value (black line) is set to 0.05, and was chosen arbitrarily. From the experiment, we observe that covariance methods accurately identify the existence of entanglement, as shown in the case of (left) W states [Eq. (47)] and (middle) GHZ states [Eq. (1)]. However, entropic methods are able to identify (right) the lack of entanglement better than covariance-based protocols.

Euclidean distance

$$d(C, C^*) = \sqrt{\text{tr}[(C^* - C)^\top (C^* - C)]}, \quad (48)$$

where C is the optimized covariance matrix and C^* is the ideal covariance matrix for the given state. The distance in Eq. (48) can similarly quantify the error in the characteristic matrix $d(Q, Q^*)$. Note that this inference error quantifier only works in our numerical experiment because we prepare a known state. In practice, the state preparation is not known and the performance of the optimization cannot be quantified by Eq. (48).

For the considered state $|\psi\rangle = |W\rangle \otimes |\Phi\rangle$, the ideal qubit covariance and characteristic matrices are

$$C^* = \begin{pmatrix} 1 & \frac{2}{3} & \frac{2}{3} & 0 & 0 \\ \frac{2}{3} & 1 & \frac{2}{3} & 0 & 0 \\ \frac{2}{3} & \frac{2}{3} & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, Q^* = \begin{pmatrix} S_W & I_W & I_W & 0 & 0 \\ I_W & S_W & I_W & 0 & 0 \\ I_W & I_W & S_W & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \quad (49)$$

where $I_W \equiv \mathcal{I}_m(q_i; q_j) \approx 0.349976$ and $S_W \equiv S(W^{q_i}) \approx 0.918296$ for any of the qubits q_i and q_j of the state $|W\rangle$. To obtain the ideal covariance matrix and mutual information in Eq. (49), it is sufficient to measure all qubits

in the $\{|+\rangle, |-\rangle\}$ basis. While this calculation is straightforward for the $|\Phi\rangle$ state, we will be more explicit with the $W^{ABC} = |W\rangle\langle W|$ state, for which the reduced density matrices are

$$W^A = \text{tr}_{BC}(W^{ABC}) = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|1\rangle\langle 1| \quad (50)$$

and

$$W^{AB} = \text{tr}_C(W^{ABC}) = \frac{1}{3}|00\rangle\langle 00| + \frac{2}{3}|\Psi^+\rangle\langle\Psi^+|, \quad (51)$$

where $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$. When the observable σ_x is measured, the covariance and variance are

$$\text{Cov}(A, B) = \text{tr}((\sigma_x \otimes \sigma_x) W^{AB}) = \frac{2}{3} \quad (52)$$

and $\text{Var}(A) = 1$. The qubit von Neumann entropy can then be explicitly calculated from Eq. (50) as

$$S_W \equiv S(W^A) = -\frac{2}{3} \log \frac{2}{3} - \frac{1}{3} \log \frac{1}{3} \approx 0.918, \quad (53)$$

while the measured mutual information is found to be $I_W = \mathcal{I}_m(A, B) \approx 0.349976$ where each party measures the observable σ_x .

Since all qubits are measured in the same basis, the measured mutual information is equivalent to the classical mutual information for all qubit pairs, that is $\mathcal{I}_m(q_i, q_j) =$

$I(q_i, q_j)$). Thus, for this example, it is sufficient to optimize the classical bipartite mutual information rather than the measured mutual information. This simplification provides significant speed-ups when running on quantum hardware because only one circuit needs to be evaluated to collect the mutual information of all qubit pairs as opposed to the ten circuits needed to optimize the measured mutual information for each qubit pair independently. Note that queue times for the IBM hardware are the main bottleneck in our variational network inference scheme.

In Fig. 8, we plot the inference error as the covariance and characteristic matrices are optimized with respect to the state $|\psi\rangle = |W\rangle \otimes |\Phi\rangle$. To investigate the relation between the number of shots and the inference error, we consider optimizations where 10, 100, 1000, and 10 000 shots are used when collecting data from the quantum computer. We expect that the inference error should decrease as the number of shots increases.

As a baseline, we first run our numerical experiment on PennyLane’s default.qubit classical simulator, which is a noiseless, finite-shot simulation of a quantum

computer. The data shown in the top row of Fig. 8 plots the mean inference error from ten independent optimizations for both the covariance and characteristic matrices. As expected, the amount of inference error decreases as the number of shots increases. Furthermore, we find that the covariance and characteristic matrices do not always find a global optimum in which all matrix terms converge to their maximal theoretical values. As a result, the mean does not approach zero, reflecting the importance of running the optimization algorithm multiple times. Moreover, we find that the covariance matrix is optimized in fewer iterations than the characteristic matrix.

We then test our optimization on the `ibmq_belem` 5-qubit quantum computer, which exhibits a considerable amount of noise. When we run the optimization on an IBM quantum computer, the optimization steps become expensive to run due to the queue wait times. As a result, we are only able to run one optimization for each number of shots. We plot the IBM hardware optimization results in the middle row of Fig. 8. For the 10-shot case, it is not clear that the inference error is decreasing. For larger numbers

