# Unconditionally verifiable blind quantum computation

# Unconditionally verifiable blind quantum computation

Joseph F. Fitzsimons[1,2,*] and Elham Kashefi[3,4]

[1]*Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372*

[2]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543*

[3]*School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom*

[4]*LIP6, Departement Informatique et Reseaux, UPMC, 4 Place Jussieu, 75252 Paris CEDEX 05, France*

(Received 5 April 2017; published 5 July 2017)

Blind quantum computing (BQC) allows a client to have a server carry out a quantum computation for them such that the client's input, output, and computation remain private. A desirable property for any BQC protocol is verification, whereby the client can verify with high probability whether the server has followed the instructions of the protocol or if there has been some deviation resulting in a corrupted output state. A verifiable BQC protocol can be viewed as an interactive proof system leading to consequences for complexity theory. We previously proposed [A. Broadbent, J. Fitzsimons, and E. Kashefi, in *Proceedings of the 50th Annual Symposium on Foundations of Computer Science, Atlanta, 2009* (IEEE, Piscataway, 2009), p. 517] a universal and unconditionally secure BQC scheme where the client only needs to be able to prepare single qubits in separable states randomly chosen from a finite set and send them to the server, who has the balance of the required quantum computational resources. In this paper we extend that protocol with additional functionality allowing blind computational basis measurements, which we use to construct another verifiable BQC protocol based on a different class of resource states. We rigorously prove that the probability of failing to detect an incorrect output is exponentially small in a security parameter, while resource overhead remains polynomial in this parameter. This resource state allows entangling gates to be performed between arbitrary pairs of logical qubits with only constant overhead. This is a significant improvement on the original scheme, which required that all computations to be performed must first be put into a nearest-neighbor form, incurring linear overhead in the number of qubits. Such an improvement has important consequences for efficiency and fault-tolerance thresholds.

DOI: 10.1103/PhysRevA.96.012303

## I. INTRODUCTION

Scalable quantum computing has proven extremely difficult to achieve, and when the technology to build large-scale quantum computers does become available it is likely that they will appear initially in small numbers at a handful of centers. How will a user interface securely with such a quantum computer? A solution to this problem is offered by blind quantum computing, which enables a classical client (Alice) with limited quantum technology to delegate a computation to the quantum server(s) (Bob) in such a way that the privacy of the computation is preserved [1–6].

Blind classical computing (the notion of computing with encrypted data) was proposed by Feigenbaum [7] and then extended by Abadi *et al.* in a client server setting [8]. They showed that a randomized classical polynomial time client can encrypt and delegate general instances of certain problems in NP[1] to a powerful but untrusted server. Remarkably, they also proved that the decision of no NP-hard function can be encrypted in this way if unconditional security is required,[2] unless the polynomial hierarchy collapses at the third level. The idea of computing known circuits on encrypted data, while requiring the encryption and decryption procedures be independent of the complexity of the function to be evaluated, was introduced earlier by Rivest, Adleman, and Dertouzous in a scenario restricted to computational security [9] shortly after the invention of RSA [10]. The problem of creating such a scheme, known as fully homomorphic encryption, remained open for 30 years before being settled by Gentry [11], leading to one of the most active areas of research in modern cryptography [12].[3]

The first example of blind quantum computation was proposed by Childs [1] based on the idea of encrypting input qubits with a quantum one-time pad [19,20]. At each step, the client sends the encrypted qubits to the server, which applies a known quantum gate. Finally, the server returns the quantum state for the client to decrypt with their key. Cycling through a fixed set of universal gates ensures that the server learns nothing about the circuit. The next quantum blind protocol with the possibility of detecting a cheating server was proposed by Arrighi and Salvail [2]. In their scheme, the client gives the server multiple quantum inputs, most of which are *decoys* (not

*Corresponding author: joe.fitzsimons@nus.edu.sg



[1]A problem is in the class NP if one can verify its answers efficiently; it is NP-hard if it is as hard as any problem in NP.

[2]A cryptosystem is unconditionally (computationally) secure if it is secure even when the adversary has unlimited (restricted) computing power.

[3]While several attempts have been made in recent years to find homomorphic encryption schemes that allow for the evaluation of certain quantum operations [13–17], a quantum analog of fully homomorphic encryption remains elusive [18].

intended to be part of the desired computation), but rather are used to detect the server's deviation. This leads to a trade-off on the server side between gaining information and not disturbing the system and achieves cheat-sensitive security against individual attacks for a set of classical functions called random verifiable, where it is possible for the client to efficiently generate random input-output pairs. Extending these results, together with Broadbent, we presented a universal blind quantum computing (UBQC) protocol [3] in the measurement-based model [21,22], where the only requirement for the client is a classical computing machine and a very weak quantum instrument, a random single-qubit generator, a currently available technology as we have demonstrated recently [23]. Aside from the cryptographic scenario, a scheme based on a quantum authentication protocol[4] was proposed by Aharonov *et al.* [4], showing that any language in BQP has an interactive proof system with a verifier accessing a constant-size quantum computer. This work was complemented by a recent result of Reichardt *et al.* on the command of quantum systems via rigidity of CHSH games [5], leading to further work on device-independent verifiable blind quantum computing [24,25].

Recent years have seen an explosion of interest in the topic of blind quantum computing. This includes, for example, the extension of measurement-based UBQC to various settings [26–30], addressing key questions regarding the effect of the noise [31,32], the creation of protocols to optimize communications requirements [33–35], the development of privacy amplification techniques, similar to those applicable to quantum key distribution, to combat the adverse effect of imperfect devices on blindness [36], experimental demonstrations [23,37,38], and cryptographic applications [6,39].

A desirable property for any UBQC protocol is verifiability, whereby the client has a mechanism to verify the correctness of a delegated computation. The motivation for this stems from the broad range of computations that can be performed on a quantum computer. For problems that are in NP, the solution can be efficiently verified, at least in principle, using a witness. However, for other problems that can be efficiently computed using quantum computation, such as quantum simulation [40], a dishonest server cannot be detected in such a way. The ability to compute with encrypted data, while hiding the underlying function, has opened up alternative approaches to the problem of verification [3–5]. The main contributions of the present paper are to make rigorous the foundations of measurement-based UBQC and to present a verification protocol that we prove to be secure against the most general adversarial behavior of the server. Using this protocol, the client can verify with high probability whether Bob has followed the instructions of the protocol and the output state is indeed in the correct form or if there has been a deviation resulting in an incorrect output state. The central idea is based on the insertion of randomly prepared single qubits (called traps), blindly isolated from the actual computation, which can act as such a witness. Here, even the computation of the test (measurement of the qubits) can be performed blindly by an untrusted server as we have demonstrated recently [38].

The verification scheme we present here makes use of similar elements as suggested in [3]: Trap computations are used to detect errors and a fault-tolerant encoding of the computation is used to amplify the detection rate. While the proof sketch for the effectiveness of verification in the original UBQC paper did not consider the most general adversary, we prove that the modified scheme we present here detects or corrects any possible deviation by the server, except with probability that is exponentially suppressed. In order to do so we introduce universal resource states beyond the original brickwork state introduced in [3]. The first such family is a simple modification of the brickwork state that allows for the embedding of an arbitrary trap qubit, which leads to an inverse polynomial probability of detecting a deviation from the computation. In order to achieve a higher rate of detection, we introduce a second resource state that overcomes the locality limitations inherent in the brickwork state. This allows for the inclusion of a polynomial number of trap qubits and fault-tolerant implementation of the target computation based on the topological scheme of Raussendorf *et al.* [41]. Together, these two features allow for the probability of failing to detect or correct a deviation from the protocol to be made exponentially small. In this work we deal only with the stand-alone security definitions, as composable security follows from recent follow-up work by Dunjko *et al.* [42].

The remainder of the paper is organized as follows. Sections II and III summarize various required concepts from measurement-based quantum computing and also the original UBQC scheme presented in [3]. In order to construct our verifiable UBQC protocol we first introduce the concept of dummy qubits in Sec. IV, where we assume Alice now can prepare a qubit randomly chosen not only in the equatorial plain, as in the original UBQC scheme, but also from the set $\{|0\rangle,|1\rangle\}$. The latter qubits are called dummy qubits as they have no effect on the actual underlying computation. However, they permit the blind construction of isolated trap qubits in the state $|+_\theta\rangle$ as explained in Sec. VI, where the core concept of verification is introduced. In order to deal with both universality and verification, in Sec. V we introduce two resource states called the cylinder brickwork and dotted-complete graph states. The use of this scheme is expected to lead to substantially increased thresholds for fault-tolerant computing in the blind setting. A threshold for fault-tolerant blind computation in the absence of verification based on this fault-tolerance scheme was previously calculated as $4.3\times10^{-3}$ by Morimae and Fujii [31]. As shown in Sec. VI, introduction of a single blind isolated trap qubit leads to a verifiable blind quantum computing protocol with security polynomial in the total number of qubits. In order to boost the security while maintaining universality a different scheme has to be constructed. This is done in Sec. VII, where we put together various constructions of the previous sections to present the main result of this paper, a universal exponentially secure verifiable blind quantum computing protocol.

## II. PRELIMINARIES

Measurement-based quantum computing (MBQC) [21,22] is a form of quantum information processing where the key twin notions that distinguish quantum information processing

[4]The parties aim to communicate messages over an untrusted channel in such a way that the receiver can authenticate the sender.

from its classical counterpart, entanglement (creating nonlocal correlations between quantum elements) and measurement (observing a quantum system), are the explicit driving force of computation. More precisely, a measurement-based computation consists of a phase in which a collection of qubits is set up in a standard entangled state. Measurements are then made on individual qubits and the outcomes of the measurements may be used to determine further adaptive measurements. Finally, again depending on measurement outcomes, local adaptive unitary operators, called corrections, are applied to some qubits; this allows the elimination of the indeterminacy introduced by measurements. Conceptually MBQC separates the quantum and classical aspects of computation; thus it clarifies, in particular, the interplay between classical control and the quantum evolution process. The UBQC protocol explores this unique feature of MBQC as it has been proven to be conceptually enlightening to reason about distributed computing tasks using this approach [43]. We begin by describing all the required elements for an MBQC protocol and then move to the particular family of distributed MBQC protocols for hiding various aspects of a given computation.

### A. Single-party (undistributed) MBQC protocol

A formal language to describe in a compact way the operations needed for the MBQC model was proposed in [22]. In this framework every MBQC algorithm (usually referred to as an MBQC pattern) involves a sequence of operations such as entangling gates, measurements, and feedforwarding of outcome results to determine further measurement bases. A measurement pattern, or simply a pattern, is defined by a choice of a set of working qubits ($V$), a subset of input qubits ($I$), another subset of output qubits ($O$), and a finite sequence of commands acting on qubits in $V$. Therefore, we consider patterns associated with the so-called open graphs.

*Definition 1*. An open graph is a triplet $(G,I,O)$, where $G = (V,E)$ is a undirected graph and $I,O \subseteq V$ are respectively called input and output vertices.

Following the terminology of [22], a single-party MBQC protocol consists of three elements.

(i) A uniform family of open graph states $\{(G_{n,m},I_n,O_n)\}_n$ over $m$ vertices is associated with individual qubits, where $n$ is the size of the input and output space of the underlying computation. In this paper we deal only with those MBQC protocols that implement a unitary operator over their input space and hence the size of the output space is the same as the input space, but this is not a restriction and we can extend this treatment to any general completely positive trace preserving map by padding the input and output spaces. Further, for simplicity, we will assume that the input is always a pure state, though again this treatment can be extended to the general case. We usually assume that $|I| = |O| = n$, however sometimes $n$ is taken to be strictly larger than the dimension of the input and output Hilbert space due to the existence of auxiliary input or output qubits (as in later protocols that incorporate trap qubits). In order to have uniform notation, for the latter case, we will still use $I$ ($O$) to be the set of all nonprepared (nonmeasured) qubits where it is strictly larger than the class of all input (output) qubits. By the term "uniform family" we simply mean that for any protocol there exists a classical Turing machine that for a given input of the size $n$ describes the required graph over $m \geqslant n$ vertices. If the underlying geometry of the graph is regular, for example, being one-dimensional lines, two-dimensional regular lattices, or brickwork graphs (as we describe later), then instead of referring to the Turing machine to define the uniform family we simply use fixed parameters such as the size of the line or lattice to specify the graphs. For any fixed input size $n$ the graph $G_{n,m}$ describes the initial quantum state of the protocol. Given an arbitrary state of the input qubits corresponding to the input vertices of the graph, one prepares $m-n$ qubits in the state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ corresponding to all noninput qubits ($I^c$) in the graph and then apply CTRL-$Z$ operator between qubits $i$ and $j$, if the corresponding vertices in $G_{n,m}$ are connected. Note that since the CTRL-$Z$ gate is symmetric the direction of the edge is not important and hence we are working with undirected graphs. We will usually refer to the obtained quantum state based on the graph $G_{n,m}$ as the graph state $G_{n,m}$, unless a different notation is more appropriate; also for simplicity we drop the indices.

(ii) A set of angles $\phi_i \in A$, where $A \subseteq [0,\ 2\pi)$ for all nonoutput qubits, describes a collection of single-qubit $(X,Y)$ measurements, that is, measurement in the bases $\frac{1}{\sqrt{2}}(|0\rangle \pm e^{i\phi_i}|1\rangle)$. For the specific class of MBQC protocols that we discuss in this paper we require the angles to specify a collection of measurement bases such that individual measurements are unbiased with respect to the initial state. This is an essential ingredient for the blindness property that we define later. Without loss of generality, we can fix the set from which the angles are chosen to be $A = \{0,\pi/4,2\pi/4,\ldots,7\pi/4\}$. We will discuss later how this combination of angles and particular families of graph states leads to approximate universality.

(iii) The last ingredient is the structure of the dependence among the measurements. It is known that despite the probabilistic nature of the measurements, an MBQC protocol can implement a unitary computation over the input space by introducing a causal structure over the measurements. This is done by allowing any measurement on qubit $i$ to be dependent on the result of some (possibly none) previously measured qubits. Let $s_i \in \{0,1\}$ be the classical result of the measurement at qubit $i$. There are two type of dependences, called $X$ and $Z$ dependences. If a measurement at qubit $i$ is $X$ or $Z$ dependent on the $s_j$ where qubit $j$ has already been measured, then the actual angle of the measurement of qubit $i$ during the protocol run is $(-1)^{s_j}\phi_i$ or $\phi_i + s_j\pi$, respectively. Naturally one needs a noncyclic structure to be able to run such dependences and for an arbitrary graph such construction (if it exists) is formalized by the notion of the *flow* of the graph [44,45]. Intuitively, flow captures the propagation of quantum information as the resource state is measured, identifying the locations where measurement-dependent corrections should be made (see Fig. 1). A flow is defined by a function ($f : O^c \to I^c$) from the measured qubits to noninput qubits and a partial order ($\preceq$) over the vertices of the graph such that $i \preceq f(i) \forall i$ and $i \preceq j \forall j \in N_G(f(i))$, where $N_G(k)$ denotes the neighborhood of vertex $k$ in $G$. This last property enforces $f$ to be one to one. Each qubit $k$ is $X$ dependent on $f^{-1}(k)$ and $Z$ dependent on all qubits $l$ such that $k \in N_G(f(l))$. Note that if the dependence set is empty, that is, there is no qubit $q$ such that $q = f^{-1}(k)$ or $q \in N_G(f(l))$,

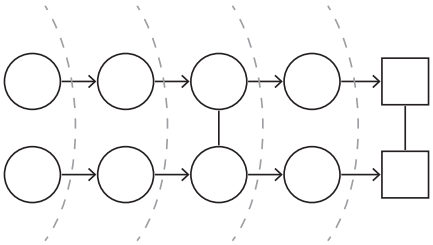

FIG. 1. Open graph state with flow. The boxed vertices are the output (nonmeasured) qubits and the circular vertices are the measured qubits. The flow function is represented as arrows (representing the $X$ dependence between measured qubits) and the partial order on the vertices (measurement order) is given by the dotted partition sets. One can see easily how the flow highlights the underlying circuit implemented by the measurement pattern.

then we set the convention that the corresponding value of $s_q$ is zero and hence we can use the same formulas $[(-1)^{s_j}\phi_i$ or $\phi_i + s_j\pi]$ to compute the dependent angles. For a given graph, once the input and output qubits have been labeled, the flow, if it exists, is uniquely determined.

The above describes only a nondistributed (single-party) MBQC protocol, that is, a protocol where a party both prepares the graph state and performs the sequence of the dependent measurements according to the order given by the flow (see [21,22] for more details on MBQC computation). One can easily extend the above definition to the distributed setting where different elements of the protocol are accessible and known only to specific parties and through classical-quantum communication the parties collaborate to perform a specific computation. Consider a simple two-party example where Alice has the information about the angles and Bob has the information about the graph and hence he can calculate the flow. Then they can collaborate to perform the corresponding computation as follows. First Bob prepares the required graph state and asks Alice to send him the classical information about the angles of the measurement. Bob then computes the dependence and performs the measurement and so forth. The purpose of this paper is to describe a family of such distributed protocols where, despite the communication, Alice can keep the measurement angles hidden from Bob. We then show that, for certain carefully chosen graph families, hiding these angles is sufficient to hide the full underlying computation together with the input and outputs.

### B. Two-party (distributed) hiding protocols

We define a specific family of two-party (Alice and Bob) MBQC protocols (which we term hiding protocols) that can be shown to be "blind" in the sense that Alice can hide information from Bob. For simplicity, instead of working with a family of graphs representing the computation over an arbitrary size input, we fix the input size to be $n$ and we define by $m \geqslant n$ the total number of vertices in the graph and hence the total number of qubits in the equivalent single-party protocol. Note that if we desire to have an efficient protocol, then we restrict the computation of the protocol to be of the polynomial size by requiring that $m = \mathrm{Poly}(n)$. However, blindness is independent of any complexity assumptions, so we do not, in general, restrict the size of $m$.

The protocol will be interactive, having $m - n$ steps if the output is quantum or $m$ steps if the output is classical, where at each step a single qubit is measured. In practice, we can parallelize the protocol to $D$ steps, where $D$ is the depth of the partial order of the flow of the graph [46,47]. This is due to the special structure of the partial order of the qubits defined by the flow function whereby all the qubits in the same class of the partial order are independent of each other and hence can be measured in parallel, i.e., at the same time. However, this parallelization will make no difference to the concept of blindness that we are concerned with, so we keep the simple convention that at each step only one qubit is measured. Furthermore, we assume for the case of classical output that all of the output qubits are measured in the final step with a Pauli $X$ measurement. Again, this is simply a convention for the discussion in our paper and in general the output qubits could be measured with any angles and in different steps depending on the flow construction. Such a convention does not affect universality, as the circuit being implemented can simply be modified to replace measurements in arbitrary bases with measurements in fixed bases preceded by an appropriate local rotation.

We will denote by $\mathbf{s}$ a sequence of length $m - n$ with value in $\{0,1\}$ describing the result of the nonoutput measurements performed so far. In the case of classical output, where output qubits are measured as the last $n$ steps, $\mathbf{s}$ is a sequence of length $m$. The value associated with a qubit that is not yet measured is set to 0 and hence at the beginning of the protocol before any measurement being performed we set $\mathbf{s} = 0,0,\ldots,0$. We will denote by $\mathbf{s}_{\leqslant i}$ the prefix of length $i$ of $\mathbf{s}$ and elements of $\mathbf{s}$ are denoted by $s_i$. Whenever adding the values of $s_i$ and $s_j$ we define their sum modulo 2. All the qubits in the protocol are enumerated in such a way that at position $i$ all qubits with label less than $i$ are measured before measuring qubit $i$. Any total ordering of the qubits consistent with partial ordering of the flow will work and as a result the measurement at qubit $i$ will depend only on the string $\mathbf{s}_{<i}$ .

We describe first a generic hiding protocol with quantum input and output (Protocol 1) and one with classical input and output (Protocol 2) and then formalize various derivatives of them to obtain universal, blind, and verifiable protocols. Protocol 2 is exactly the same as Protocol 1 except that the steps for encoding input are removed and all the output qubits are measured in the Pauli $X$ basis. We retain the common text between the protocols so that they can be understood individually. Note that the reason we choose the measurement of the output qubits to be in the Pauli $X$ basis is purely for simplicity of presentation so that the same evaluation function $C$ of the nonoutput measurements, in Protocol 1, can be used for the output qubits. However, one could add a separate evaluation function for the output qubit measurement to perform Pauli $Z$ measurement over them.

The outline of the main protocol is as follows. Alice has in mind a unitary operator $U$ that is implemented with a measurement pattern on some graph state $G$ with its unique flow function $f$ and measurement angles in $A = \{0,\pi/4,2\pi/4,\ldots,7\pi/4\}$. This pattern could have been designed either directly within the MBQC framework or via translation from a circuit construction. The pattern assigns a

measurement angle $\phi_i$ to each qubit in $G$; however, during the execution of the pattern, the actual measurement angle $\phi_i'$ is a modification of $\phi_i$ that depends on previous measurement outcomes instructed by $f$ in the following way [44,45]:

$$\phi_i' = (-1)^{s_{f^{-1}(i)}}\phi_i + \sum_{j:i\in N_G(f(j))} s_j\pi.$$

As said before, in a standard MBQC pattern all the noninput qubits are prepared in the state $|+\rangle$ and all the input qubits in the desired input state $|I\rangle$. Considering such quantum input allows for the possibility of Alice having additional capabilities allowing her to produce arbitrary input states or for the possibility that the input state is supplied on Alice's behalf by a third party.

In our protocols, in order to hide the information about the angles some randomness has to be added to the preparation and consequently the measurements have to be adjusted to compensate for this initial randomness to obtain the correct outcome. This randomization has three components: (i) a set of random angles $\theta$ used to hide the true measurement angles $\phi$, (ii) a set of random bits $r$ used to hide measurement outcomes, and (iii) a set of random bits $x$ used, along with $\theta$, to hide any quantum input via a one-time pad.

Alice prepares all the noninput qubits in $|+_{\theta_i}\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\theta_i}|1\rangle)$ for some randomly chosen $\theta_i \in A$ and also applies a modified version of a full quantum one-time pad encryption over the input qubits using random keys $x_i \in \{0,1\}$ and $\theta_i \in A$ as

$$|e\rangle = X_1^{x_1}Z_1(\theta_1)\otimes\cdots\otimes X_n^{x_n}Z_n(\theta_n)|I\rangle$$

before sending all qubits to Bob. After that, Bob entangles qubits according to $G$. Note that this unavoidably reveals upper bounds on the dimensions of the underlying quantum computation, corresponding to the length of the input and depth of the computation. The computation stage involves interaction: For each qubit, Alice sends Bob a classical message $\delta_i \in A$ to tell him in which basis [in the $(X,Y)$ plane] he should measure the qubit. This angle is computed in such a way as to correct for the one-time padding of the input qubits and the random rotation of the noninput qubits as

$$\delta_i = (-1)^{x_i+s_{f^{-1}(i)}}\phi_i + \sum_{j:i\in N_G(f(j))} s_j\pi + \theta_i + r_i\pi,$$

where the last term $r_i\pi$, with a randomly chosen $r_i \in \{0,1\}$, is added to hide the correct classical outcome of the measurement from Bob without affecting the overall computation (see the correctness proof below). Bob then performs the measurement and communicates the outcome $b_i$ to Alice. Alice's choice of angles in future rounds will depend on these values, hence she will correct the obtained outcome by setting $s_i := b_i \oplus r_i$. If Alice is computing a classical function, the protocol finishes when all qubits are measured (Protocol 2), as the classical outputs are encoded in the measurement outcomes sent to Alice. If she is computing a quantum function, Bob returns to her the final qubits (Protocol 1) and it is taken that the quantum output is encoded in these qubits. Note that in Protocol 2 we take the input to be $|+\rangle\otimes\cdots\otimes|+\rangle$, an encoding of the fixed classical input $0\cdots0$; any other arbitrary classical input $i_1\cdots i_n$ is prepared by applying appropriate $Z$ on the corresponding qubit to create

$$|e\rangle = Z_1^{i_1}\otimes\cdots\otimes Z_n^{i_n}(|+_{\theta_1}\rangle\otimes\cdots\otimes|+_{\theta_n}\rangle).$$

For classical input there is no need for a full one-time padding of the input, hence no need for the $x_i$ random variables as $\theta_i$ rotation completely hides the input. The above explanation is the basis for the correctness of all of the protocols presented in this paper.

*Definition 2.* A hiding protocol with quantum input is *correct* if the quantum output state is $U|I\rangle$ or if the classical outputs are the result of Pauli $X$ measurements on the state $U|I\rangle$, where $U$ is the unitary operator corresponding to the implementation of the measurement pattern of the hiding protocol. Similarly, one could define correctness for protocols with classical input.

*Theorem 1 (correctness).* Assume Alice and Bob follow the steps of Protocols 1 and 2. Then the outcome is correct.

*Proof.* The correctness of these protocol follows from the correctness of standard measurement-based quantum computation [22], as we now show. We explicitly give a proof only for the case of quantum input and output, as the remaining cases have virtually identical proofs. The protocol deviates in three ways from the standard implementation of the desired measurement pattern defined by a graph state $G$ with measurement angles $\phi_i$: a random $Z(\theta_i)$ rotation over all qubits, a random $X^{x_i}$ rotation over the input qubits, and measuring with angles $\delta_i$. However, since CTRL-$Z$ commutes with $Z$ rotations, Alice's preparation does not change the underlying graph state; only the phase of each qubit is locally changed and it is as if Bob had done the $Z$ rotation after the CTRL-$Z$. Let $\phi_i'$ be the adapted angles of the measurement $\phi_i$ according to the flow structure of the desired measurement pattern defined by $G$. Note that a measurement in the $\{|+_{\phi_i'}\rangle,|-_{\phi_i'}\rangle\}$ basis on a state $|\psi\rangle$ is the same as a measurement in the $\{|+_{\phi_i'+\theta_i}\rangle,|-_{\phi_i'+\theta_i}\rangle\}$ basis on $Z(\theta_i)|\psi\rangle$. Also a measurement in the $\{|+_{\phi_i'}\rangle,|-_{\phi_i'}\rangle\}$ basis on a state $|\psi\rangle$ is the same as a measurement in the $\{|+_{-\phi_i'}\rangle,|-_{-\phi_i'}\rangle\}$ basis on $X|\psi\rangle$. Finally, since $\delta_i = (-1)^{x_i}\phi_i' + \theta_i + \pi r_i$, if $r_i = 0$, Bob's measurement has the same effect as Alice's target measurement; if $r_i = 1$, all Alice needs to do is to flip the outcome. Therefore, all the deviation from the actual implementation of the measurement pattern are corrected and the quantum output is the desired state corresponding to the action of the unitary operator implemented by the graph state $G$ over the input state. ■

Note that, in practice, if Alice has the description of a unitary $V$ such that $V(\otimes_i|+\rangle) = |I\rangle$, then trivially a hiding protocol that blindly computes $UV$ over the input states $\otimes_i|+\rangle$ will prepare the desired output state of the form $U|I\rangle$. Therefore, for such a scenario Alice can follow the step of Protocol 1 with classical input without having to prepare the encoded state $X_1^{x_1}Z_1(\theta_1)\otimes\cdots\otimes X_n^{x_n}Z_n(\theta_n)|I\rangle$ herself. However, we have presented the full protocol for an arbitrary, possibly unknown, quantum input state, since the general scheme proved useful for dealing with input supplied by a third party [39].

## III. BLINDNESS

We say a hiding protocol is blind if Bob cannot tell anything relating to the angles of measurements. In considering this it

is worth noting that Bob can run the protocol only once with fixed values for Alice's parameters $\phi_i$, $\theta_i$, $r_i$, and $x_i$. Later we will show how for generic graphs this will lead to hiding the output of the computation as well. Following the convention of [8], we use the notation of a leakage function, denoted by $L(X)$, to formalize what Bob learns during the interaction. We present a stand-alone security definition that is equivalent to the original definition of blindness provided in [3].

*Definition 3*. A hiding protocol $\mathsf{P}$ with input $X$ is blind while leaking at most $L(X)$ if the distribution of messages obtained by Bob in $\mathsf{P}$ is dependent only on $L(X)$.

*Theorem 2 (blindness)*. Protocol 1 is blind while leaking at most $G$ and $n$ and Protocol 2 is blind while leaking at most $G$.

*Proof*. We first give a proof for the blindness of Protocol 1. We show that given $G$ and $n$ and independent of the actions of Bob, the message registers he receives are always in a maximally mixed state. We begin by introducing a new variable $\theta'_i = \theta_i + r_i\pi$ for all $i$. Thus, any quantum input received by Bob during a run of the protocol is given by $|e\rangle = X_1^{x_1} Z_1^{r_1} Z_1(\theta'_1) \otimes \cdots \otimes X_n^{x_n} Z_n^{r_n} Z_n(\theta'_n)|I\rangle$, while the remaining qubits he receives are in states $|+_{\theta'_i + r_i\pi}\rangle$ for $n < i \leqslant m$. Expressed in terms of $\theta'_i$, $\delta_i$ becomes independent of $r_i$ for all $i$, since

$$\delta_i = (-1)^{s_{f^{-1}(i)}}\phi_i + \sum_{j:\, i \in N_G(f(j))} s_j\pi + \theta'_i.$$

Thus, only the $i$th qubit received by Bob is dependent on $r_i$ and so tracing over the secret values $r$ simply dephases every qubit in the computational basis. Similarly, only qubit $i$ is dependent on $x_i$ for $1 \leqslant i \leqslant n$ and so tracing over $x$ completes the depolarization of the quantum input. Thus every qubit received by Bob is in the maximally mixed state and uncorrelated with all other qubits.

Next consider the classical communication used to convey measurement angles during the protocol. The computation of $\delta_i$ is composed of three terms. The first two terms $(-1)^{s_{f^{-1}(i)}}\phi_i$ and $\sum_{j:\, i \in N_G(f(j))} s_j\pi$ may depend implicitly on $b_k$ and $\delta_k$ for $k < i$, and on $r$ and $x$. However, note that the communication received up to step $i$ is independent of $\theta'_i$, the third term of $\delta_i$. Since $\theta'_i$ is uniformly random over $A$, $\delta_i$ must also be uniformly random and uncorrelated with previous communication sent to Bob. Thus, all communication in the protocol is uniformly random and uncorrelated, once the random keys ($x$, $r$, and $\theta$) are traced out, independent of the actions of Bob. An identical argument holds for Protocol 2, except that all $m$ qubits are assigned measurements, and hence $n$ is not revealed. ■

We note that the above definition is equivalent to a simulator-based definition, since once $L(X)$ is fixed, the distribution of messages Bob receives is also fixed. Hence, Alice could be replaced by a simulator with access only to $L(X)$ and this substitution could not be detected by Bob. A more detailed treatment of simulator-based definitions and composable security can be found in [42].

## IV. DUMMY QUBITS

In order to obtain an intuitive method for achieving verification, we construct an extension of Protocol 1 where Alice can also prepare qubits in the state $|z\rangle$ where $z$ is chosen uniformly at random from $\{0,1\}$. These qubits are called dummy qubits, as they will not be part of actual computation. A dummy qubit remains disentangled from the rest of the qubits of the graph state and, as we prove later, the addition of these dummy qubits does not affect the correctness or blindness of the hiding protocol. These dummy qubits are measured with random angles, which again will not affect the actual computation due to the fact that they are disentangled from the rest of the qubits. However, as we demonstrate in the next section, these dummy qubits allow Alice to easily create isolated trap qubits within the resource state to enable verification of the computation. Note that Alice must keep the position of the dummy qubits hidden from Bob (i.e., part of the secret) in order to keep the position of any trap qubits hidden. The addition of the dummy qubits can also be viewed as a method for the blind implementation of the Pauli $Z$ basis measurements. This is due to the fact that their position is hidden from Bob and from his point of view they are measured in the $(X,Y)$ plane as well. However, due to their preparation state ($|0\rangle$ or $|1\rangle$) through the entangling step, they have the same effect of measuring the corresponding qubit in the Pauli $Z$ basis. Therefore, we use the term blind Pauli $Z$ measurement interchangeably with dummy qubits in the rest of the paper. Due to the addition of dummy qubits, we will assume from now on that $n$ is an upper bound over the number of the input or output qubits. This is required to allow the possibility of having hidden trap or dummy qubits as part of the input or output system. Therefore, in the design of the measurement pattern, auxiliary qubits are added to the input and output space in such a way that the actual computation remains intact.

Protocol 1. Generic hiding protocol with quantum input and output.