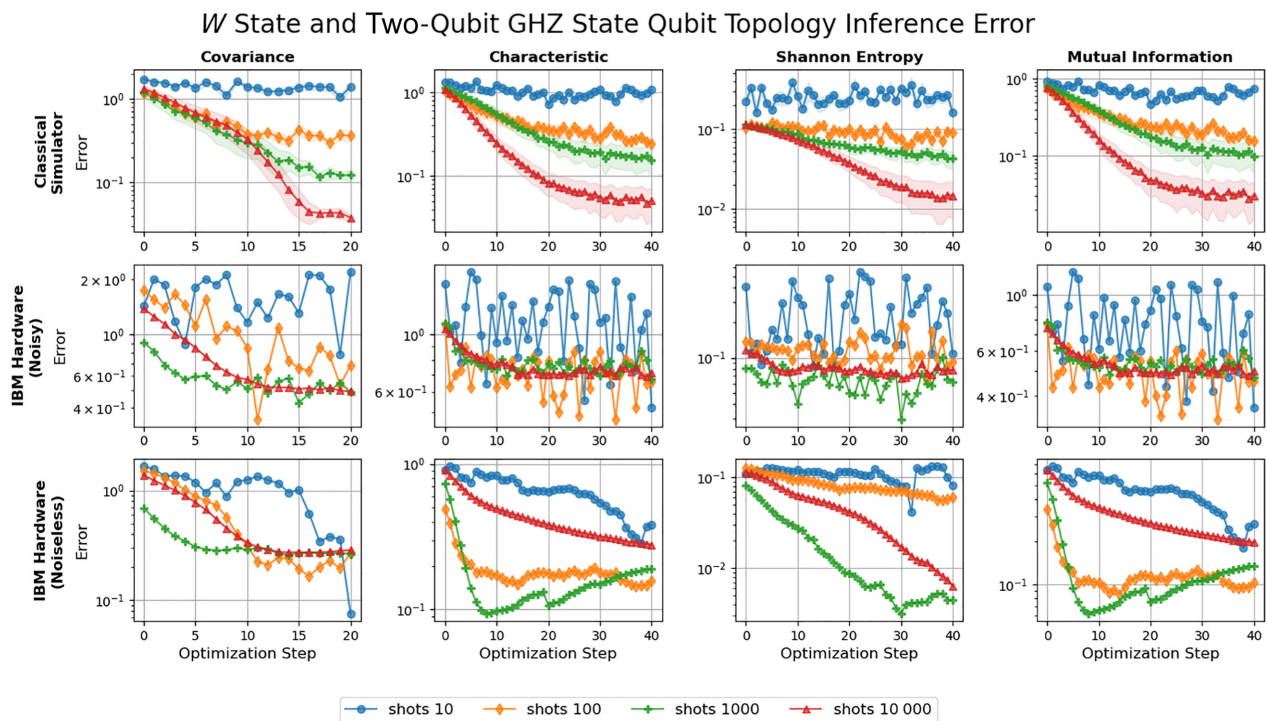


FIG. 8. Variational quantum optimization of the covariance matrices and characteristic matrices for a W state and two-qubit GHZ state prepared on IBM Hardware. From left to right each column plots the covariance matrix optimization, the characteristic matrix optimization, the Shannon entropy minimization, and the classical mutual information maximization. Note that the Shannon entropy and mutual information optimizations are combined to construct the characteristic matrix. In each plot the blue circles show the ten-shot optimization, the orange diamonds show the 100-shot optimization, green plus signs show the 1000-shot optimization, and red triangles show the 10 000-shot optimization. The x axis shows the optimization step while the y axis shows the inference error calculated as the Euclidean distance between the ideal covariance and characteristic matrix and the matrix in each optimization step. The first row shows the optimization data averaged over ten runs on a finite-shot noiseless classical simulator, the second row shows the optimization data collected from the `ibmq_belem` quantum computer, and the third row shows the data collected when the settings from the noisy IBM hardware optimization are re-evaluated on a noiseless classical simulator.

of shots, the inference error decreases throughout the optimization. To further validate the optimization results, we take the settings optimized on the noisy IBM hardware and re-evaluate them on a noiseless, infinite-shot classical simulator. Overall, there is a general improvement in the inference error as shown in the bottom row of Fig. 8 when compared with the data collected from the noisy IBM hardware. Although the 1000-shot case seems to be an exception with the error increasing slightly when evaluating the same parameters on the simulator, it is worth noting that the optimal parameters on noisy hardware might differ slightly from optimal on a simulator; hence, we view this run as mostly a statistical outlier. As further confirmed by additional experimental results in the Appendix, while the data from the quantum hardware may seem noisy, the optimization over qubit measurements is indeed decreasing the inference error.

In Fig. 9 we show the optimal covariance and characteristic where the optimal matrix values are taken over all optimization steps. As the number of shots increases, so does the optimization’s ability to resolve the correlation structure with greater accuracy. On the other hand, when the number of shots is small, statistical fluctuations can lead to stronger correlations than present, leading to false-positive correlations. For example, in the ten-shot case of Fig. 9 the zero terms of the covariance and characteristic matrices are optimized to be quite large, which would lead a researcher to infer that the two qubits are correlated. Also note that the IBM hardware can have significant errors on a given qubit for example, in the 10 000-shot characteristic matrix the top left matrix term remains close to zero despite taking more shots than other more successful trials. From experiments on the simulator, we know that mutual information generally is more difficult to optimize than covariance. Moreover, a noisy

device might not be stable for the amount of time needed to acquire 10 000 shots. This instability leads to noise that inhibits the optimization’s ability to make progress toward an optimum, particularly so for computing the mutual information. Therefore, unless one finds oneself working in the low-shot regime, covariance-based method is generally preferred as it exhibits more reliable convergence when presented with adequate resources.

VI. CONCLUSION

In this work, we introduced protocols for distinguishing and inferring the topology of n -local quantum networks. The protocols construct matrices that encode entanglement structures, i.e., the characteristic matrix [cf. Eqs. (14) and (32)] and the covariance matrix. The entries of these matrices can be estimated using only local measurements, which allows for easy implementation on quantum hardware. Assuming sources prepare GHZ states, the characteristic matrix can uniquely determine the topology of a quantum network. Moreover, if one is capable of making qubitwise measurements, the topology of the network can be inferred in polynomial time from both the characteristic matrix and the covariance matrix. Furthermore, the protocol is robust to noise and can be implemented on quantum hardware using quantum variational optimization. From experiments on both classical simulators and quantum hardware, we found that covariance-based methods are generally more stable during optimization, leading to more reliable discoveries of quantum entanglement. However, with limited shots, entropic protocols are more effective at avoiding false positives.

It is worth noting that the characteristic matrix cannot distinguish between quantum entanglement and shared

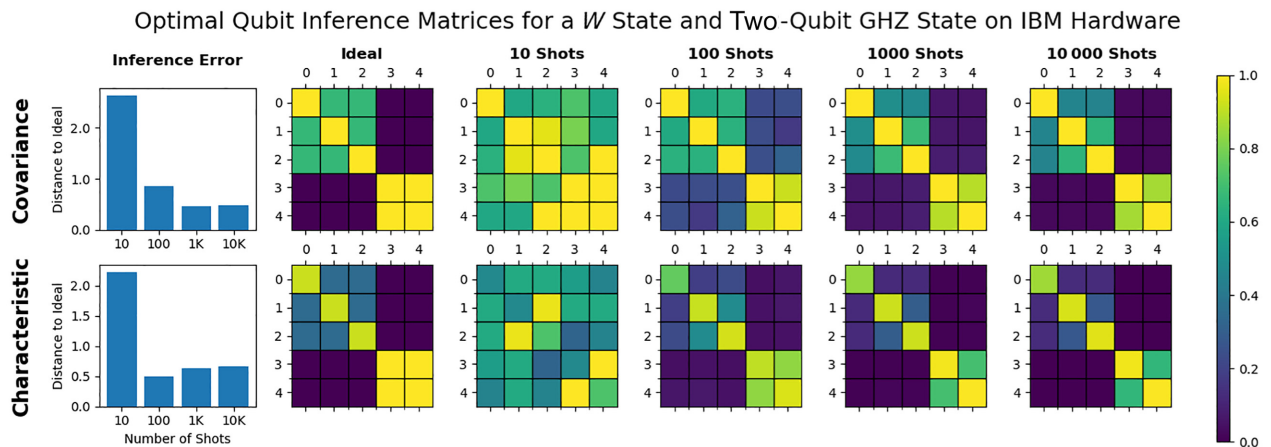


FIG. 9. In the top row, we plot the maximal value achieved for each term of the covariance matrix across all optimization steps. In the bottom row, we plot the maximal mutual information (off diagonals) and minimal Shannon entropies (diagonal) across all optimization steps. The bar graph plots the Euclidean distance between the ideal and inferred matrices for each distinct number of shots.

randomness. However, the characteristic matrix does indicate which qubits are correlated. Thus, an entanglement witness can be tailored to the network's topology. One approach might be to then test each source independently using an entanglement witness of choice [11,12].