(1) Alice's resources
  (i) Graph $G$ over $m$ vertices where labeling of vertices is in such a way that the first $n$ qubits are input and the last $n$ qubits are output.
  (ii) An $n$-qubit input state $|I\rangle$.
  (iii) A sequence of nonoutput measurement angles $\phi = (\phi_i)_{1 \leqslant i \leqslant (m-n)}$ with $\phi_i \in A$.
  (iv) $m$ random variables $\theta_i$ with values taken uniformly at random from $A$.
  (v) $n$ random variables $x_i$ and $m-n$ random variables $r_i$ with values taken uniformly at random from $\{0,1\}$.
  (vi) A fixed function $C_G$ that for each nonoutput qubit $i$ ($1 \leqslant i \leqslant m-n$) computes the angle of the measurement of qubit $i$ to be sent to Bob. This function depends on $\phi_i, \theta_i, r_i, x_i$, and the result of the measurements that have been performed so far ($\mathbf{s}_{<i}$). The function $C_G$ also depends on the flow $(f, \preceq)$ of the graph $G$. However, since the flow of the graph $G$ is unique (if it exists), we need not take flow as

a parameter of the function $C_G$. We have

$$C_G : \{1,\ldots,(m-n)\}\times A\times A\times\{0,1\}\times\{0,1\}\times\{0,1\}^{m-n} \to A,$$

$$(i,\phi_i,\theta_i,r_i,x_i,\mathbf{s}) \mapsto (-1)^{x_i+s_{f^{-1}(i)}}\phi_i + \sum_{j:i\in N_G(f(j))} s_j\pi + \theta_i + r_i\pi$$

where $x_k$ for $n+1 \leqslant k \leqslant m$ and also $s_k$ for any nondefined value of $k$ is set to zero.

(2) Initial step
   (i) Alice's move: Alice sends Bob the graph $G$ and sets all the values in $\mathbf{s}$ to be 0. Next she sends $m$ qubits in the order of the labeling of the vertices of the graph as follows: First, Alice encodes the $n$-qubit input state as
   $$|e\rangle = X_1^{x_1}Z_1(\theta_1)\otimes\cdots\otimes X_n^{x_n}Z_n(\theta_n)|I\rangle$$
   and sends them as the first $n$ qubits to Bob. She then prepares $m-n$ single qubits in the state $|+_{\theta_i}\rangle$ $(n+1 \leqslant i \leqslant m)$ and sends them to Bob as the remaining qubits.
   (ii) Bob's move: Bob receives $m$ single qubits and entangles them according to $G$.

(3) Step $i$: $1 \leqslant i \leqslant m-n$
   (i) Alice's move: Alice computes the angle $\delta_i = C_G(i,\phi_i,\theta_i,r_i,x_i,\mathbf{s})$ and sends it to Bob.
   (ii) Bob's move: Bob measures qubit $i$ with angle $\delta_i$ and sends Alice the result $b_i$.
   (iii) Alice's move: Alice sets the value of $s_i$ in $\mathbf{s}$ to be $b_i \oplus r_i$.

(4) Step $i$: $m-n+1 \leqslant i \leqslant m$
   (i) Bob's move: Bob sends qubit $i$ to Alice.
   (ii) Alice's move: Alice applies $X^{s_{f^{-1}(i)}}Z^{\sum_{j:i\in N_G(f(j))}s_j}Z(\theta_i)$ over qubit $i$.

*Theorem 3*. Assume Alice and Bob follow the steps of Protocol 3. Then the outcome obtained is the same as if the computation took place over the graph $G$ after removal of the dummy vertices in $D$, the set of positions of dummy qubits in $G$.

Protocol 2. Generic hiding protocol with classical input and output.

(1) Alice's resources
   (i) Graph $G$ over $m$ vertices where labeling of vertices are in such a way that the first $n$ qubits are input and the last $n$ qubits are output.
   (ii) An $n$-bit input string $c_1,\ldots,c_n$.
   (iii) A sequence of nonoutput measurement angles $\phi = (\phi_i)_{1\leqslant i\leqslant(m-n)}$ with $\phi_i \in A$.
   (iv) $m$ random variables $\theta_i$ with values taken uniformly at random from $A$.
   (v) $m$ random variables $r_i$ with values taken uniformly at random from $\{0,1\}$.
   (vi) A fixed function $C_G$ that for each nonoutput qubit $i$ $(1 \leqslant i \leqslant m)$ computes the angle of the measurement of qubit $i$ to be sent to Bob,
   $$C_G : \{1,\ldots,m\}\times A\times A\times\{0,1\}\times\{0,1\}^m \to A,$$
   $$(i,\phi_i,\theta_i,r_i,\mathbf{s}) \mapsto (-1)^{s_{f^{-1}(i)}}\phi_i + \sum_{j:i\in N_G(f(j))} s_j\pi + \theta_i + r_i\pi,$$
   where $s_k$ for any nondefined value of $k$ is set to zero and also $\phi_i = 0$ for $m-n+1 \leqslant i \leqslant m$.

(2) Initial step
   (i) Alice's move: Alice sends Bob the graph $G$ and sets all the values in $\mathbf{s}$ to be 0. Next she sends $m$ qubits in the order of the labeling of the vertices of the graph as follows: First, Alice encodes the $n$-bit string classical input $c_1,\ldots,c_n$ as state
   $$|e\rangle = Z_1^{c_1}\otimes\cdots\otimes Z_n^{c_n}(|+_{\theta_1}\rangle\otimes\cdots\otimes|+_{\theta_n}\rangle) = |+_{\theta_1+i_1\pi}\rangle\otimes\cdots\otimes|+_{\theta_n+i_n\pi}\rangle$$
   and sends them as the first $n$ qubits to Bob. She then prepares $m-n$ single qubits in the state $|+_{\theta_i}\rangle$ $(n+1 \leqslant i \leqslant m)$ and sends them to Bob as the remaining qubits.
   (ii) Bob's move: Bob receives $m$ single qubits and entangles them according to $G$.

(3) Step $i$: $1 \leqslant i \leqslant m$
   (i) Alice's move: Alice computes the angle $\delta_i = C_G(i,\phi_i,\theta_i,r_i,\mathbf{s})$ and sends it to Bob.
   (ii) Bob's move: Bob measures qubit $i$ with angle $\delta_i$ and sends Alice the result $b_i$.
   (iii) Alice's move: Alice sets the value of $s_i$ in $s$ to be $b_i \oplus r_i$.

*Proof*. The proof is similar to the proof of Theorem 1; the only new element is the effect of the dummy qubits. If a dummy qubit is in the state $|0\rangle$, then in the entangling step this qubit does not affect the state of the other qubits. However, if the dummy qubit is in the state $|1\rangle$ then the entangling operation will introduce a Pauli $Z$ rotation on all the neighboring qubits in $G$. Hence a qubit $i \notin D$ will be affected by the operator $\prod_{j\in N_G(i)\cap D} Z^{d_j}$. In the initial step, Alice already applied the operation $\prod_{j\in N_G(i)\cap D} Z^{d_j}$ over the prepared qubits and therefore all qubits $i \notin D$ are in the desired

Protocol 3. Generic hiding protocol with quantum input and output and dummy qubits.

(1) Alice's resources
(i) Graph $G$ over $m$ vertices where labeling of vertices is in such a way that all the $l$ input qubits are located among the first $n \geqslant l$ qubits and all the $l$ output qubits are located among the last $n$ qubits.
(ii) An $l$-qubit input state $|I\rangle$.
(iii) The dummy qubits positions, set $D$, chosen among all possible vertices except the $l$ input and $l$ output qubits.
(iv) A sequence of nonoutput measurement angles $\phi = (\phi_i)_{1\leqslant i\leqslant(m-n)}$ with $\phi_i \in A$, where $\phi_i = 0$ for all $i \in D$.
(v) $m$ random variables $\theta_i$ with values taken uniformly at random from $A$.
(vi) $l$ random variables $x_i$, $m-n$ random variables $r_i$, and $|D|$ random variables $d_i$ with values taken uniformly at random from $\{0,1\}$.
(vii) A fixed function $C_G$ that for each nonoutput qubit $i$ ($1 \leqslant i \leqslant m-n$) computes the angle of the measurement of qubit $i$ to be sent to Bob,

$$C_G : \{1,\ldots,(m-n)\}\times A\times A\times\{0,1\}\times\{0,1\}\times\{0,1\}^{m-n} \to A,$$

$$(i,\phi_i,\theta_i,r_i,x_i,\mathbf{s}) \mapsto (-1)^{x_i+s_{f^{-1}(i)}}\phi_i + \sum_{j:i\in N_G(f(j))} s_j\pi + \theta_i + r_i\pi,$$

where $x_k$ for $n+1 \leqslant k \leqslant m$ and $s_k$ for any nondefined value of $k$ are set to zero.

(2) Initial step
(i) Alice's move: Alice sends Bob the graph $G$ and sets all the values in $\mathbf{s}$ to be 0. Alice encodes the $l$-qubit input state as

$$|e\rangle = X_1^{x_1}Z_1(\theta_1)\otimes\cdots\otimes X_n^{x_l}Z_n(\theta_l)|I\rangle$$

and positions them among the first $n$ qubits. She then prepares the remaining qubits in the following form:

$$|d_i\rangle \forall i \in D,$$

$$\prod_{j\in N_G(i)\cap D} Z^{d_j}|+_{\theta_i}\rangle = \left|+_{\theta_i+\sum_{j\in N_G(i)\cap D} d_j\pi}\right\rangle \forall i \notin D.$$

Then Alice sends Bob all $m$ qubits in the order of the labeling of the vertices of the graph.
(ii) Bob's move: Bob receives $m$ single qubits and entangles them according to $G$.

(3) Step $i$: $1 \leqslant i \leqslant m-n$
(i) Alice's move: Alice computes the angle $\delta_i = C_G(i,\phi_i,\theta_i,r_i,\mathbf{s})$ and sends it to Bob.
(ii) Bob's move: Bob measures qubit $i$ with angle $\delta_i$ and sends Alice the result $b_i$.
(iii) Alice's move: Alice sets the value of $s_i$ in $\mathbf{s}$ to be $b_i \oplus r_i$.

(4) Step $i$: $m-n+1 \leqslant i \leqslant m$
(i) Bob's move: Bob sends qubit $i$ to Alice.
(ii) Alice's move: Alice applies $X^{s_{f^{-1}(i)}}Z^{\sum_{j:i\in N_G(f(j))} s_j}Z(\theta_i)$ to qubit $i$.

state $|+_{\theta_i}\rangle$, since $Z$ operator is self-inverse. Moreover, all the dummy qubits are unentangled with the rest of qubits and are measured in a random basis with no consequences for the part of the computation taking place over the graph $G$ after removing vertices $D$. ■

*Theorem 4*. The hiding protocol with dummy qubits, Protocol 3, is blind while leaking $G$ and $n$.

*Proof*. Proof follows along similar lines of Theorem 2. We define $\theta_i' = \theta_i + \pi r_i + \pi\sum_{j\in N_G(i)\cap D} d_i$. Alice's total communication to Bob consists of the initial quantum states, which we can rewrite as $|+_{\theta_i'-\pi r_i}\rangle$ if the qubit is not a dummy qubit or $\in_R \{|0\rangle,|1\rangle\}$ if it is a dummy qubit, and the measurement angles, which are set to be $\delta_i = \phi_i' + \theta_i' - \pi\sum_{j\in N_G(i)\cap D} d_i$. As before, the values of $\delta_i$ are uniformly random since $\theta_i'$ are uniformly random, and for any fixed values of $\delta_i$ tracing over all $r_i$, we obtain the initial quantum state for each qubit as either

$$\frac{1}{2}|+_{\theta_i'}\rangle\langle+_{\theta_i'}| + \frac{1}{2}|-_{\theta_i'}\rangle\langle-_{\theta_i'}| = \frac{\mathbb{I}}{2}$$

if the qubit was not a dummy or

$$\frac{1}{2}|0\rangle\langle0| + \frac{1}{2}|1\rangle\langle1| = \frac{\mathbb{I}}{2}$$

if the qubit was a dummy. Hence the qubits obtained by Bob are always in the maximally mixed state and are not correlated with each other. ■

## V. UNIVERSAL RESOURCE STATES

During a hiding protocol Bob learns the graph of entanglement $G$; however, it was shown in [3] that it is possible for Alice to choose a family of graphs corresponding to what were termed brickwork states such that blindness of the angles, as defined before, will permit Alice to hide the unitary operator that the protocol is implementing, revealing only an upper bound on the dimensions of the circuit required to implement it. The key element to achieve this is the use of those universal resources for MBQC [48] that are generic, hence revealing no information about the structure of the underlying computation, except the bounds on the size of input and the depth of the computation. Moreover, to make the protocol practical from Alice's point, it is desirable to restrict the class of measurement angles so that the required class of random qubits prepared by Alice is also restricted. Note that exact universal blind quantum computing could be achieved if Alice could prepare separable single-qubit states $|+_\theta\rangle$ with $\theta$ chosen randomly in

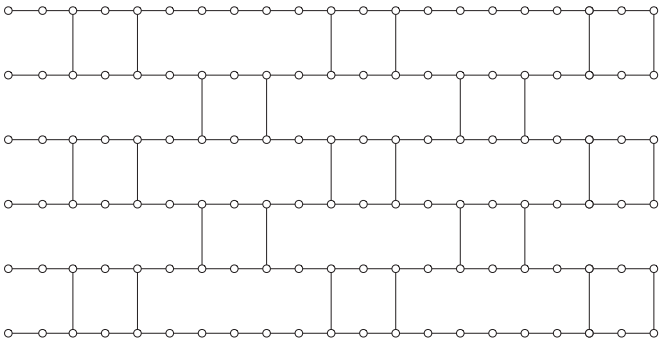

FIG. 2. Brickwork state $\mathcal{G}_{6\times 21}$. Qubits are arranged according to layer $x$ and row $y$, corresponding to the vertices in the above graph, and are originally in state $|+\rangle$. CTRL-$Z$ gates are then performed between qubits that are joined by an edge. A similar resource state was proposed in [50].

$[0,2\pi)$ and if Bob could make any measurement with angles in $[0,2\pi)$. Such a model requires Alice to communicate random real angles to Bob and hence such a setting is unattractive from a communications resources point of view. Similar to the quantum circuit scenario, by the Solovay-Kitaev theorem, a finite set of angles (for instance, a set that corresponds to Hadamard and $\frac{\pi}{8}$-phase gates) can be used to efficiently approximate any single-qubit unitary operator.[5] For the rest of this paper we will restrict our attention to approximate universality and we use the fact that a large family of graph states is approximately universal if one restricts the set of angles to be in the set $\{0, \pm\pi/4, \pm\pi/2\}$ [49]. We give two such examples below.

*Definition 4*. A brickwork state $\mathcal{G}_{n\times m}$, where $m \equiv 5$ or $1 \pmod 8$, is an entangled state of $n\times m$ qubits constructed as follows.

(i) Prepare all qubits in state $|+\rangle$ and assign to each qubit an index $(i,j)$, $i$ being a row ($i \in [n]$) and $j$ being a column ($j \in [m]$).

(ii) For each row, apply the operator CTRL-$Z$ on qubits $(i,j)$ and $(i,j+1)$ where $1 \leqslant j \leqslant m-1$.

(iii) For each column $j \equiv 3 \pmod 8$ and each odd row $i$, apply the operator CTRL-$Z$ on qubits $(i,j)$ and $(i+1,j)$ and also on qubits $(i,j+2)$ and $(i+1,j+2)$.

(iv) For each column $j \equiv 7 \pmod 8$ and each even row $i$, apply the operator CTRL-$Z$ on qubits $(i,j)$ and $(i+1,j)$ and also on qubits $(i,j+2)$ and $(i+1,j+2)$.

We will refer to the underlying graph of a brickwork state as the brickwork graph and denote it with the same notation by $\mathcal{G}_{n\times m}$ (see Fig. 2).

*Theorem 5 (universality)*. The brickwork state $\mathcal{G}_{n\times m}$ is universal for quantum computation. Furthermore, we only require single-qubit measurements under the angles $\{0, \pm\pi/4, \pm\pi/2\}$ to achieve approximate universality [3] and measurements can be done layer by layer.

*Proof*. The proof is straightforward (see details in [3]) based on constructing measurement patterns for elements of an approximate set of universal gates that could be tiled together as brickwork states as depicted in Fig. 3. ■

Let us denote vertices of a brickwork graph $\mathcal{G}_{n\times m}$ by $(i,j)$ (where $1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m$). Then it is easy to verify that the unique flow function of $\mathcal{G}$ is defined by

$$f_{\mathcal{G}}((i,j)) = (i,j+1).$$

That is to say, the flow of each vertex in the graph is from its immediate left neighbor in the same row. The corresponding partial order $\prec_{\mathcal{G}}$ is defined as the collection of sets $L_j$ of all vertices in the $j$th column of the brickwork graph

$$L_j = \{(x,y) | 1 \leqslant x \leqslant n, y = j\}.$$

Now suppose Alice has in mind a unitary operator $U$ of size $2^n \times 2^n$ and the $n$-qubit input state $|I\rangle$. Due to Theorem 5 there exist an integer $m$ and angles $\{\phi_{i,j}\}_{1\leqslant i\leqslant n, 1\leqslant j\leqslant m} \in A$ such that the measurement pattern with angles $\{\phi_{i,j}\}$ over the brickwork state $\mathcal{G}_{n\times m}$, where the first $n$ qubits are set to be in the state $|I\rangle$, approximates $U|I\rangle$. Therefore, the last $n$ qubits after the measurements of the first $m-n$ qubits and application of the corresponding corrections induced by flow are in a state that can be made arbitrarily close to $U|I\rangle$. We can simply adapt the generic hiding protocol to implement this measurement pattern blindly as presented in [3].

As mentioned in Sec. IV, in order to construct a verification scheme we make use of dummy qubits. While this presents

[5]More precisely, the Solovay-Kitaev theorem states that if the subgroup generated by some subset of SU(2) operators is dense in SU(2), then the approximation converges exponentially quickly to any element of SU(2) in the number of these operators from a smaller set one uses to approximate.

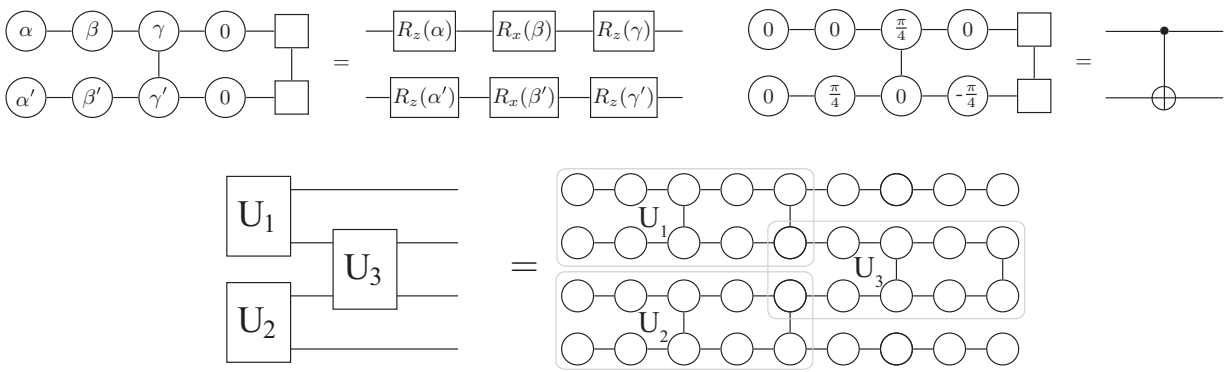


FIG. 3. Measurement patterns implementing arbitrary single-qubit rotations and the CNOT operator. These patterns can be composed within the brickwork state, as shown in the lower portion of the figure.

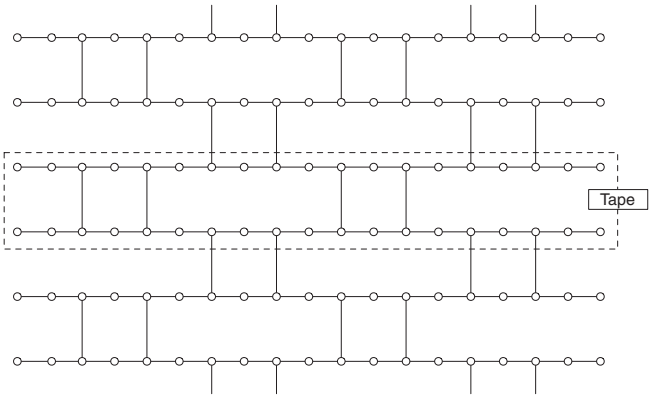


FIG. 4. Cylinder brickwork state $\mathcal{G}^{C}_{6\times 19}$.

a simple mechanism to achieve isolated trap qubits, the presence of trap and dummy qubits disrupts the computation. However, this can be fixed through a simple modification of the brickwork state.

*Definition 5*. A cylinder brickwork state $\mathcal{G}^{C}_{n\times m}$ is a modification of the brickwork state of size $n\times m$, for even $n$, where the first and the last rows are connected such that the regular brickwork structure is preserved, while introducing rotational symmetry. We will refer to the underlying graph of a cylinder brickwork state as the cylinder brickwork graph and denote it with the same notation by $\mathcal{G}^{C}_{n\times m}$ (see Fig. 4). A *tape* $\mathcal{T}_i$ in a cylinder brickwork graph is the subgraph induced by all the nodes of $i$th and $(i+1)$th rows.

The cylinder brickwork state allows for a simple construction for trap-based verification, as discussed in Sec. VI. Next we introduce another generic family called dotted-complete graph states (see Fig. 5), which enables significant amplification of the probability of detecting deviations from the computation, particularly in the case of quantum output, as discussed in Sec. VII. The basic idea behind this universal resource state is that it can be partitioned blindly into smaller universal resource states, one of which will be used for the computation, while the others will be used as traps for verification purposes. To begin with, we need to introduce the graphs that we will use and prove that they have some special properties.

*Definition 6*. We define the operator $\sim(G)$ on graph $G$ to be the operator that transforms a graph $G$ to a new graph denoted by $\tilde{G}$ by replacing every edge in $G$ with a new vertex connected to the two vertices originally joined by that edge. Let $K_N$ denote the complete graph of $N$ vertices. We call the quantum state corresponding to the graph $\tilde{K}_N$ the dotted-complete graph

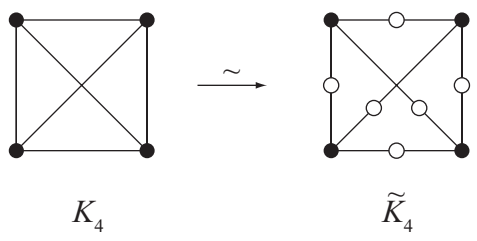


FIG. 5. Example of the relationship between a complete graph $K_4$ and the corresponding dotted-complete graph $\tilde{K}_4$. The vertices in black in $\tilde{K}_4$ denote the set $P(\tilde{K}_4)$, while the white vertices correspond to $A(\tilde{K}_4)$.

state denoted by $\tilde{\mathcal{K}}_N$. We denote the set of vertices of $\tilde{K}_N$ previously inherited from $K_N$ by $P(\tilde{K}_N)$ and the vertices added by the $\sim()$ operation by $A(\tilde{K}_N)$. The number of the vertices in the $\tilde{K}_N$ graph is then equal to $N(N+1)/2$.

The following definition and lemmas will be used in manipulation of dotted-complete graph states.

*Definition 7*. We define the bridge operator on a vertex $v$ of degree 2 on graph $G$ to be the operator that connects the two neighbors of $v$ and then removes vertex $v$ and any associated edges from $G$. We define the break operator on a vertex $v$ of graph $G$ to be the operator that removes vertex $v$ and any associated edges from $G$. Let $G$ be a graph on $m$ vertices. Then we say that $G$ is $n$ universal, for $n \leqslant m$, if and only if any graph of $n$ vertices can be obtained from $G$ through a sequence of bridges and breaks.

*Lemma 1*. $\tilde{K}_N$ is $N$ universal and the bridge and break operations used to obtain a target graph need only be performed on vertices in $A(\tilde{K}_N)$.

*Proof*. Given any graph $G$ on $N$ vertices, associate each vertex $u_i$ in $G$ with a vertex $v_i$ in $P(\tilde{K}_N)$. Each pair of vertices $(v_i, v_j)$ in $P(\tilde{K}_N)$ is connected through an intermediate vertex of degree 2 in $A(\tilde{K}_N)$. Thus, by bridging over the intermediate vertex if $u_i$ and $u_j$ are joined by an edge and breaking the intermediate vertex otherwise, $\tilde{K}_N$ reduces to $G$. As this is true for all graphs $G$ on $N$ vertices, $\tilde{K}_N$ is $N$ universal. ■

*Lemma 2*. Given a partitioning of the vertices $P(\tilde{K}_N)$ into $n$ sets $\{P_i\}$ containing $N_i$ vertices, respectively, by applying a sequence of break operations only, it is possible to transform $\tilde{K}_N$ into $n$ disconnected graphs $\tilde{k}_i$ such that each one of them is of the form $\tilde{K}_{N_i}$ and $P(\tilde{k}_i) = P_i$.

*Proof*. As the vertices $P(\tilde{K}_N)$ are associated with a corresponding vertex in $K_N$, the vertices of $K_N$ can be partitioned into the sets $\{P_i\}$. As $K_N$ is the complete graph the vertices within each partition $P_i$ form a clique. Thus, by removing edges between the partitions the resulting graph is composed of $n$ disconnected graphs $\{k_i = K_{N_i}\}$ such that the vertices in $k_i$ are the vertices in $P_i$. As removing an edge before applying the $\sim()$ operator is equivalent to applying a break operation after the $\sim()$ operator there exists a corresponding sequence of break operations such that the resulting graph is $\sim(\{k_i\}) = \{\tilde{k}_i\}$. As $\tilde{k}_i = \sim(k_i)$, it follows that $P(\tilde{k}_i) = P_i$ and since $k_i = K_{N_i}$ then $\tilde{k}_i = \tilde{K}_{N_i}$ as required. ■

*Lemma 3*. Given a graph $\tilde{K}_N$, by applying break operators to every vertex in $P(\tilde{K}_N)$ or $A(\tilde{K}_N)$, the resulting graph is composed of the vertices of $A(\tilde{K}_N)$ or $P(\tilde{K}_N)$, respectively, and contains no edges.

*Proof*. As the $\sim()$ operation only introduces vertices connected to vertices in $P(\tilde{K}_N)$, every vertex in $A(\tilde{K}_N)$ shares edges only with vertices in $P(\tilde{K}_N)$. Thus, when the vertices in $P(\tilde{K}_N)$ and their associated edges are removed by the break operators, the vertices in $A(\tilde{K}_N)$ become disconnected. Similarly, since $\sim()$ removes all edges between vertices in $P(\tilde{K}_N)$, every vertex in $P(\tilde{K}_N)$ shares edges only with vertices in $A(\tilde{K}_N)$. Thus, when the vertices in $A(\tilde{K}_N)$ and their associated edges are removed by the break operators, the vertices in $P(\tilde{K}_N)$ become disconnected. ■

We now extend these results to graph states.

*Lemma 4*. Given two graph states $|\psi_{G_1}\rangle$ and $|\psi_{G_2}\rangle$ corresponding to graphs $G_1$ and $G_2$, respectively, if it is possible to obtain $G_2$ from $G_1$ through a sequence of bridge

and break operations, then it is possible to obtain $|\psi_{G_2}\rangle$ from $|\psi_{G_1}\rangle$ through a sequence of Pauli measurements and local rotations about the $Z$ axis through angles from the set $\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$.

*Proof.* By measuring any qubit in a graph state with a Pauli $Z$ operator, we obtain a state equivalent up to local Pauli $Z$ corrections to the graph state obtained from the graph when that vertex and its associated edges are removed. To see this, we consider the operations this qubit undergoes: It is first prepared in a state $|+\rangle$, then it interacts with its neighbors via CTRL-$Z$ gates, and then it is measured in the $Z$ basis. As the measurement commutes with the entangling operation, this result is identical to the case where the CTRL-$Z$ gates are applied to the measured eigenstate of $Z$. Thus, when the complete sequence of events is taken into account, this operation is equivalent to the identity when the measurement outcome is 0 and equivalent to local Pauli $Z$ operators applied to the neighbors of the measured site when the measurement outcome is 1. This is then the graph state equivalent of the break operation defined on the associated graph.

If a vertex is of degree 2, then measuring the associated qubit with the Pauli $Y$ operator yields the graph state corresponding to the graph obtained by applying a bridge operation to that vertex, up to local $Z$ rotations through an angle $\pm\frac{\pi}{2}$. To see this, we again consider the sequence of operations the qubit undergoes: It is prepared in the state $|+\rangle$, interacts with its neighbors, and then is measured in the $Y$ basis. Immediately prior to measurement, the net operator applied is $\frac{1}{\sqrt{2}}|0\rangle \otimes \mathbb{I} + \frac{1}{\sqrt{2}}|1\rangle \otimes Z_1 \otimes Z_2$, where the subscripts 1 and 2 denote the neighbors of the measured qubit. Thus, if the measurement result is 0, then this is equivalent to directly applying the operator $e^{i(\pi/4)Z_1\otimes Z_2}$ to the neighboring qubits, whereas if the measurement result is 1 this is equivalent to applying the operator $e^{-i(\pi/4)Z_1\otimes Z_2}$ to these qubits. Since the CTRL-$Z$ gate can be written as either $e^{i(\pi/4)(\mathbb{I}-Z\otimes\mathbb{I}-\mathbb{I}\otimes Z+Z\otimes Z)}$ or $e^{-i(\pi/4)(\mathbb{I}-Z\otimes\mathbb{I}-\mathbb{I}\otimes Z+Z\otimes Z)}$, the effect on the neighboring qubits is equivalent to a CTRL-$Z$, up to local $Z$ rotations by $\frac{\pi}{2}$ (for a measurement result of 0) or $-\frac{\pi}{2}$ (for a measurement result of 1). This could also be derived via the stabilizer formalism. For a more detailed discussion of the effect of Pauli measurements in the measurement-based model, the reader is referred to [51]. ■

*Theorem 6 (universality).* The dotted-complete graph state $\tilde{\mathcal{K}}_N$ is universal for quantum computation. Furthermore, we only require single-qubit measurements under the angles $\{0, \pm\pi/4, \pm\pi/2\}$ and in the Pauli $Z$ basis to achieve approximate universality, and measurements can be done layer-by-layer.

*Proof.* Due to Lemmas 1 and 4, by choosing $N$ big enough, we could construct the brickwork state $\mathcal{G}_{n\times m}$ from $\tilde{\mathcal{K}}_N$ using only Pauli measurements. Hence, from Theorem 5 we obtain the universality of dotted-complete graph states and approximate universality with only single-qubit measurements under the angles $\{0, \pm\pi/4, \pm\pi/2\}$ (which include the Pauli $Y$ measurements required to implement bridge operations) and the Pauli $Z$ basis measurements required to implement break operations. ■

From this result we can construct a new universal hiding protocol based on dotted-complete graph states, as given in Protocol 4. Interestingly, in the case of classical input and output this new protocol does not even reveal the circuit dimensions, but instead a single integer that is an upper bound on the number of qubits required to implement the computation in the measurement-based model.

Protocol 4. Dotted-complete graph state universal hiding protocol with quantum input and output.

(1) Alice's resources
(i) Parameter $N$ such that the desired computation could be obtained from the state $\tilde{\mathcal{K}}_N$ after a sequence of break and bridge operators (Theorem 6). The labeling of vertices is in such a way that the first $n$ qubits are input and the last $n$ qubits are output.
(ii) The dummy qubits position, set $D$, is set to be the position of all the qubits that are required to be Pauli $Z$ measured for performing the break operators.
(iii) A sequence of nonoutput measurement angles $\phi = (\phi_i)_{1\leqslant i\leqslant(m-n)}$ with $\phi_i \in A$, where $\phi_i = \frac{\pi}{2}$ for all $i \in D$ and also for all the qubits that are required to be Pauli $Y$ measured to perform the bridge operators.
(iv) The rest of the resources are the same as Protocol 3.
Follow the steps of Protocol 3 where $G$ is replaced with $\tilde{K}_N$.

*Theorem 7.* Protocol 4 is blind, while leaking at most $n$ and $N$.

*Proof.* As Bob entangles according to $\tilde{\mathcal{K}}_N$, clearly the parameter $N$ is leaked. Additionally, in the case of quantum output, Bob must be instructed how many qubits to return to Alice and hence knows $n$. However, fixing these parameters, due to Theorem 2, all the measurement angles including the measurements for the bridge operators are blind to Bob. Similarly, from Theorem 4 we have blindness for the measurement corresponding to the break operators. Together these guarantee the blindness of the operations required to prepare a brickwork state from $\tilde{\mathcal{K}}_N$. Finally, Theorem 2 proved the blindness of the remaining measurements performed on the prepared brickwork state. ■

## VI. VERIFICATION

This section deals with another property of the hiding protocol called verification. This property requires that Alice can verify with high probability whether Bob has followed the instructions of the protocol and hence if the quantum or classical output state is indeed in the correct form or whether there has been a deviation and she should therefore reject the output state. The main idea is to exploit blindness so that Alice can expand the protocol to include *trap qubits* where Alice knows in advance the classical outcome of these specific measurements (i.e., the correct message from Bob for these measurements), where the blindness ensures that the position of these traps remains hidden from Bob. At the

end Alice will accept the quantum or classical output only if Bob has produced all of the *expected* outcomes for these trap qubits measurements. The subtlety in verification is to prove that the accepted quantum or classical output is indeed correct.