Future work should also focus on relaxing assumptions made in the paper for even broader applications. For example, we assumed all measurement nodes are observed and measurements can be performed on all nodes. In reality, the known set of measurement nodes can merely be a subset of the entire network. Exploring the limits and extensions of our protocol for the case of partially observed networks can be a fruitful future direction. On the other hand, the protocol for distinguishing network topology (cf. Sec. III) relies on preparing the GHZ state, which is delicate and lacks robustness in practice. Extending the analysis to sources potentially distributing states with greater entanglement robustness, such as the W-class states [42,43], cluster states, or other partially entangled states would increase the applicability of this protocol. Another direction is to extend the protocol to infer network topology in more complex scenarios, for example, networks with intermediate processing nodes between source and measurement nodes, or communication between measurement devices and sources.

CODE AVAILABILITY

Our data, numerical methods, and software tools are publicly available on GitHub [36].

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, and the Office of Advanced Scientific Computing Research, Accelerated Research for Quantum Computing program under Contract No. DE-AC02-06CH11357. We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

APPENDIX A: PROOF OF LEMMA 3

In order to show Lemma 3, we use the following established results reviewed in Lemmas 5 and 6.

Lemma 5.—Let σ_n be an n -qubit shared uniform random bit, and let the classical probability distribution upon measuring σ_n be $\mathbb{P}(\vec{a})$. Then, the following inequality is true for any measurement basis:

$$H(\mathbb{P}(\vec{a})) \geq 1 \quad (\text{A1})$$

with equality occurring when measured in the computational basis.

Proof.—Recall that the entropy of the measurement of a state is minimized when measured in its eigenbasis [34]. An eigenbasis of σ_n is the computational basis, which behaves like a classical coin flip upon measuring. Thus, the Shannon entropy is lower bounded by 1. ■

Lemma 6.—Consider the distribution acquired through local measurements on a Bell state, that is

$$\mathbb{P}(\vec{a}) = \langle \Phi | \Pi_{a_1} \otimes \Pi_{a_2} | \Phi \rangle \quad (\text{A2})$$

for projective operators $\{\Pi_{a_1}\}$ and $\{\Pi_{a_2}\}$. Then, for any choice of Π_{a_1} and Π_{a_2} ,

$$H(\mathbb{P}(\vec{a})) \geq 1, \quad (\text{A3})$$

where the equality holds when measured in the computational basis.

Proof.—By subadditivity [34], the Shannon entropy of measuring $|\Phi\rangle$ is at least the entropy of its marginal, which is σ_1 . Since σ_1 is a classical coin flip, it has a Shannon entropy of 1 and the entropy of $|\Phi\rangle$ is lower bounded by 1. ■

Now, we present the proof for Lemma 3 below.

Proof.—Recall the definition of measured mutual information in Eq. (9). We must first determine the basis to measure in at each node, that is $\Pi_{\vec{a}_i}^{A_i}$ and $\Pi_{\vec{a}_j}^{A_j}$. Moreover, for any two nodes A_i and A_j , the qubits received can be either maximally entangled or independent (maximally mixed). By Assumption 1A, the measurement basis does not influence the entropies at each node. Thus, we can achieve the lower bounds in Lemmas 5–6 by measuring in the computational basis.

Knowing the basis of choice, we proceed to understand the graph-theoretic properties of the mutual information. Let $N_s^{A_i}$ be the number of sources connected to device A_i . Then, we know that $H(\mathbb{P}(\vec{a}_i)) = N_s^{A_i}$ and similarly with $H(\mathbb{P}(\vec{a}_j))$ where $\mathbb{P}(\vec{a}_i)$ and $\mathbb{P}(\vec{a}_j)$ are probability distributions upon measuring at node A_i and A_j , respectively.

For the joint entropy $H(\mathbb{P}(\vec{a}_i, \vec{a}_j))$, partition $A_i \cup A_j$ into sets:

$$S_1 = \{(q_k, q_\ell) : q_k \in A_i, q_\ell \in A_j, \{q_k, q_\ell\} \subseteq \Lambda_i \text{ for some } i\}, \quad (\text{A4})$$

$$S_2 = \{q_k : \forall q_\ell, \{q_k, q_\ell\} \notin S_1\}. \quad (\text{A5})$$

Each pair of qubits in S_1 are entangled and act jointly as a fair coin flip. On the other hand, each qubit in S_2 acts independently of one another also like a fair coin flip. Moreover, each element in S_1 and S_2 is independent of each other and Shannon entropies are additive. Thus, $H(\mathbb{P}(\vec{a}_i, \vec{a}_j)) = |S_1| + |S_2|$. Since the total number of qubits is $2|S_1| + |S_2| = H(\mathbb{P}(\vec{a}_i)) + H(\mathbb{P}(\vec{a}_j))$, the joint

entropy can be expressed as

$$H(\mathbb{P}(\vec{a}_i, \vec{a}_j)) = H(\mathbb{P}(\vec{a}_i)) + H(\mathbb{P}(\vec{a}_j)) - |S_1|. \quad (\text{A6})$$

By definition of S_1 , we find the measured mutual information to be

$$\mathcal{I}_m(A_i; A_j) = |S_1| = N_s^{A_i, A_j}. \quad (\text{A7})$$

■

APPENDIX B: PROOF OF THEOREM 1

Proof.—For sufficiency, observe that if two networks have the same topology, then the number of sources connected to $A_i^{(1)}$ is the same as $A_i^{(2)}$, and the number of sources shared between $A_i^{(1)}$ and $A_j^{(1)}$ is the same as $A_i^{(2)}$ and $A_j^{(2)}$. By Lemmas 2 and 3, the von Neumann entropy of each node and measured mutual information of pairs of nodes are the same.

We now show necessity. Suppose the von Neumann entropy at each measurement node and the measured mutual information between any pair of measurement nodes for the two networks are identical.