It is essential that Alice keeps the position of these trap qubits unknown to Bob so that he cannot attempt to interfere with the actual computation of $U$ while keeping the trap qubits untouched. We will present a protocol where every qubit of the underlying graph could potentially be an isolated (unentangled) trap qubit in an unknown state $|+_\theta\rangle$ for $\theta \in A$. In order to do so, it is enough to prepare all the neighboring vertices of the trap qubit as dummy qubits; hence these dummy qubits together with the trap qubits remain disentangled from the rest of the graph during the preparation stage. Building on this simple construction, by adding more traps and adding error detection elements, we will present a final protocol in which the probability of not detecting an incorrect outcome is exponentially small.

In order to first demonstrate the main idea of this method of verification, we ignore the universality property and only later will we present a concrete universal blind quantum computing protocol with the verification property. Hence, to obtain a generic hiding protocol with a random unknown trap it is sufficient to use Protocol 3, where Alice chooses a random position $t$ to be an isolated trap qubit (Protocol 5).

Protocol 5. Generic hiding quantum computation for unitary with dummy, trap, and quantum input and output.

(1) Alice's resources
   (i) Graph $G$ over $m$ vertices and a random position $t$ among the vertices of $G$.
   (ii) The rest of the resources are the same as Protocol 3, where $\phi_i = 0$ for $i = t$ and $i \in D$, where $D$ contains the set of all neighbors of position $t$ in the original graph to create an isolated trap qubit at position $t$.
(2) Follow the steps of Protocol 3.
(3) Accept or reject
   (i) After obtaining all the output qubits from Bob, if the trap qubit $t$ is an output qubit, Alice measures it with angle $\delta_t = \theta_t + r_t\pi$ to obtain $b_t$.
   (ii) Alice accepts if $b_t = r_t$.

Theorem 4 directly implies that Protocol 5 is blind and the position of the trap qubits $t$ remains unknown to Bob. Recall that at each stage $i$ only qubit $i$ is measured. We present some intermediate definitions before formalizing the definition of verification. All the protocols presented so far describe the expected behavior of Alice and Bob in a hiding protocol. Since we are concerned with the secrecy of Alice's resources we can assume that Alice always follows the steps of the protocol. In fact, after the initial step when Alice draws all the random variables $\theta_i$ and $r_i$ her behavior, for a fixed run of the protocol, is deterministic. This means that at each step the next move of Alice is determined completely by the past; however; a malicious Bob might deviate in any way he desires. We will define a run of protocol to be honest (Bob has behaved as expected) or correct (the output is correct despite Bob's deviations) based on the outcome of all measurements and the quantum output state if it exists.

Recall that in a generic hiding protocol with quantum input and output the messages sent by Bob to Alice depend on a collection of outcome measurements, $s_i \in \{0,1\}$. In fact, Bob will send the outcome value $b_i$ and then Alice, depending on $r_i$, will reset them to their corrected values $s_i$. In what follows we will deal with the corrected outcome measurement, that is, $s_i$. Similarly, at the end of the protocol Bob will send Alice some quantum output state in the output Hilbert space $\mathcal{H}_O$ that needs to be corrected depending on all the measurements outcomes. In what follows we consider the corrected quantum output state $\rho$. Note that the values of $s_i$ and $\rho$ depend on Alice's specific random choices and also Bob's general strategy of deviation. We treat this information as a single density operator to deal uniformly with both classical and quantum output. Finally, in order to consider the most general deviation that Bob can perform during a run of protocol we consider a collection of unitary operators, each acting at a stage of the protocol on the private qubits of Bob and all the other qubits and classical bits sent by Alice to Bob.

*Definition 8.* Consider a particular run of a generic hiding protocol, where all the following parameters are fixed: Alice's angles of measurements $\phi = (\phi_i)_{1\leqslant i\leqslant(m-n)}$; Alice's random variables $x = (x_i)_{1\leqslant i\leqslant n}$, $r = (r_i)_{1\leqslant i\leqslant(m-n)}$, $\theta = (\theta_i)_{1\leqslant i\leqslant m}$, and $d = (d_i)_{i\in D}$; Alice's input state $|I\rangle$; the number of Bob's private qubits $B$; and Bob's deviation unitaries at each stage of the protocol $\mathcal{U} = \{U_i\}_{0\leqslant i\leqslant m+1}$ acting on all quantum and classical messages. We define the outcome density operator (of all classical and quantum messages sent by Bob to Alice) as follows:

$$\mathcal{B}_j(\nu) = \sum_{\vec{s}\in\{0,1\}^{|O^c|}} p_{\nu,j}(\vec{s})|\vec{s}\rangle\langle\vec{s}| \otimes \rho_{\nu,j}^{\vec{s}},$$

where $\nu$ collectively denotes Alice's choice of variables $t$, $x$, $r$, $\theta$, and $d$; $j$ ranges over Bob's choices $B$ and $\mathcal{U}$; $\vec{s}$ ranges over all possible values of the corrected values $\{s_i\}$ of the measurement outcomes $\{b_i\}$ sent by Bob to Alice; and $\rho_{\nu,j}^{\vec{s}}$ is the reduced density operator for the nonmeasured qubits with the corresponding correction operators for the measurement outcomes $\vec{s}$ has been applied. We call the outcome density operator $\mathcal{B}_0(\nu)$, obtained from a run of the protocol where all $U_i$ are set to be the identity operator, the exact outcome density operator. This is the outcome density operator obtained from a run where Bob exactly follows the step of the protocol.

Note that if we were dealing only with a deterministic pattern over a connected graph state then the outcome density operator could have been simplified to a fixed pure state of the output qubits, independent of the measurement outcomes. Moreover, in such a scenario the probability of each branch of the computation would have been the same. However, the

above definition aims to capture any general deviation by Bob that could affect the determinism and probability of the branches. Also, since we will have dummy and trap qubits, then not all the possible branches will be equally probable. The outcome density operator, depending on all the random choices of Alice and Bob, can be classified as follows below. Although not all mentioned categories will be used in the remainder of the paper, we give them here for completeness and to highlight the subtle differences between possible outcomes.

*Definition 9*. We say the outcome density operator $\mathcal{B}_j(\nu)$ is honest if it is indistinguishable from the exact outcome density operator

$$\|\mathcal{B}_j(\nu) - \mathcal{B}_0(\nu)\|_{\mathrm{tr}} = 0,$$

where $\|\cdot\|_{\mathrm{tr}}$ denotes the trace norm. It is called correct if the quantum output state and the trap outcome measurement is indistinguishable from the corresponding value of the exact outcome density operator:

$$\|\mathrm{Tr}_{i\notin O, i\neq t}[B_j(\nu)] - \mathrm{Tr}_{i\notin O, i\neq t}[B_0(\nu)]\|_{\mathrm{tr}} = 0.$$

It is called lucky if $b_t = r_t$ and finally it is called incorrect if it is lucky but the quantum output state $\mathrm{Tr}_{i\notin\{O\setminus\{t\}\}}[B_j(\nu)]$ is orthogonal to the corresponding subsystem of the exact outcome density operator. Note that for the classical output scenario, any bit flip implies orthogonality.

Alice should not care if Bob's deviation leads to a correct outcome density operator, as the final quantum or classical output is in the correct state. Therefore, in the definition of a verifiable blind quantum computation we aim to bound the probability of Alice being fooled, i.e., the probability of Alice accepting an incorrect outcome density operator. Any outcome density operator either results in $s_t \neq r_t$ or is contained within the subspace of correct and incorrect outcome states. Hence, intuitively, a protocol is defined to be verifiable if the corresponding outcome state is *far from* any incorrect outcome states. Following the approach of [52], we first define the notion of correctness. Recall that for simplicity we have assumed that the computation is deterministic and the input is in a pure state and hence the ideal output will necessarily be a pure state. This restriction to pure states mirrors the approach of [52].

*Definition 10*. Let $P^{\nu}_{\mathrm{incorrect}}$ be the projection onto the subspace of all the possible incorrect outcome density operators for the fixed choice of Alice's random variables $\nu$. It will be convenient to divide $\nu$ into two subsets depending on whether the secret variables correspond to the trap setting or the remainder of the computation. Thus we define $\nu_T = \{t, r_t, \theta_t\}$ and $\nu_C = \nu/\nu_T$. When the output state is a pure state, $P^{\nu}_{\mathrm{incorrect}}$ is given by

$$(\mathbb{I} - |\Psi_{\mathrm{ideal}}\rangle\langle\Psi_{\mathrm{ideal}}|) \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right|,$$

where $|\Psi_{\mathrm{ideal}}\rangle\langle\Psi_{\mathrm{ideal}}| = \mathrm{Tr}_{i\notin\{O\setminus\{t\}\}}[B_0(\nu)]$ and where $|\eta_t^{\nu_T}\rangle = |+_{\theta_t}\rangle$ when $t \in O$ and $|\eta_t^{\nu_T}\rangle = |r_t\rangle$ otherwise. Let $p(\nu)$ be the probability of Alice choosing random variables parametrized by $\nu$, that is, the probability of choosing a particular vertex, among all possible vertices of the graph, to be the trap position (denoted by a random variable $t$) and the probability of choosing random variables $r$, $x$, $\theta$, and $d$ (as defined in Definition 8). Given $0 \leqslant \epsilon < 1$, we define a protocol to be $\epsilon$ verifiable if for any choice of Bob's strategy (defined as in Definition 8 and denoted by the index $j$) the probability of Alice accepting an incorrect outcome density operator is bounded by $\epsilon$:

$$\mathrm{Tr}\left(\sum_{\nu} p(\nu) P^{\nu}_{\mathrm{incorrect}}\ B_j(\nu)\right) \leqslant \epsilon.$$

Recall that $B_0(\nu)$ is the output density operator of an honest run after the corrections have been performed. Hence, in the above definition $|\Psi_{\mathrm{ideal}}\rangle$ is independent of $\nu$, since for an honest run of the protocol the output state is independent of Alice's secret parameters, via the correctness theorem.

*Theorem 8*. Protocol 5 is $1 - \frac{1}{2m}$ verifiable in general, and in the special case of purely classical output the protocol is also $1 - \frac{1}{m}$ verifiable, where $m$ is the total number of qubits in the protocol.

*Proof*. At the beginning of the protocol, Alice chooses the independent and uniform random variables for $\nu$. Next Alice prepares the input qubits in the form

$$|e^{\nu}\rangle = X_1^{x_1} Z_1(\theta_1) \otimes \cdots \otimes X_n^{x_l} Z_n(\theta_l)|I\rangle$$

and positions them among the first $n$ qubits. Recall that $n > |I|$ and hence the trap qubit might be among this set of qubits. She then prepares the remaining qubits in the form (where $D$ is the index of the dummy qubits)

$$|d_i\rangle \forall i \in D,$$
$$\prod_{j\in N_G(i)\cap D} Z^{d_j}|+_{\theta_i}\rangle = \left|+_{\theta_i + \sum_{j\in N_G(i)\cap D} d_j \pi}\right\rangle \forall i \notin D$$

and sends all $m$ qubits in the order of the labeling of the vertices of the graph. We represent the whole $m$ qubit state as $|M^{\nu}\rangle$. We can treat all the measurement angles $\delta_i$ as orthogonal quantum states $|\delta_i\rangle$. For a fixed choice of Alice's random variables ($\nu$) and Bob's strategy ($j$), Bob's output from the computation can be written in the form of the output of a circuit computation as depicted in Fig. 6. Note that this is the state of the system before the relevant corrections for Alice's secret key have been applied to yield the outcome density operator $B_j(\nu)$.

While in the actual protocol, at step $i$, Alice computes $\delta_i$ as a function of $s_{<i}$, which in turn is calculated from $b_{<i}$ and $r_{<i}$, we can rewrite the circuit from Fig. 6 in such a way that the values $\delta_i$ are part of the initial state, without affecting causality as they do not interact with anything until after the corresponding $b_i$ has been generated. This intuition is made rigorous in Eq. (1) via the inclusion of projections to ensure consistency. This will allow us to reorder all the operators $U_i$ to the end to obtain the new circuit shown in Fig. 7. Note that Fig. 7 is not an actual run of the protocol; it is a mathematical equivalent of Fig. 6 where the values of $b_i$ have been fixed to permit us to commute the operators as depicted. However, in the following proof we have considered any general deviation performed by Bob, that is to say, we consider any arbitrary $U_i$ operators.

In the rest of this proof we will use $t$ to represent both the random variable and also the position of the trap qubit. We define by $\Omega = U'_{m-n} U'_{m-n-1} \cdots U'_1$ the overall action of Bob's deviation and by $\mathcal{P} = [\bigotimes_{1\leqslant i\leqslant m-n} H_i Z_i(\delta_i)]E_G$ the action of the exact protocol prior to measurement. Here and in Fig. 7 we have taken $U'_i = \mathcal{P}_i U_i \mathcal{P}_i^{\dagger}$, where $\mathcal{P}_i = \bigotimes_{i+1\leqslant j\leqslant m-n} H_j Z_j(\delta_j)$. Further we define by

$$|\Psi^{\nu,b}\rangle = \bigotimes_{1\leqslant i\leqslant m} |M^{\nu}\rangle \bigotimes_{1\leqslant j\leqslant m-n} \left|\delta_j^b\right\rangle$$

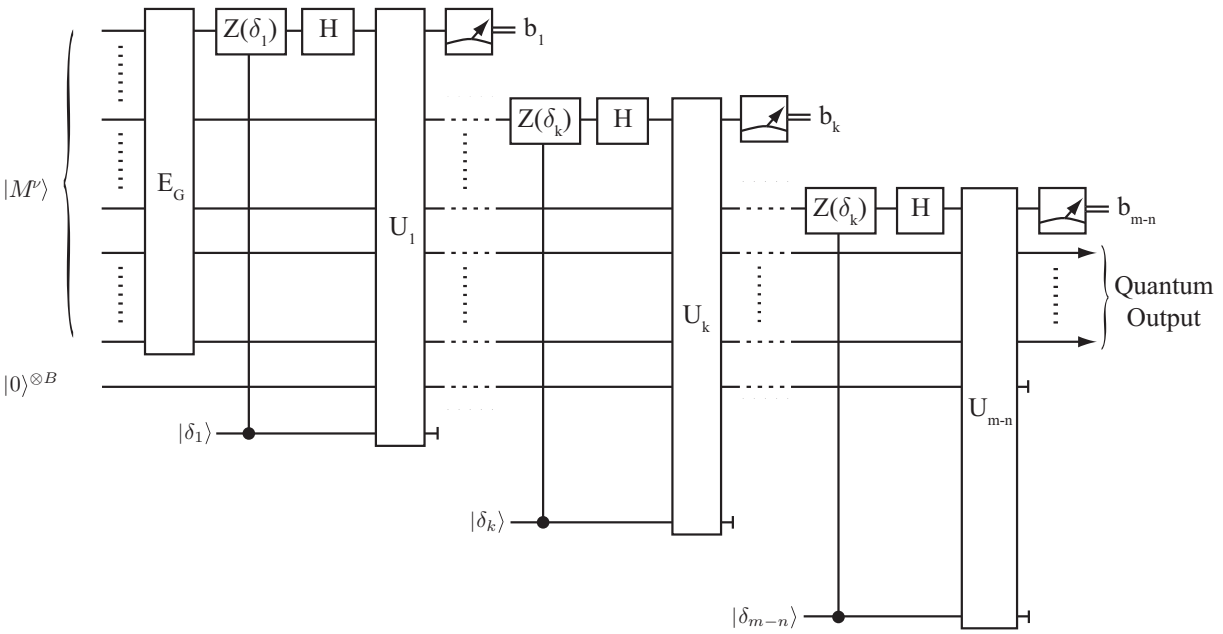


FIG. 6. Run of protocol together with Bob's deviation represented as $U_i$ operators. The entangling operator $E_G$ is the collection of all the required CTRL-$Z$ operators corresponding to the graph edges. Note that in Definition 8 we also considered an operator $U_0$ representing Bob's initial deviation. In the figure, for simplicity, we have commuted $U_0$ and combined it with $U_1$. Trivially, if all the $U_i$ operators are set to be identity the above circuit converges to the exact run of the protocol, where a measurement in the basis $|\pm_{\delta_i}\rangle$ is implemented using the controlled $Z$ rotation followed by a Hadamard gate and finally a Pauli $Z$ basis (computation basis) measurement on the corresponding qubits.