First, we note that the two networks must have the same number of nodes; an immediate contradiction is reached otherwise. On the other hand, the two networks will also have the same number of links, which can be written alternatively using the von Neumann entropy as $N_\ell = \sum_i S(A_i)$ using Assumption 1A. Lastly, the number of sources must also be the same. Suppose not, and $N_s^{(1)} < N_s^{(2)}$. Since the number of links present in either network is the same, there must be at least one link ℓ connected to the preparation node in $\mathcal{N}^{(1)}$ that is not connected to the same preparation node in $\mathcal{N}^{(2)}$. This means that the measurement device A_i that was connected to A_j via ℓ in $\mathcal{N}^{(1)}$ must have lost the connection to A_j in $\mathcal{N}^{(2)}$. Thus, the $\mathcal{I}_m(A_i^{(1)}; A_j^{(1)}) > \mathcal{I}_m(A_i^{(2)}; A_j^{(2)})$ and we have reached a contradiction.

Now, focus on the case where the two networks have the same number of sources, links, and nodes. Define a *triplet* in a network to be a tuple of three elements, (A_i, Λ_k, A_j) , such that A_i and A_j share Λ_k . For two n -local quantum networks, we construct the map ξ such that

- (i) $\xi((A_i^{(1)}, \Lambda_k^{(1)}, A_j^{(1)})) = (A_i^{(2)}, \Lambda_k^{(2)}, A_j^{(2)})$ where $(A_i^{(1)}, \Lambda_k^{(1)}, A_j^{(1)})$ is in $\mathcal{N}^{(1)}$ and $(A_i^{(2)}, \Lambda_k^{(2)}, A_j^{(2)})$ is in $\mathcal{N}^{(2)}$, and
- (ii) performing the map ξ on all triplets t_n in $\mathcal{N}^{(1)}$ yields $\mathcal{N}^{(2)}$.

Think of ξ as the map that “moves” $\mathcal{N}^{(1)}$ to $\mathcal{N}^{(2)}$. In particular, when a triplet is present in both networks, we take ξ to be the identity; when a triplet is only present in one network, then ξ moves the corresponding triplet into a new location. This restriction is important since this removes

the case that ξ cyclically moves edges, e.g., $(1, 1, 2) \mapsto (2, 1, 3), (2, 1, 3) \mapsto (3, 1, 1),$ and $(3, 1, 1) \mapsto (1, 1, 2)$.

This map can induce relabeling of indices. Let $\phi : [N_s] \rightarrow [N_s]$ be the map that relabels the source indices, and $\varphi : [N_m] \rightarrow [N_m]$ be the one for measurement nodes. Then, we can define ϕ and φ using ξ : if $\xi(A_i^{(1)}, \Lambda_k^{(1)}, A_j^{(1)}) = (A_i^{(2)}, \Lambda_k^{(2)}, A_j^{(2)})$, then define

$$\phi(\Lambda_k^{(1)}) = \Lambda_k^{(2)}, \quad \varphi(A_i^{(1)}) = A_i^{(2)}, \quad \varphi(A_j^{(1)}) = A_j^{(2)}. \quad (\text{B1})$$

If ϕ and φ are well defined, then they are immediately bijective because we are comparing networks of equal sizes—nodes have to be mapped to (onto) and two nodes cannot be “squished” and become indistinguishable (one-to-one). Therefore, we will show that if ξ preserves the characteristic matrix, then ϕ and φ are well defined and the two networks have the same topology. We do both by contradiction.

If ϕ is not well defined, then there are i, j, m, n ($i \neq m$) such that

$$\xi(A_i^{(1)}, \Lambda_k^{(1)}, A_j^{(1)}) = (A_i^{(2)}, \Lambda_k^{(2)}, A_j^{(2)}) \quad (\text{B2})$$

but

$$\xi(A_m^{(1)}, \Lambda_k^{(n)}, A_n^{(1)}) = (A_m^{(2)}, \Lambda_\ell^{(2)}, A_n^{(2)}) \quad (\text{B3})$$

for some $\ell \neq k$. However, this means that connections between $A_i^{(1)}$ and $A_m^{(1)}$ through $\Lambda_k^{(1)}$ are moved away upon applying ξ and

$$\mathcal{I}_m(A_i^{(2)}, A_m^{(2)}) \leq \mathcal{I}_m(A_i^{(1)}, A_m^{(1)}) - 1, \quad (\text{B4})$$

which contradicts the assumption that ξ preserves the measured mutual information.

On the other hand, if φ is not well defined, then there are j, k, ℓ, m ($k \neq \ell$) such that

$$\xi(A_i^{(1)}, \Lambda_k^{(1)}, A_j^{(1)}) = (A_i^{(2)}, \Lambda_k^{(2)}, A_j^{(2)}) \quad (\text{B5})$$

but

$$\xi(A_i^{(1)}, \Lambda_\ell^{(1)}, A_n^{(1)}) = (A_m^{(2)}, \Lambda_\ell^{(2)}, A_n^{(2)}) \quad (\text{B6})$$

for some $m \neq i$. However, this means that the connection to $A_i^{(1)}$ from $\Lambda_\ell^{(1)}$ is moved away so

$$S(A_i^{(1)}) \leq S(A_i^{(2)}) - 1, \quad (\text{B7})$$

which contradicts the assumption that ξ preserves the Shannon entropy. Thus, if two networks have the same characteristic matrix, then the two networks have the same topology. ■

APPENDIX C: MUTUAL INFORMATION AND COVARIANCE OF CHANNEL NOISE

Let ρ be the density matrix of the Bell state. We hope to calculate the mutual information and covariance between the two qubits after applying depolarizing and amplitude-damping noise independently on each qubit. We approach the calculation the same way. We first write down the density operator through the respective quantum channels, $\mathcal{E}(\rho)$, via the Kraus operator formalism. Then, we will find the classical distribution induced by measurement in the basis that maximizes the correlation quantity of interest.