the joint state of the initial (input, dummy, and prepared) qubits sent by Alice to Bob and the classical angles $\delta_i^b$, where $b$ represents a possible branch of the computation as parametrized by the measurement results $\{b_i\}$ sent by Bob to Alice. Finally, in line with Definition 10, we define $C_{\nu_C,b}$ to be the Pauli operator that maps the final quantum output state to the correct one depending on the random variable $\nu_C$ and computation branch $b$. Hence we have

$$B_j(\nu) = \mathrm{Tr}_B\Bigg( \sum_b |b + c_r\rangle\langle b| C_{\nu_C,b}\Omega\mathcal{P}[(\otimes^B|0\rangle\langle 0|) \otimes |\Psi^{\nu,b}\rangle \times \langle\Psi^{\nu,b}|]\mathcal{P}^\dagger\Omega^\dagger C^\dagger_{\nu_C,b}|b\rangle\langle b + c_r|\Bigg), \tag{1}$$

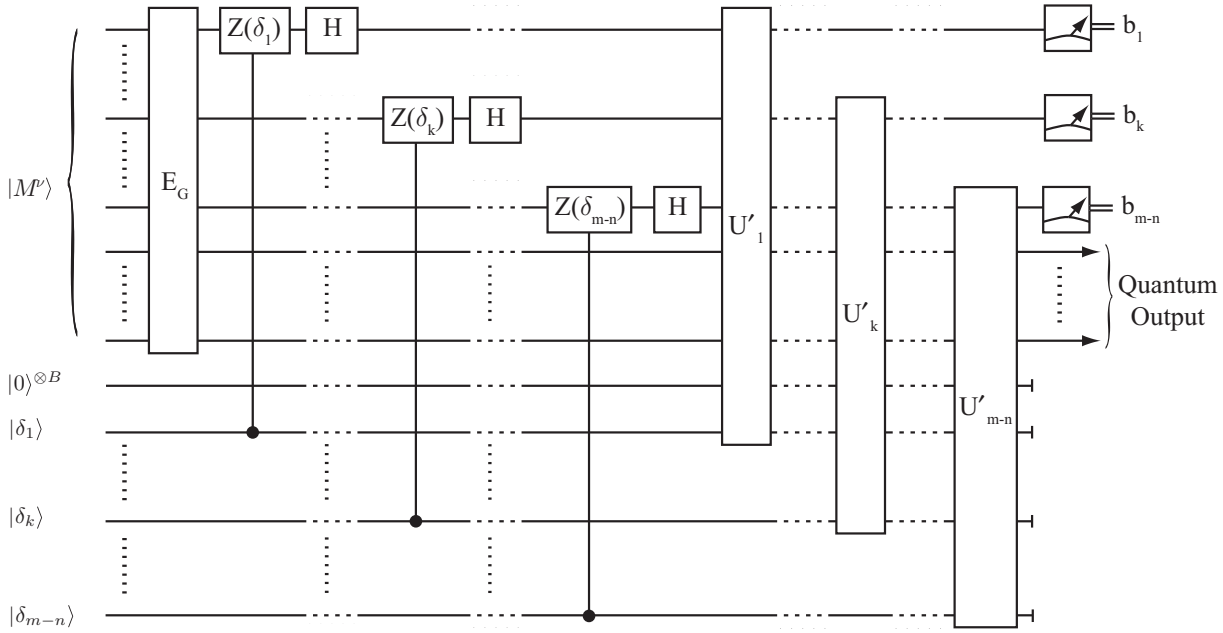


FIG. 7. The fact that any $U_j$ in Fig. 6 is independent of all $\delta_{i>j}$ allows us to reposition the deviation to the end of the circuit as shown above. Hence we can rewrite Bob's deviation as $U'_i = \mathcal{P}_i U_i \mathcal{P}_i^\dagger$, where $\mathcal{P}_i = \bigotimes_{i+1\leqslant j\leqslant m-n} H_j Z_j(\delta_j)$.

where $(c_r)_i = r_i$ for all $i \neq t$, $(c_r)_t = 0$, and the subscript $B$ denotes that the partial trace is taken over Bob's private register. Here $c_r$ is used to compactly deal with the fact that in the protocol all measured qubits are decrypted by applying an XOR operation with $r$, except for the trap qubit, which remains uncorrected. Note that, in the above, the operator $\langle b| \cdots |b\rangle$ acts upon the subspace of all measured qubits and $|b + c_r\rangle \cdots \langle b + c_r|$ store the corrected outcome of the measurement. The above equation includes the dependence of $\delta_i$ on previous measurement results via the inclusion of the parameter $b$ in the initial state $|\Psi^{\nu,b}\rangle$. The projectors $|b + c_r\rangle\langle b|$ and $|b\rangle\langle b + c_r|$ then enforce consistency, by ensuring that measurement results match the values used in the computation of subsequent $\delta_i$.

We take $P_\perp$ to be the projection onto the subspace of incorrect states for the nontrap qubits, after Alice's final corrections have been applied to any quantum output. Hence

$$P^{\nu}_{\text{incorrect}} = P_\perp \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right|,$$

where $|\eta_t^{\nu_T}\rangle = |r_t\rangle_t$ for $1 \leqslant t \leqslant m - n$ and $|\eta_t^{\nu_T}\rangle = |+_{\theta_t}\rangle_t$ for $m - n + 1 \leqslant t \leqslant m$. Here we use the subscript on the ket to identify the relevant qubit. Thus we have

$$\mathrm{Tr}\big[P^{\nu}_{\text{incorrect}} B_j(\nu)\big] = \mathrm{Tr}\left[ P_\perp \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \left( \sum_b |b + c_r\rangle\langle b| C_{\nu_C,b} \Omega \mathcal{P}[(\otimes^B |0\rangle\langle 0|) \otimes |\Psi^{\nu,b}\rangle\langle\Psi^{\nu,b}|] \mathcal{P}^\dagger \Omega^\dagger C^\dagger_{\nu_C,b} |b\rangle\langle b + c_r| \right) \right].$$

As Bob's private register is traced out, the net result of $\Omega$ is to apply a completely positive trace preserving map of the other qubits. Taking the Kraus operators associated with this operator to be $\{\chi_k\}$, with $\sum_k \chi_k \chi_k^\dagger = \mathbb{I}$, we have

$$\mathrm{Tr}\big[P^{\nu}_{\text{incorrect}}\ B_j(\nu)\big] = \sum_k \sum_b \mathrm{Tr}\big[ \big(P_\perp \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \big) |b + c_r\rangle\langle b| C_{\nu_C,b} \chi_k \mathcal{P} |\Psi^{\nu,b}\rangle\langle\Psi^{\nu,b}| \mathcal{P}^\dagger \chi_k{}^\dagger C^\dagger_{\nu_C,b} |b\rangle\langle b + c_r| \big].$$

Since any Kraus operator can be written as a linear combination of Pauli operators with complex coefficients, we have $\chi_k = \sum_i \alpha_{ki} \sigma_i$, where $\sum_k \sum_i \alpha_{ki} \alpha^*_{ki} = 1$ and $\sigma_i$ is a Pauli operator acting on the joint quantum state of the system. Therefore, the above equation can be written as

$$\begin{aligned}
\mathrm{Tr}\big[P^{\nu}_{\text{incorrect}} B_j(\nu)\big] &= \sum_k \sum_b \mathrm{Tr}\left[ \big(P_\perp \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \big) |b + c_r\rangle\langle b| C_{\nu_C,b} \left( \sum_{i,j} \alpha_{ki} \alpha^*_{kj} \sigma_i \mathcal{P} |\Psi^{\nu}\rangle\langle\Psi^{\nu}| \mathcal{P}^\dagger \sigma_j \right) C^\dagger_{\nu_C,b} |b\rangle\langle b + c_r| \right] \\
&= \sum_k \sum_b \mathrm{Tr}\left( \sum_{i,j} \alpha_{ki} \alpha^*_{kj} \big(P_\perp \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \big) |b + c_r\rangle\langle b| C_{\nu_C,b} \sigma_i \mathcal{P} |\Psi^{\nu,b}\rangle\langle\Psi^{\nu,b}| \mathcal{P}^\dagger \sigma_j C^\dagger_{\nu_C,b} |b\rangle\langle b + c_r| \right).
\end{aligned}$$

In order to determine which $\sigma_i$ terms have a nonzero contribution in the above sum after the projection operator is taken into account, it will be necessary to look at the structure of each such Pauli operator. To this end, we will denote by $\sigma_{i|\gamma}$ the action of $\sigma_i$ on qubit $\gamma$ and hence $\sigma_{i|\gamma} \in \{I, X, Y, Z\}$. For simplicity, we assume each $\delta_i$ is encoded across three qubits (since there are only eight possible angles). Thus, we have $1 \leqslant \gamma \leqslant m + 3(m - n)$, where $1 \leqslant \gamma \leqslant m$ identifies qubits received from Alice and the remaining $\gamma$ values identify the qubits containing $\delta_i$. Without loss of generality, we can assume that the qubits representing the values of $\delta$ remain unchanged by Bob's deviation and hence we can take $\sigma_{i|\gamma} \in \{I, Z\}$ for all $m < \gamma$.

The probability of Alice accepting an incorrect outcome density operator is given by

$$p_{\text{incorrect}} = \mathrm{Tr}\left( \sum_\nu p(\nu) P^{\nu}_{\text{incorrect}} B_j(\nu) \right).$$

This can be calculated via the expression for $\mathrm{Tr}[P^{\nu}_{\text{incorrect}}\ B_j(\nu)]$ obtained earlier,

$$\begin{aligned}
p_{\text{incorrect}} &= \sum_\nu p(\nu) \mathrm{Tr}\big[P^{\nu}_{\text{incorrect}}\ B_j(\nu)\big] \\
&= \sum_{k,b} \mathrm{Tr}\left( \sum_\nu p(\nu) \sum_{i,j} \alpha_{ki} \alpha^*_{kj} \big(P_\perp \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \big) |b + c_r\rangle\langle b| C_{\nu_C,b} \sigma_i \mathcal{P} |\Psi^{\nu,b}\rangle\langle\Psi^{\nu,b}| \mathcal{P}^\dagger \sigma_j C^\dagger_{\nu_C,b} |b\rangle\langle b + c_r| \right) \\
&= \sum_{b,i,j,k} \mathrm{Tr}\left( \sum_\nu p(\nu) \alpha_{ki} \alpha^*_{kj} \big(P_\perp \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \big) |b + c_r\rangle\langle b| C_{\nu_C,b} \sigma_i \mathcal{P} |\Psi^{\nu,b}\rangle\langle\Psi^{\nu,b}| \mathcal{P}^\dagger \sigma_j C^\dagger_{\nu_C,b} |b\rangle\langle b + c_r| \right).
\end{aligned}$$

By noting that $|b_j + c_{r_j}\rangle$ commutes with $|\Psi^{\nu,b}\rangle\langle\Psi^{\nu,b}|$ for all $j \neq t$, the above expression can be rewritten as

$$p_{\text{incorrect}} = \sum_{b,i,j,k} \mathrm{Tr}\left( \sum_\nu p(\nu) \alpha_{ki} \alpha^*_{kj} \big(P_\perp \otimes \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \big) |b_t\rangle\langle b| C_{\nu_C,b} \sigma_i \mathcal{P} |\Psi^{\nu,b}\rangle\langle\Psi^{\nu,b}| \mathcal{P}^\dagger \sigma_j C^\dagger_{\nu_C,b} |b\rangle\langle b_t| \right).$$

In order to obtain an upper bound for the above expression we make use of sets of indices $\gamma$ of qubits such that the action of $\sigma_i$ at that position $\sigma_{i|\gamma}$ is a particular Pauli operator, which we define as follows:

$$A_i = \{\gamma : \sigma_{i|\gamma} = I,\ 1 \leqslant \gamma \leqslant m\}, \quad B_i = \{\gamma : \sigma_{i|\gamma} = X,\ 1 \leqslant \gamma \leqslant m\},$$
$$C_i = \{\gamma : \sigma_{i|\gamma} = Y,\ 1 \leqslant \gamma \leqslant m\}, \quad D_i = \{\gamma : \sigma_{i|\gamma} = Z,\ 1 \leqslant \gamma \leqslant m\}.$$

Note that in the above we restrict attention to the set of qubits originally sent from Alice to Bob (which is why $1 \leqslant \gamma \leqslant m$) and disregard the action on Bob's private qubits. Additionally, we will make use of a superscript $O$ to denote subsets of the above sets subject to the constraint that $\gamma$ is an output qubit ($m - n < \gamma$). Thus, for example, $D_i^O = \{\gamma : \sigma_{i|\gamma} = Z,\ m - n + 1 \leqslant \gamma \leqslant m\}$. We note that only $\sigma_i$ and $\sigma_j$ operators for which $\mathrm{Tr}(P_\perp \sigma_i \mathcal{P}|\Psi^{v,b}\rangle\langle\Psi^{v,b}|\mathcal{P}^\dagger \sigma_j) \neq 0$ contribute to $p_{\text{incorrect}}$. With the above definitions in place, we can express succinctly a necessary condition for this to hold as $|B_i| + |C_i| + |D_i^O| \geqslant 1$ (with $i \in E_i$) and $|B_j| + |C_j| + |D_j^O| \geqslant 1$ (with $j \in E_j$). That is to say, one or both of the following has happened: $\sigma_i$ ($\sigma_j$) has produced an incorrect outcome for one or more of the measurement results and hence $|B_i \setminus B_i^O| + |C_i \setminus C_i^O| \geqslant 1$ ($|B_j \setminus B_j^O| + |C_j \setminus C_j^O| \geqslant 1$) or $\sigma_i$ ($\sigma_j$) acts nontrivially on the quantum output and hence $|B_i^O| + |C_i^O| + |D_i^O| \geqslant 1$ ($|B_j^O| + |C_j^O| + |D_j^O| \geqslant 1$). Using this set notion and by taking the trace over the subspace of the measurement results except for the trap qubit we obtain

$$p_{\text{incorrect}} = \sum_{k,b} \sum_{i\in E_i} \sum_{j\in E_j} \mathrm{Tr}\left( \sum_v p(v)\alpha_{ki}\alpha_{kj}^* \left(P_\perp \otimes \left|\eta_t^{v_T}\right\rangle\!\left\langle\eta_t^{v_T}\right|\right) |b_t\rangle\langle b| C_{v_C,b}\sigma_i \mathcal{P}|\Psi^{v,b}\rangle\langle\Psi^{v,b}|\mathcal{P}^\dagger \sigma_j C_{v_C,b}^\dagger |b\rangle\langle b_t| \right),$$

where we take $|b_t\rangle$ to have have unit dimension if $t \in O$. The reason for doing this is to allow a uniform treatment of trap qubits independent of whether or not the trap occurs on a measured qubit. Taking $b' = \{b_i\}_{i\neq t}$, a substring of $b$ that excludes the value for the trap measurement, the above equation can be written as

$$p_{\text{incorrect}} = \sum_{k,b} \sum_{i\in E_i} \sum_{j\in E_j} \mathrm{Tr}\left( \sum_v p(v)\alpha_{ki}\alpha_{kj}^* \left[P_\perp \otimes \left(\left|\eta_t^{v_T}\right\rangle\!\left\langle\eta_t^{v_T}\right| b_t\right\rangle\langle b_t|\right)\right] \langle b'| C_{v_C,b}\sigma_i \mathcal{P}|\Psi^{v,b}\rangle\langle\Psi^{v,b}|\mathcal{P}^\dagger \sigma_j C_{v_C,b}^\dagger |b'\rangle \right).$$

Note in the above that if the trap is measured we have $\langle\eta_t^{v_T}|b_t\rangle = \delta_{\eta_t^{v_T},b_t}$; otherwise $|b_t\rangle\langle b_t| = 1$. Hence we have

$$\begin{aligned}
p_{\text{incorrect}} &= \sum_{k,b'} \sum_{i\in E_i} \sum_{j\in E_j} \mathrm{Tr}\left( \sum_v p(v)\alpha_{ki}\alpha_{kj}^* \left(P_\perp \otimes \left|\eta_t^{v_T}\right\rangle\!\left\langle\eta_t^{v_T}\right|\right) |b'\rangle\langle b'| C_{v_C,b'}\sigma_i \mathcal{P}|\Psi^{v,b'}\rangle\langle\Psi^{v,b'}|\mathcal{P}^\dagger \sigma_j C_{v_C,b'}^\dagger \right) \\
&= \sum_{k,b'} \sum_v p(v)\mathrm{Tr}\left[ \left(P_\perp \otimes \left|\eta_t^{v_T}\right\rangle\!\left\langle\eta_t^{v_T}\right|\right) |b'\rangle\langle b'| C_{v_C,b'} \left(\sum_{i\in E_i} \alpha_{ki}\sigma_i\right) \mathcal{P}|\Psi^{v,b'}\rangle\langle\Psi^{v,b'}|\mathcal{P}^\dagger \left(\sum_{i\in E_i} \alpha_{ki}\sigma_i\right)^\dagger C_{v_C,b'}^\dagger \right] \\
&\leqslant \sum_{k,b'} \sum_v p(v)\mathrm{Tr}\left[ \left(\left|\eta_t^{v_T}\right\rangle\!\left\langle\eta_t^{v_T}\right| \otimes |b'\rangle\langle b'|\right) C_{v_C,b'} \left(\sum_{i\in E_i} \alpha_{ki}\sigma_i\right) \mathcal{P}|\Psi^{v,b'}\rangle\langle\Psi^{v,b'}|\mathcal{P}^\dagger \left(\sum_{i\in E_i} \alpha_{ki}\sigma_i\right)^\dagger C_{v_C,b'}^\dagger \right] \\
&= \sum_{k,b'} \sum_v p(v)\mathrm{Tr}\left[ \left(\left|\eta_t^{v_T}\right\rangle\!\left\langle\eta_t^{v_T}\right| \otimes |b'\rangle\langle b'|\right) \left(\sum_{i\in E_i} \alpha_{ki}\sigma_i\right) \mathcal{P}|\Psi^{v,b'}\rangle\langle\Psi^{v,b'}|\mathcal{P}^\dagger \left(\sum_{i\in E_i} \alpha_{ki}\sigma_i\right)^\dagger \right],
\end{aligned}$$

where the inequality follows from the fact that the projector $P_\perp$ acts on a positive semidefinite matrix and the last equality follows from the fact that both remaining projectors act as the identity on qubits in $O$.

Next we attempt to show that a necessary requirement for a term in the above summation over $i$ and $j$ to be nonzero is that $i = j$. As per the proof of blindness, summing over $\nu_C$ yields the maximally mixed state of the system received from Alice. Hence we have

$$\begin{aligned}
p_{\text{incorrect}} &\leqslant \sum_{k,b',\nu_T} \sum_{i\in E_i} \sum_{j\in E_j} \alpha_{ik}\alpha_{jk}^* p(\nu_T)\mathrm{Tr}\left[ \left(\left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \otimes |b'\rangle\langle b'|\right)\sigma_i \left(\left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \otimes |\delta_t\rangle\langle\delta_t| \otimes \frac{I}{\mathrm{Tr}(I)}\right)\sigma_j \right] \\
&= \sum_{k,\nu_T} \sum_{i\in E_i} \sum_{j\in E_j} \alpha_{ik}\alpha_{jk}^* p(\nu_T)\mathrm{Tr}\left[ \left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \sigma_i \left(\left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \otimes |\delta_t\rangle\langle\delta_t| \otimes \frac{I}{\mathrm{Tr}(I)}\right)\sigma_j \right] \\
&= \sum_{k,\nu_T} \sum_{i\in E_i} \sum_{j\in E_j} \alpha_{ik}\alpha_{jk}^* p(\nu_T)\mathrm{Tr}\left[ \left\langle\eta_t^{\nu_T}\right| \sigma_i \left(\left|\eta_t^{\nu_T}\right\rangle\!\left\langle\eta_t^{\nu_T}\right| \otimes |\delta_t\rangle\langle\delta_t| \otimes \frac{I}{\mathrm{Tr}(I)}\right)\sigma_j \left|\eta_t^{\nu_T}\right\rangle \right].
\end{aligned}$$

As all Pauli matrices other than the identity are traceless, any terms in the sum that are nonzero necessarily have $\sigma_{i|\gamma} = \sigma_{j|\gamma}$ everywhere except for $\gamma = t$ and the corresponding $\delta$ register. We then consider separately the two cases corresponding

to whether the trap is located in the quantum output or not. If $t \in O$ then the $\delta$ register does not exist and using the fact that $\sum_{\theta_t,r_t} \mathrm{Tr}(\langle\eta_t^{v_T}|\sigma_i|\eta_t^{v_T}\rangle\langle\eta_t^{v_T}|\sigma_j|\eta_t^{v_T}\rangle) = 0$, unless $\sigma_{i|t} = \sigma_{j|t}$, we arrive at the conclusion that the only terms that contribute to $p_{\text{incorrect}}$ are those where $\sigma_i = \sigma_j$. If, on the other hand, $t \notin O$, then averaging over $r_t$ alone is sufficient to give $\mathrm{Tr}(\langle\eta_t^{v_T}|\sigma_i|\eta_t^{v_T}\rangle\langle\eta_t^{v_T}|\sigma_j|\eta_t^{v_T}\rangle) = 0$ and hence $\sigma_{i|t} = \sigma_{j|t}$. In this case, averaging over $\theta$ yields the $\delta_t$ register in the maximally mixed state and hence, as before, $\sigma_i$ and $\sigma_j$ must act identically on these qubits too, in order to avoid contributing zero to the value of $p_{\text{incorrect}}$. Consequently, the only terms that contribute are those for which $\sigma_i = \sigma_j$. Using this identity with our previous expression for $p_{\text{incorrect}}$, we obtain

$$\begin{aligned}
p_{\text{incorrect}} &\leqslant \sum_{k,v_T}\sum_{i\in E_i} \alpha_{ik}\alpha_{ik}^* p(v_T)\mathrm{Tr}\left[\langle\eta_t^{v_T}|\sigma_i\left(|\eta_t^{v_T}\rangle\langle\eta_t^{v_T}| \otimes |\delta_t\rangle\langle\delta_t| \otimes \frac{I}{\mathrm{Tr}(I)}\right)\sigma_i|\eta_t^{v_T}\rangle\right] \\
&= \sum_{k,v_T}\sum_{i\in E_i} |\alpha_{ik}|^2 p(v_T)\mathrm{Tr}\left(\langle\eta_t^{v_T}|\sigma_{i|t}|\eta_t^{v_T}\rangle\langle\eta_t^{v_T}|\sigma_{i|t}|\eta_t^{v_T}\rangle\right) \\
&= \sum_{k,v_T}\sum_{i\in E_i} |\alpha_{ik}|^2 p(v_T)\left(\langle\eta_t^{v_T}|\sigma_{i|t}|\eta_t^{v_T}\rangle\right)^2 \\
&= \frac{1}{16m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 \sum_{t,r_t,\theta_t}\left(\langle\eta_t^{v_T}|\sigma_{i|t}|\eta_t^{v_T}\rangle\right)^2 \\
&= \frac{1}{16m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 \left(\sum_{t\leqslant m-n,\theta_t,r_t}\left(\langle\eta_t^{v_T}|\sigma_{i|t}|\eta_t^{v_T}\rangle\right)^2 + \sum_{m-n<t,\theta_t,r_t}\left(\langle\eta_t^{v_T}|\sigma_{i|t}|\eta_t^{v_T}\rangle\right)^2\right) \\
&= \frac{1}{16m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 \left(\sum_{t\leqslant m-n,\theta_t,r_t}(\langle r_t|\sigma_{i|t}|r_t\rangle)^2 + \sum_{m-n<t,\theta_t,r_t}(\langle +_{\theta_t}|\sigma_{i|t}|+_{\theta_t}\rangle)^2\right) \\
&= \frac{1}{16m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 \left[\left(16\left|A_i \setminus A_i^O\right| + 16\left|D_i \setminus D_i^O\right|\right) + \left(8\left|B_i^O\right| + 8\left|C_i^O\right| + 16\left|A_i^O\right|\right)\right] \\
&= \frac{1}{2m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 \left(2|A_i| + 2\left|D_i \setminus D_i^O\right| + \left|B_i^O\right| + \left|C_i^O\right|\right).
\end{aligned}$$

This can be further simplified, since $|A_i| + |B_i| + |C_i| + |D_i| = m$, giving

$$\begin{aligned}
p_{\text{incorrect}} &\leqslant \frac{1}{2m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 \left[2m - 2\left(|B_i| + |C_i| + \left|D_i^O\right|\right) + \left|B_i^O\right| + \left|C_i^O\right|\right] \\
&\leqslant \frac{1}{2m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 \left(2m - |B_i| - |C_i| - 2\left|D_i^O\right|\right) \\
&\leqslant \frac{1}{2m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 (2m-1) \\
&\leqslant 1 - \frac{1}{2m}
\end{aligned}$$

for the general case. However, for the specific case of only classical output, this bound can be made tighter by performing the simplification in a different way, since $|B_i^O| = |C_i^O| = |D_i^O| = 0$, and hence

$$\begin{aligned}
p_{\text{incorrect}} &\leqslant \frac{1}{2m}\sum_{k}\sum_{i\in E_i} |\alpha_{ki}|^2 \left(2|A_i| + 2\left|D_i \setminus D_i^O\right| + \left|B_i^O\right| - \left|C_i^O\right|\right) \\
&= \frac{1}{m}\sum_{k}\sum_{i:|B_i|+|C_i|\geqslant 1} |\alpha_{ki}|^2 (|A_i| + |D_i|) \\
&= \frac{1}{m}\sum_{k}\sum_{i:|B_i|+|C_i|\geqslant 1} |\alpha_{ki}|^2 (m - |B_i| - |C_i|) \\
&\leqslant \frac{1}{m}\sum_{k}\sum_{i:|B_i|+|C_i|\geqslant 1} |\alpha_{ki}|^2 (m-1) \\
&\leqslant 1 - \frac{1}{m}.
\end{aligned}$$

■

This single trap construction will be generalized in the next section to allow for exponential suppression of the probability of accepting an incorrect outcome even in the case of quantum output. We finish this section by showing that even this simple construction can be used to verify universal quantum computation, using the cylinder brickwork state presented in Sec. V.

It is easy to verify that if Alice chooses a random row of a cylinder graph $\mathcal{G}^C_{n\times m}$ (Fig. 4) and prepares all the qubits of that row in the states $|z_i\rangle$ where $z_i \in_R \{0,1\}$ and the rest of nodes in the state $|+\rangle$, then after entangling according to the cylinder brickwork graph the obtained state is $\mathcal{G}_{(n-1)\times m}\bigotimes_{i=1}^m |z_i\rangle$. By choosing a random trap location and a dummy tape that contains its neighborhood we can construct a single-trap verifiable universal blind quantum computing protocol, given by Protocol 6 and illustrated in Fig. 8.

*Corollary 1.* Protocol 6 is universal, blind while leaking at most $m$ and $n$, and $1 - \frac{1}{2m}$ verifiable in general and $1 - \frac{1}{m}$ verifiable in the case of classical output.

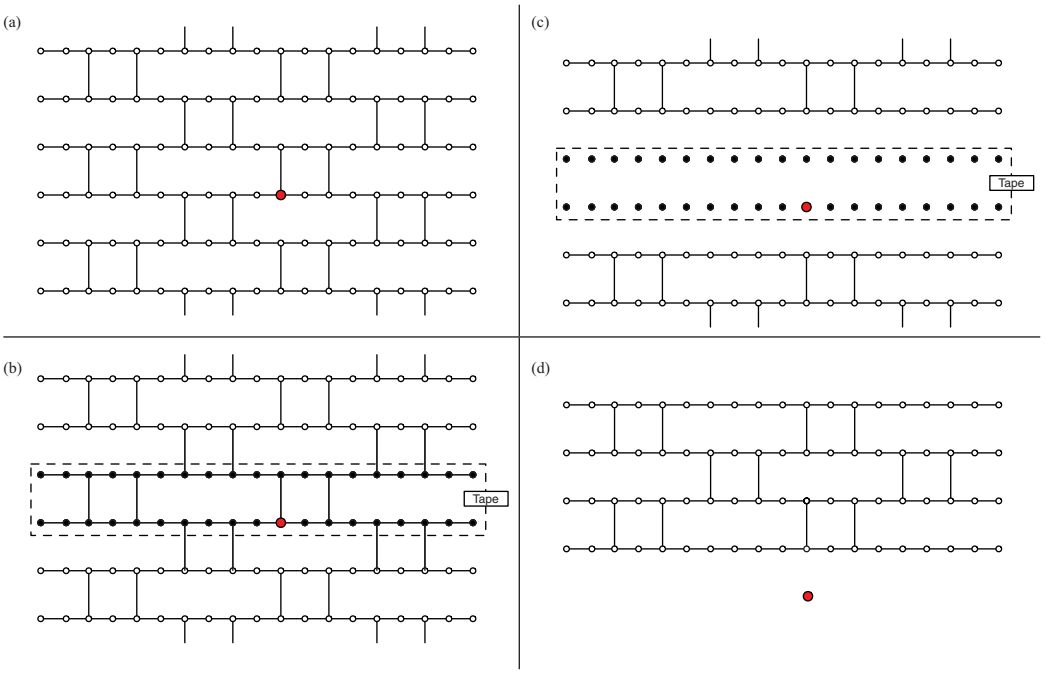


FIG. 8. Single-trap verifiable universal blind quantum computation using the cylinder brickwork state. (a) A random qubit is chosen to be the trap qubit, the (red) filled node. (b) All other vertices in the tape containing the trap qubit, the solid (black) nodes, are set to be dummy qubits. (c) This results in an isolated trap qubit in the state $|+_\theta\rangle$ together with many dummy qubits after entanglement operations are applied by the server. (d) The net result, after discarding the dummy qubits, is a disentangled trap qubit in a product state with a brickwork state.

*Proof.* Since the dummy qubits are prepared in eigenstates of the Pauli $Z$ operator, they remain in a product state with the rest of the system after the entangling operations are applied by Bob. The result, as depicted in Fig. 8, is that the trap qubit also remains in a product state and a brickwork state is prepared in the subsystem excluding $T$. The universality property then follows directly from the universality of the brickwork state from Theorem 5. As Protocol 6 is a special case of Protocol 3, the blindness property follows directly from Theorem 4 and therefore the angles of measurement $\phi_i$ remain secret from Bob. Moreover, the universality of the cylinder brickwork state guarantees that Bob's knowledge of $\mathcal{G}^C_{n\times m}$ does not reveal anything about the underlying computation except $n$ and $m$. As Protocol 6 is also a special case of Protocol 5, the verifiability property follows directly from Theorem 8. ■

## VII. PROBABILITY AMPLIFICATION FOR UNIVERSAL VERIFIABLE BLIND QUANTUM COMPUTATION

In the preceding section we presented a very simple verifiable protocol where the probability of Bob succeeding in making Alice accept an incorrect outcome density operator was strictly less than 1. Building upon that simple construction, by adding more traps and making the computation fault tolerant, we can make the probability of Alice accepting an incorrect outcome density operator as

Protocol 6. Single-trap verifiable universal blind quantum computation.

(1) Alice's resources
  (i) A graph $G = \mathcal{G}^C_{n\times m}$ and a randomly chosen vertex $t$ of $G$.
  (ii) The rest of the resources are the same as Protocol 3, where $\phi_i = 0$ for $i = t$ and $i \in D$, where $D$ contains the set of all vertices in a tape $T$ that contains position $t$ and all of its neighbors.
(2) Follow the steps of Protocol 3.
(3) Accept or reject
  (i) After obtaining all the output qubits from Bob, if the trap qubit $t$ is an output qubit, Alice measures it with angle $\delta_t = \theta_t + r_t\pi$ to obtain $b_t$.
  (ii) Alice accepts if $b_t = r_t$.

small as required. The central idea is to design a protocol with $O(N)$ many traps in essentially random locations, where $N$ is the number of qubits in the protocol, to increase the probability of any local error being detected. The fault tolerance is added to increase the minimum weight of any operator that leads to an incorrect outcome and hence further increase the probability of detection. Here and in what follows, the weight of a Pauli operator is defined to be the number of qubits upon which it acts nontrivially. First, given such a protocol, we show how it amplifies the verification parameter. We then present the

central contribution of this paper, a universal verifiable blind quantum computing protocol that achieves the probability amplification without any such assumptions.

*Theorem 9.* Let $\mathcal{P}$ be a blind quantum computing protocol on $N$ qubits with $N_T$ isolated traps in the states $|+_{\theta_t}\rangle$ at a set of positions $T$ chosen uniformly at random. Let $N_T/N$ be a constant $c$. Take $\sigma = \{\sigma^i\}$ to be a set of Pauli operators such that for at most $d$ distinct indices $i$ we have $\sigma^i \in \{X,Y,Z\}$ and for the remaining indices $\sigma^i = I$. Assume that the underlying computation of $\mathcal{P}$ is encoded in such a way that for any such $\sigma$, if each measurement result or unmeasured qubit $i$ is modified by applying $\sigma^i$, then either the computation is correct or an error is detected when the output is decoded. Then the protocol is $(1-\frac{c}{2})^d$ verifiable in general and $(1-c)^d$ verifiable in the case of purely classical output.