1. Depolarizing noise

Quantum channel that applies qubit-depolarizing noise can be written as the following summation:

$$\mathcal{E}(\rho) = \sum_{ij} (K_i \otimes K_j) \rho (K_i \otimes K_j)^\dagger, \quad (\text{C1})$$

where $\{K_i\}$ are the Kraus operators of a single-qubit depolarizing channel

$$\begin{aligned} K_1 &= \sqrt{\frac{3\gamma}{4}} \mathbb{I}_2, K_2 = \sqrt{\frac{\gamma}{4}} \sigma_x, \\ K_3 &= \sqrt{\frac{\gamma}{4}} \sigma_y, K_4 = \sqrt{\frac{\gamma}{4}} \sigma_z. \end{aligned} \quad (\text{C2})$$

Expanding the summation, the quantum state $\mathcal{E}(\rho)$ can be compactly written as

$$\mathcal{E}(\rho) = \frac{\gamma(2-\gamma)}{4} \mathbb{I}_4 + (1-\gamma)^2 \rho. \quad (\text{C3})$$

Measuring the above state in the σ_z basis yields the following probability distribution:

$$\mathbb{P}(x_1 x_2) = \begin{cases} (1 + (1-\gamma)^2)/4 & \text{if } x_1 = x_2 \\ \gamma(2-\gamma)/4 & \text{otherwise} \end{cases}. \quad (\text{C4})$$

The remainder calculation follows the definition of measured mutual information and covariance, which will lead to Eqs. (42) and (43).

2. Amplitude damping

We follow a similar procedure in the case of applying amplitude-damping noise. Again, we can write the noisy

quantum state similar to Eq. (C1) but with Kraus operators

$$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1-\gamma \end{pmatrix}, K_2 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}. \quad (\text{C5})$$

Expanding the summation yields the following density matrix:

$$\mathcal{E}(\rho_\Phi) = \frac{1}{2} \begin{pmatrix} 1+\gamma^2 & 0 & 0 & 1-\gamma \\ 0 & \gamma(1-\gamma) & 0 & 0 \\ 0 & 0 & \gamma(1-\gamma) & 0 \\ 1-\gamma & 0 & 0 & (1-\gamma)^2 \end{pmatrix}, \quad (\text{C6})$$

and measuring in the σ_x basis—the basis that maximizes the observed correlation—is equivalent to measuring the state

$$\frac{1}{4} \begin{pmatrix} 2-\gamma & \gamma & \gamma & 2\gamma^2-3\gamma+2 \\ \gamma & \gamma & 2\gamma^2-\gamma & \gamma \\ \gamma & 2\gamma^2-\gamma & \gamma & \gamma \\ 2\gamma^2-3\gamma+2 & \gamma & \gamma & 2-\gamma \end{pmatrix} \quad (\text{C7})$$

in the computational basis, that is, the induced classical distribution is encoded in the diagonal. Following definitions of mutual information and covariances, respectively, again yields Eqs. (45) and (46).

APPENDIX D: INFERENCE OF FIVE-QUBIT ZERO STATE AND FIVE-QUBIT GHZ STATE IBM HARDWARE

In this section, we present the remainder of the results from our numerical experiments on IBM quantum hardware. Namely, we show the inference error when our variational scheme is applied to a five-qubit zero state and a five-qubit GHZ state. The former isolates the optimization of the von Neumann entropy, while the latter isolates the optimization of covariance and mutual information. We follow the same approach used for the inference of the $|W\rangle \otimes |\Phi\rangle$ as discussed in the main body of the paper.

In Figs. 10 and 11 we plot the network inference for a five-qubit zero state preparation $|\psi\rangle = |00000\rangle$. Since $|\psi\rangle$ is separable, for any two-qubit measurement $\Pi^{q_i} \otimes \Pi^{q_j}$, the mutual information $I(q_i; q_j) = 0$ and the covariance is $\text{Cov}(q_1, q_2) = 0$. As a result, randomly initialized measurements are optimal as shown in the simulated results in the top-right plot of Fig. 10. On the other hand, the variance of any qubit is $\text{Var}(q_i) \in [0, 1]$ and the Shannon entropy $H(\mathbb{P}(a_i)) \in [0, 1]$ where the maximal variance and minimal Shannon entropy are both achieved only when the state $|0\rangle\langle 0|$ is measured in the computational basis. Therefore, randomly initialized measurements must

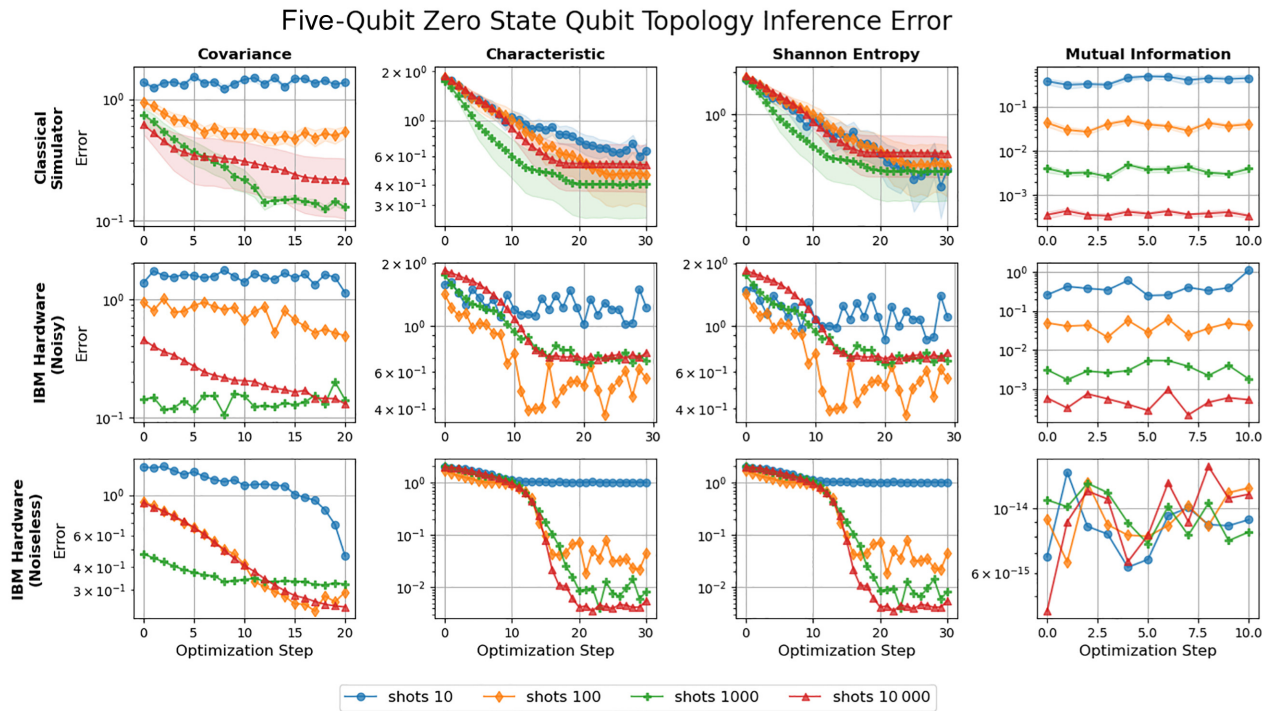


FIG. 10. Variational quantum optimization of the covariance matrices and characteristic matrices for a five-qubit zero state prepared on IBM Hardware. In each plot the blue circles show the ten-shot optimization, the orange diamonds show the 100-shot optimization, green plus signs show the 1000-shot optimization, and red triangles show the 10 000-shot optimization. The x axis shows the optimization step while the y axis shows the inference error calculated as the Euclidean distance between the ideal covariance and characteristic matrix and the matrix in each optimization step. The first row shows the optimization data averaged over ten runs on a finite-shot noiseless classical simulator, the second row shows the optimization data collected from the `ibmq_belem` quantum computer, and the third row shows the data collected when the settings from the noisy IBM hardware optimization are re-evaluated on a noiseless classical simulator. From left to right each column shows the covariance matrix optimization, the characteristic matrix optimization, the Shannon entropy optimization, and the classical mutual information optimization.

be optimized to achieve the optimal covariance matrix and characteristic matrix, which are the 5×5 identity and zero matrix, respectively.

As the number of shots increases we expect to see the network inference error decrease. This trend can easily be seen in the classical simulator data in the top row of Fig. 10. First, note in the rightmost column that a $10\times$ increase in the number of shots corresponds to roughly a $10\times$ decrease in the amount of error. In the covariance, characteristic, and Shannon entropy optimizations, we find a similar pattern in which more shots leads to less error, however, the separation is not as profound as in the mutual information case because the covariance and Shannon entropy optimizations over measurements have some error. This error is reflected in the standard error bar on the classical simulator optimization, which plots the average optimization over ten independent optimization runs. When we run the same optimizations on the IBM hardware, we find a similar trend where increasing the number of shots decreases the error.

In Fig. 11, we visualize the network inference performance by taking the optimal matrix elements for the covariance and characteristic matrices across the whole optimization on IBM hardware. For the covariance-based inference case, we find that the inference error decreases as the number of shots increases. This trend does not hold for the characteristic matrix, which shows the error increases when more than 100 shots are considered. We largely attribute this feature to the noise on the IBM hardware, which is quite dynamic and can vary considerably over the course of an optimization.

In Figs. 12 and 13 we plot the network inference for a five-qubit GHZ state preparation $|\psi\rangle = 1/\sqrt{2}(|00000\rangle + |11111\rangle)$. For any two-qubit measurement $\Pi^{q_i} \otimes \Pi^{q_j}$, the mutual information $I(q_i, q_j) \in [0, 1]$ and covariance $\text{Cov}(q_i, q_j) \in [0, 1]$ where these values are maximal when $\Pi^{q_i} = \Pi^{q_j}$ are the same measurement. As a result, randomly initialized measurements must be optimized to obtain their maximal values. On the other hand, the variance of any qubit is $\text{Var}(q_i) = 1$ and the Shannon entropy

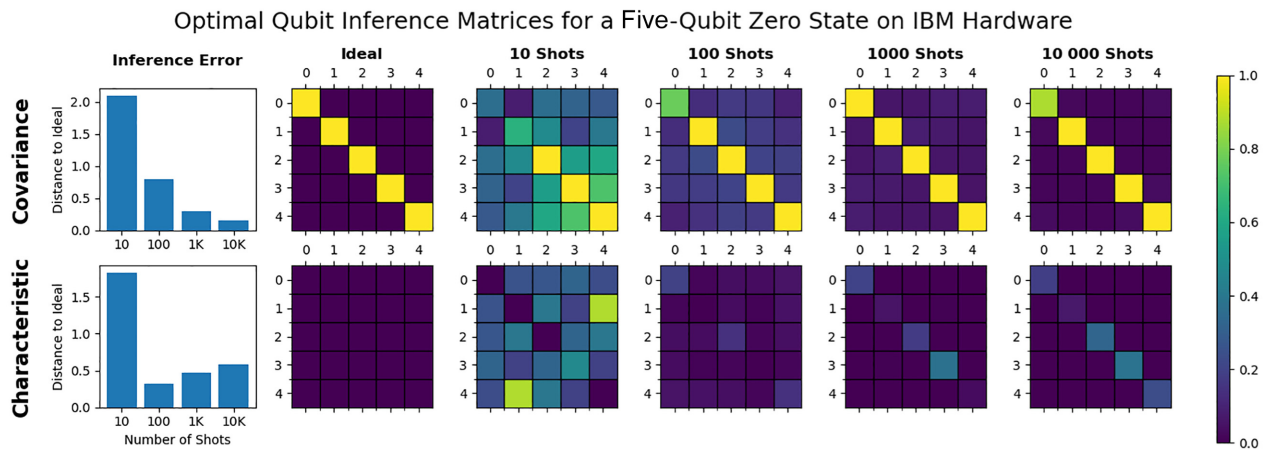


FIG. 11. In the top row, we plot the maximal value achieved for each term of the covariance matrix across all optimization steps. In the bottom row, we plot the maximal mutual information of each qubit pair (off diagonals) and minimal Shannon entropy of each qubit (diagonal) across all optimization steps. The bar graph plots the Euclidean distance between the ideal and inferred matrices for each distinct number of shots.

$H(\mathbb{P}(a_i)) = 1$ for any choice of measurement, hence randomly initialized measurements are optimal. Note that in the Shannon entropy plots of Fig. 12. The ideal qubit covariance and characteristic matrices are both 5×5 matrices of ones.

As the number of shots increases we expect to see the network inference error decrease. This trend can easily be seen in the classical simulator data in the top row of Fig. 12. For instance, note that for the Shannon entropy

optimization, a $10 \times$ increase in the number of shots corresponds to roughly a $10 \times$ decrease in the inference error, which is a result of the Shannon entropy being constant for all measurements. For the other optimizations, the 10 000-shot case does not necessarily perform the best, but this can largely be attributed to finding local optima during optimization, causing the mean inference error to level off with wide error bars. When we run the same optimizations on the IBM hardware, we find that more shots do

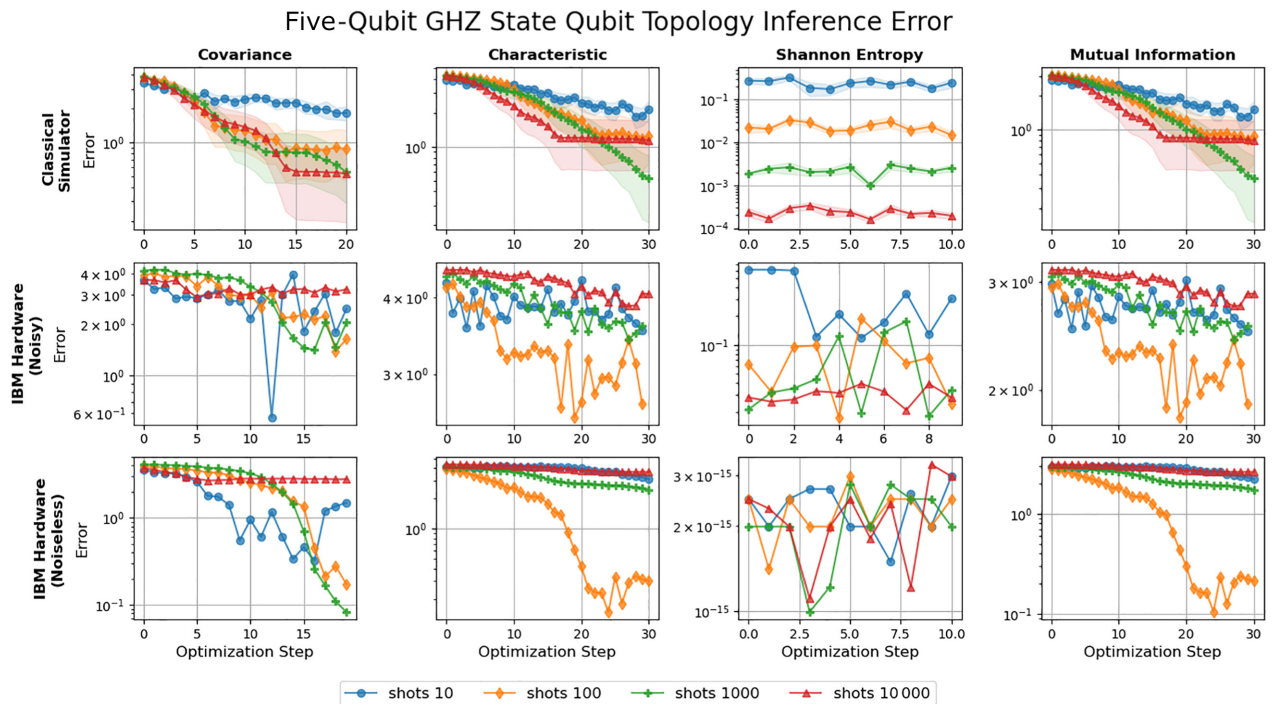


FIG. 12. Variational quantum optimization of the covariance and characteristic matrices for a five-qubit GHZ state prepared on IBM Hardware. See Fig. 10 for descriptions of individual plots.

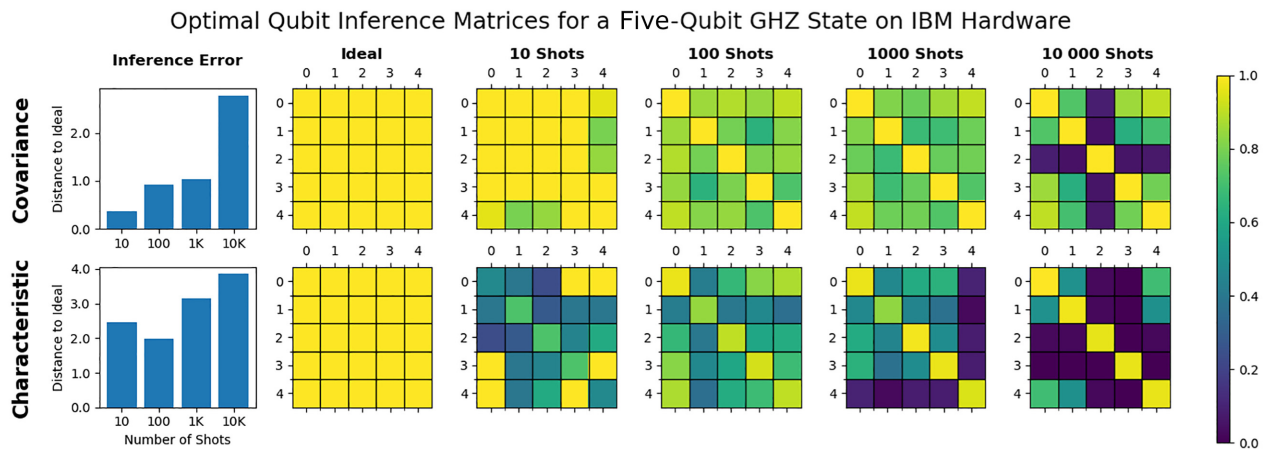


FIG. 13. In the top row, we plot the maximal value achieved for each term of the covariance matrix across all optimization steps. In the bottom row, we plot the maximal mutual information of each qubit pair (off diagonals) and minimal Shannon entropy of each qubit (diagonal) across all optimization steps. The bar graph plots the Euclidean distance between the ideal and inferred matrices for each distinct number of shots.

not necessarily decrease the inference error. We attribute this feature to the fact the IBM quantum hardware is noisy. Furthermore, the noise is not necessarily constant over the course of an optimization run. Thus, we suspect that considering too many shots allowed the noise to drift during optimization skewing the optimization. Furthermore, the numerical experiments were run serially from fewer shots to more shots, meaning that the performance of the IBM device might have also deteriorated considerably after the completion of the 100-shot experiment.

The noise in the IBM hardware is visualized in Fig. 13. For both the covariance and characteristic matrices, we find the inference error to increase as the number of shots increases. First, this trend results from the fact that we take the maximal covariance and mutual information for each qubit pair, thus, in the ten-shot case statistical fluctuations lead to a larger covariance and mutual information, implying less error. Second, we observe in the 1000- and 10 000-shot cases that the inference scheme fails for certain qubits. In these cases, either local optima are being found, or the IBM hardware is failing to produce entangled states. From our classical simulator data, we find that characteristic matrix optimizations are particularly susceptible to finding local optima, meaning that the results in Fig. 13 could simply be the optimization finding local optima. In practice, these errors can be mitigated by running the optimization repeatedly from different initial settings.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
 [2] L. Amico, R. Fazio, A. Osterloh, and V. Vedral, Entanglement in many-body systems, *Rev. Mod. Phys.* **80**, 517 (2008).

[3] J. P. Dowling and G. J. Milburn, Quantum technology: The second quantum revolution, *Philos. Trans. R. Soc. London, Ser. A: Mathematical, Physical and Engineering Sciences* **361**, 1655 (2003).
 [4] H. J. Kimble, The quantum internet, *Nature* **453**, 1023 (2008).
 [5] C. Simon, Towards a global quantum network, *Nat. Photonics* **11**, 678 (2017).
 [6] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, *Science* **362**, eam9288 (2018).
 [7] Q.-C. Sun, Y.-L. Mao, S.-J. Chen, W. Zhang, Y.-F. Jiang, Y.-B. Zhang, W.-J. Zhang, S. Miki, T. Yamashita, H. Terai, X. Jiang, T.-Y. Chen, L.-X. You, X.-F. Chen, Z. Wang, J.-Y. Fan, Q. Zhang, and J.-W. Pan, Quantum teleportation with independent sources and prior entanglement distribution over a network, *Nat. Photonics* **10**, 671 (2016).
 [8] S. L. N. Hermans, M. Pompili, H. K. C. Beukers, S. Baier, J. Borregaard, and R. Hanson, Qubit teleportation between non-neighbouring nodes in a quantum network, *Nature* **605**, 663 (2022).
 [9] G. Carvacho, E. Roccia, M. Valeri, F. B. Basset, D. Poderini, C. Pardo, E. Polino, L. Carosini, M. B. Rota, J. Neuwirth, S. F. C. da Silva, A. Rastelli, N. Spagnolo, R. Chaves, R. Trotta, and F. Sciarrino, Quantum violation of local causality in an urban network using hybrid photonic technologies, *Optica* **9**, 572 (2022).
 [10] J. Chung, E. M. Eastman, G. S. Kanter, K. Kapoor, N. Lauk, C. Peña, R. K. Plunkett, N. Sinclair, J. M. Thomas, R. Valivartha, S. Xie, R. Kettimuthu, P. Kumar, P. Spentzouris, and M. Spiropulu, Design and implementation of the illinois express quantum metropolitan area network, *IEEE Trans. Quantum Eng.* **3**, 1 (2022).
 [11] B. M. Terhal, Detecting quantum entanglement, *Theor. Comput. Sci.* **287**, 313 (2002).
 [12] O. Gühne and G. Tóth, Entanglement detection, *Phys. Rep.* **474**, 1 (2009).
 [13] M. Huber, F. Mintert, A. Gabriel, and B. C. Hiesmayr, Detection of high-dimensional genuine multipartite

- entanglement of mixed states, *Phys. Rev. Lett.* **104**, 210501 (2010).
- [14] C. Harney, S. Pirandola, A. Ferraro, and M. Paternostro, Entanglement classification via neural network quantum states, *New J. Phys.* **22**, 045001 (2020).
- [15] M.-X. Luo and S.-M. Fei, Robust multipartite entanglement without entanglement breaking, *Phys. Rev. Res.* **3**, 043120 (2021).
- [16] K. Hansenne, Z.-P. Xu, T. Kraft, and O. Gühne, Symmetries in quantum networks lead to no-go theorems for entanglement distribution and to verification techniques, *Nat. Commun.* **13**, 496 (2022).
- [17] F. Scala, S. Mangini, C. Macchiavello, D. Bajoni, and D. Gerace, in *International Joint Conference on Neural Networks* (IEEE, Pauda, 2022).
- [18] J. Henson, R. Lal, and M. F. Pusey, Theory-independent limits on correlations from generalized Bayesian networks, *New J. Phys.* **16**, 113043 (2014).
- [19] R. Chaves, C. Majenz, and D. Gross, Information-theoretic implications of quantum causal structures, *Nat. Commun.* **6**, 1 (2015).
- [20] M. Weilenmann and R. Colbeck, Analysing causal structures with entropy, *Proc. R. Soc. A: Mathematical, Physical and Engineering Sciences* **473**, 20170483 (2017).
- [21] R. Chaves, Polynomial Bell inequalities, *Phys. Rev. Lett.* **116**, 010402 (2016).
- [22] A. Tavakoli, Bell-type inequalities for arbitrary noncyclic networks, *Phys. Rev. A* **93**, 030101 (2016).
- [23] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, Nonlinear Bell inequalities tailored for quantum networks, *Phys. Rev. Lett.* **116**, 010403 (2016).
- [24] A. Tavakoli, A. Pozas-Kerstjens, M.-X. Luo, and M.-O. Renou, Bell nonlocality in networks, *Rep. Prog. Phys.* **85**, 056001 (2022).
- [25] M.-O. Renou, Y. Wang, S. Boreiri, S. Beigi, N. Gisin, and N. Brunner, Limits on correlations in networks for quantum and no-signaling resources, *Phys. Rev. Lett.* **123**, 070403 (2019).
- [26] M.-X. Luo, Network configuration theory for all networks, *arXiv:2107.05846* (2021).
- [27] A. Kela, K. V. Prillwitz, J. Aberg, R. Chaves, and D. Gross, Semidefinite tests for latent causal structures, *IEEE Trans. Inf. Theory* **66**, 339 (2020).
- [28] J. Åberg, R. Nery, C. Duarte, and R. Chaves, Semidefinite tests for quantum network topologies, *Phys. Rev. Lett.* **125**, 110505 (2020).
- [29] T. Kraft, C. Spee, X.-D. Yu, and O. Gühne, Characterizing quantum networks: Insights from coherence theory, *Phys. Rev. A* **103**, 052405 (2021).
- [30] E. Wolfe, R. W. Spekkens, and T. Fritz, The inflation technique for causal inference with latent variables, *J. Causal Inference* **7**, 20170020 (2019).
- [31] E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín, and M. Navascués, Quantum inflation: A general approach to quantum causal compatibility, *Phys. Rev. X* **11**, 021043 (2021).
- [32] X. Yang, Y.-H. Yang, and M.-X. Luo, Strong entanglement distribution of quantum networks, *Phys. Rev. Res.* **4**, 013153 (2022).
- [33] B. Doolittle, R. T. Bromley, N. Killoran, and E. Chitambar, Variational quantum optimization of nonlocality in noisy quantum networks, *IEEE Trans. Quantum Eng.* **4**, 1 (2023).
- [34] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2009).
- [35] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, edited by K. Kraus, A. Böhm, J. D. Dollard, and W. H. Wootters (Springer, Berlin, Heidelberg, 1983).
- [36] B. Doolittle, J. Larson, and D. Chen, qNetTI: the quantum network topology inferrer (2023), <https://github.com/ChitambarLab/qNetVO>.
- [37] V. Bergholm *et al.*, PennyLane: Automatic differentiation of hybrid quantum-classical computations, *arXiv:1811.04968* (2018).
- [38] B. Doolittle and T. Bromley, qNetVO: the Quantum Network Variational Optimizer, v0.1.3 (2022), <https://github.com/ChitambarLab/qNetTI>.
- [39] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [40] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [41] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [42] G. Vidal and R. Tarrach, Robustness of entanglement, *Phys. Rev. A* **59**, 141 (1999).
- [43] W. Dür, G. Vidal, and J. I. Cirac, Three qubits can be entangled in two inequivalent ways, *Phys. Rev. A* **62**, 062314 (2000).