*Proof.* In order to exploit Theorem 8, we notionally partition the qubits into independent sets with one single trap qubit in each set. These partitions amount to extra information about the location of the trap qubits and hence their inclusion can only serve to increase the probability of Bob convincing Alice to accept an incorrect state. Thus the bound we obtain with this additional information is still an upper bound on the probability of Alice accepting an incorrect output when these partitions are unknown. There are $N_T$ many such sets $S_\gamma$ with $1/c$ many qubits in each set. We adopt a similar proof strategy to that used to prove Theorem 8, taking

$$P^{v}_{\text{incorrect}} = P_\perp \bigotimes_{t\in T} \left|\eta_t^{v_T}\right\rangle\!\left\langle\eta_t^{v_T}\right|$$

as the projection onto the subspace of incorrect outcomes. As in the proof of Theorem 8, only those Pauli operators contribute to $p_{\text{incorrect}}$ where one or both of the following has happened: $\sigma_i$ has produced an incorrect outcome for some of the measurement results $b_i$ or $\sigma_i$ acts nontrivially on the quantum output. Now due to the error-detection property of the encoding assumed in the statement of the theorem, we need to consider only those $\sigma_i$ where $|B_i| + |C_i| + |D_i^O| \geqslant d$. Following the steps of the proof of Theorem 8, we obtain

$$p_{\text{incorrect}} = \sum_v p(v)\text{Tr}\left[P^{v}_{\text{incorrect}} B_j(v)\right] \leqslant \sum_k \sum_{i:|B_i|+|C_i|+|D_i^O|\geqslant d} |\alpha_{ki}|^2 \sum_T p(T) \prod_{t\in T}\left(\sum_{\theta_t,r_t} p(\theta_t)p(r_t)\left(\left\langle\eta_t^{v_T}\right|\sigma_{i|t}\left|\eta_t^{v_T}\right\rangle\right)^2\right).$$

Here we can exploit the structure we have introduced through the sets $S_\gamma$

$$p_{\text{incorrect}} \leqslant \sum_k \sum_{i:|B_i|+|C_i|+|D_i^O|\geqslant d} |\alpha_{ki}|^2 \prod_{\gamma=1}^{N_T} \sum_{t_\gamma,\theta_{t_\gamma},r_{t_\gamma}} p(t_\gamma)p\left(\theta_{t_\gamma}\right)p\left(r_{t_\gamma}\right)\left\langle\eta^v_{t_\gamma}\right|\sigma_{i|t_\gamma}\left|\eta^v_{t_\gamma}\right\rangle^2,$$

where $t_\gamma$ is taken to be the location of the trap qubit in set $S_\gamma$. Rearranging the above and substituting in the values of $p(t_\gamma)$, $p(\theta_{t_\gamma})$, and $p(r_{t_\gamma})$, we obtain

$$p_{\text{incorrect}} \leqslant \sum_k \sum_{i:|B_i|+|C_i|+|D_i^O|\geqslant d} |\alpha_{ki}|^2 \prod_{\gamma=1}^{N_T} \sum_{t_\gamma,\theta_{t_\gamma},r_{t_\gamma}} \frac{c}{16}\left\langle\eta^v_{t_\gamma}\right|\sigma_{i|t_\gamma}\left|\eta^v_{t_\gamma}\right\rangle^2.$$

Note that within each set the position of the trap is chosen uniformly at random and so the probability of detection by that trap corresponds to the bound obtained for Theorem 8. Going through the steps of the proof of Theorem 8, we obtain

$$\begin{aligned} p_{\text{incorrect}} &\leqslant \sum_k \sum_{i:|B_i|+|C_i|+|D_i^O|\geqslant d} |\alpha_{ki}|^2 \prod_{\gamma=1}^{N_T} \frac{c}{2}\left(2|A_{i\gamma}| + 2\left|D_{i\gamma} \setminus D^O_{i\gamma}\right| + \left|B^O_{i\gamma}\right| + \left|C^O_{i\gamma}\right|\right) \\ &= \sum_k \sum_{i:|B_i|+|C_i|+|D_i^O|\geqslant d} |\alpha_{ki}|^2 \prod_{\gamma=1}^{N_T} \frac{c}{2}\left(\frac{2}{c} - 2\left|D^O_{i\gamma}\right| - |B_{i\gamma}| - |C_{i\gamma}| - \left|B_{i\gamma} \setminus B^O_{i\gamma}\right| - \left|C_{i\gamma} \setminus C^O_{i\gamma}\right|\right), \end{aligned}$$

where we use the additional $\gamma$ subscript on sets $|A_{i\gamma}|, \ldots, |D_{i\gamma}|$ to indicate subsets of the respective sets, subject to the restriction that the elements are also in $S_\gamma$. For convenience we define $w_{i\gamma} = |B_{i\gamma}| + |C_{i\gamma}| + |D^O_{i\gamma}|$ and $w_i = |B_i| + |C_i| + |D_i^O|$. Thus we obtain

$$p_{\text{incorrect}} \leqslant \sum_k \sum_{i:w_i\geqslant d} |\alpha_{ki}|^2 \prod_{\gamma=1}^{N_T} \frac{c}{2}\left(\frac{2}{c} - w_{i\gamma} - \left|B_{i\gamma} \setminus B^O_{i\gamma}\right| - \left|C_{i\gamma} \setminus C^O_{i\gamma}\right| - \left|D^O_{i\gamma}\right|\right) \leqslant \sum_k \sum_{i:w_i\geqslant d} |\alpha_{ki}|^2 \prod_{\gamma=1}^{N_T}\left(1 - \frac{cw_{i\gamma}}{2}\right).$$

We now make use of the fact that, for any positive $a$, $1 - \frac{ac}{2} \leqslant [1-(a-1)\frac{c}{2}](1-\frac{c}{2})$. As $w_{i\gamma}$ is a non-negative integer, we can recursively apply this identity to obtain

$$\begin{aligned} p_{\text{incorrect}} &\leqslant \sum_k \sum_{i:w_i\geqslant d} |\alpha_{ki}|^2 \prod_{\gamma=1}^{N_T}\left(1-\frac{c}{2}\right)^{w_{i\gamma}} = \sum_k \sum_{i:w_i\geqslant d} |\alpha_{ki}|^2 \left(1-\frac{c}{2}\right)^{\sum_{\gamma=1}^{N_T} w_{i\gamma}} \\ &= \sum_k \sum_{i:w_i\geqslant d} |\alpha_{ki}|^2 \left(1-\frac{c}{2}\right)^{w_i} \leqslant \sum_k \sum_{i:w_i\geqslant d} |\alpha_{ki}|^2 \left(1-\frac{c}{2}\right)^{d} \leqslant \left(1-\frac{c}{2}\right)^{d}. \end{aligned}$$

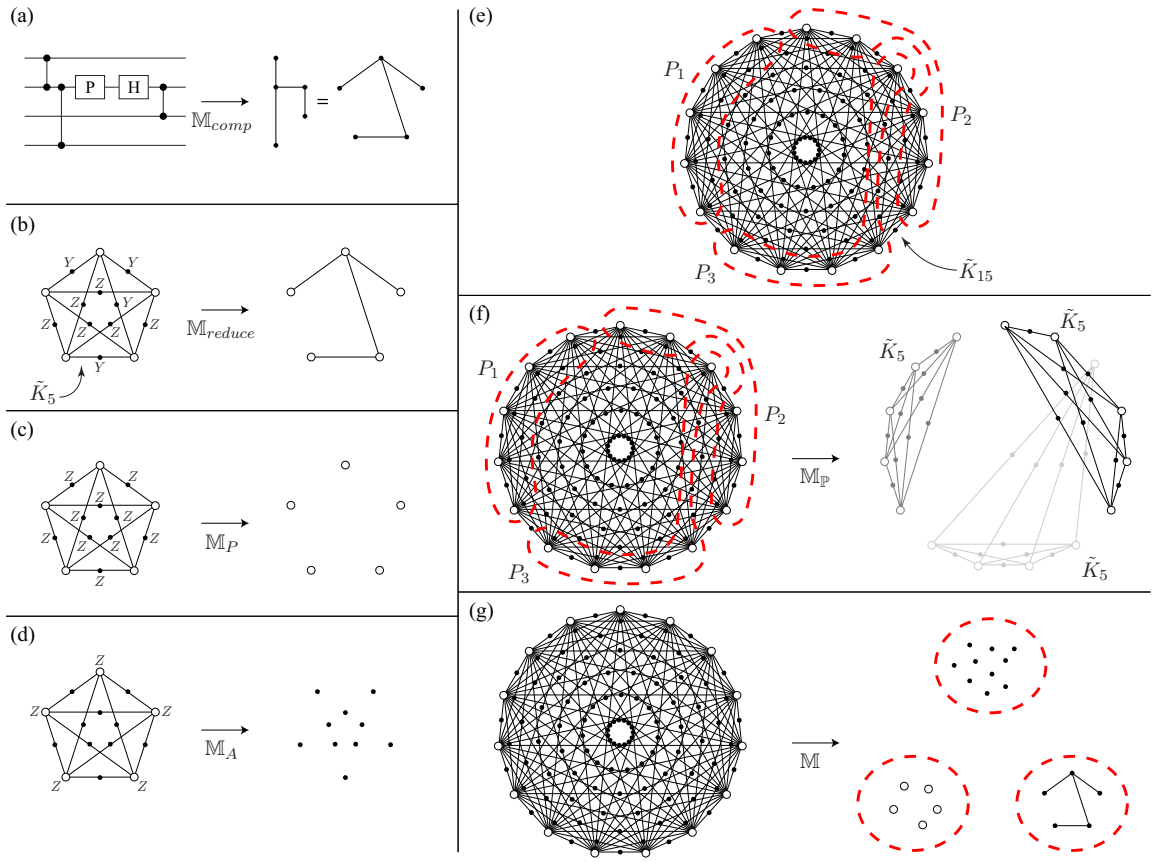


FIG. 9. Graphical depiction of Protocol 8. In this figure we replace the Raussendorf *et al.* encoding in the first step with a simpler computation, as to include a full encoding yields graphs too large to reasonably draw.

In the case of purely classical output this bound can be improved, since $|B_i^O| = |C_i^O| = |D_i^O| = 0$. Going through the same steps with this additional constraint gives

$$p_{\text{incorrect}} \leqslant \sum_k \sum_{i:w_i \geqslant d} |\alpha_{ki}|^2 \prod_{\gamma=1}^{N_T} (1 - c w_{i\gamma}) \leqslant (1-c)^d. \quad \blacksquare$$

We can now present the final contribution of this paper, a scheme for blind quantum computing that has all the previously described properties: correctness, universality, blindness of angles, input, output, computation, and more importantly verifiability with exponentially small probability of error. Roughly speaking, universality and correctness will be obtained by using dotted-complete graph states (similar to Protocol 4). In order to achieve verification we exploit the idea of dummy qubits (similar to Protocol 3) to create, blindly, out of a dotted-complete graph state $\tilde{\mathcal{K}}_{3N}$ three disconnected smaller dotted-complete graph states $\tilde{\mathcal{K}}_N$. Then we use two of these graph states to create $O(N)$ isolated trap qubits at random positions (similar to Protocol 5). The final step is to perform the actual computation over the remaining dotted-complete graph state in such a way that the stated property in Theorem 9 is also satisfied, that is, to have the measurement pattern encoded in such a way that any Pauli error with weight less than $d$ will be either corrected or detected. Such an encoding exists through the fault-tolerant one-way quantum computing scheme of [41]. All that is needed is to create a three-dimensional cluster state from the dotted-complete graph state and proceed with the fault-tolerant computation scheme of Raussendorf *et al.* [41,53].[6]

We first give a concrete protocol for choosing the required parameters for the Raussendorf *et al.* scheme, given the desired security threshold for the verification (see Protocol 7). This will fix the size of the dotted-graph state $N$ required for the actual computation. However, as stated above, we will start with a dotted-complete graph state of size $3N$ and break it into three smaller dotted-complete graph states of size $N$ each (see Fig. 9). We will refer to these graphs as the white trap graph, the black trap graph, and the computation graph. In the white trap graph all the vertices in $P(\tilde{K}_N)$ will become isolated traps (called white traps) by choosing all the vertices in $A(\tilde{K}_N)$ to be dummy qubits. Similarly, in the black trap graph all the vertices in $A(\tilde{K}_N)$ will become isolated traps (called black traps) by choosing all the vertices in $P(\tilde{K}_N)$ to be dummy

[6]In its original form, this scheme requires $Z$-basis measurements to be made adaptively, which is not easily implementable using dummy qubits. However, the location of the dummy qubits can be fixed by always including a correction step for each gate teleportation in the logical circuit, where the angle of the correction is adapted based on the outcome of the teleportation. An alternative option is to use a slightly modified version of the scheme due to Morimae and Fujii [31], which requires only measurements in the $X$-$Y$ plane. Although we assume the first scenario here, an almost identical proof applies to the second scenario.

qubits. We have to choose both type of vertices [$A(\tilde{K}_{3N})$ and $P(\tilde{K}_{3N})$] to be potentially isolated traps, otherwise Bob could choose to cheat on one type rather than the other one. In order to make the position of traps random, Alice will choose a random partition of $P(\tilde{K}_{3N})$ into three equal size sets and will choose appropriate dummy qubits (similar to Lemma 2) to obtain the three disconnected graphs. Note that this will lead to random positions for trap qubits; however, the positions of trap qubits will be also correlated with each other and we will take care of this issue when we present the proof of the verification. The above procedure is formalized in Protocol 7 and finally Protocol 8 presents a hiding protocol that is universal, verifiable, and blind.

As a high-level overview of the fault-tolerance scheme, qubits are encoded topologically as chains of defects (qubits to be measured in the $Z$ basis) of finite thickness and separation (referred to as the scale parameter) that trace out a path through the three-dimensional structure of the resource state. The encoding forces nondetectable errors to be topologically nontrivial chains, either connecting or encircling defect chains. Certain Clifford group operations are implemented directly by braiding these defect chains. For the remaining operations required for universality it is necessary to implement the gate by first distilling a suitable resource state that is then used to implement the gate via teleportation (all within the topologically encoded computation). While the teleportation can be done with Clifford group operations, the distillation is implemented on a concatenated encoding where at each level of concatenation the corresponding distillation step is topologically encoded with progressively higher defect thicknesses and scale parameters. At the lowest level, however, the operations are performed directly on physical qubits and so the defect chains are only a single qubit in diameter.

*Theorem 10.* Assume Alice and Bob follow the steps of Protocol 8. Then Alice always accepts the output and the outcome density operator is correct.

*Proof.* First we note that it is always possible to choose measurement patterns $\mathbb{M}_{\mathcal{P}}$ by Lemma 2 and $\mathbb{M}_{\text{reduce}}$ by Lemma 1. Further, by the universality of the Raussendorf *et al.* encoding, it is always possible to choose $\mathbb{M}_{\text{comp}}$. As the measurements composing $\mathbb{M}_{\mathcal{P}}$, $\mathbb{M}_{\text{reduce}}$, $\mathbb{M}_P$, and $\mathbb{M}_A$ are composed entirely of Pauli basis measurements, there is no partial time ordering imposed on the sequence of measurements and so the times at which these measurements are made have no effect on the outcome of the protocol. Thus, for any honest run of the protocol, the result will be the same as if the measurements from $\mathbb{M}_{\mathcal{P}}$ were made first. By construction this measurement pattern splits the graph state into three separate graph states $\tilde{\mathcal{K}}_N$.

The dummy qubits in $\mathbb{M}_P$ and $\mathbb{M}_A$ correspond to break operations in their respective graphs by Lemma 4 and hence after the initial step all the trap qubits remain unentangled from the rest. Recall that for these trap qubits $\phi_i = 0$ and since the qubit is prepared in the state $|+_{\theta_i}\rangle$ and measured in basis $\{|+_{\theta_i}\rangle, |-_{\theta_i}\rangle\}$, the measurement result communicated to Alice is $s_i = r_i$ for all such qubits. Thus, Alice always accepts, satisfying the first criterion.

By definition $\mathbb{M}_{\text{reduce}}$ transforms the graph state corresponding to $\tilde{K}_N$ to the resource state necessary to implement $\mathbb{M}_{\text{comp}}$. Finally, measuring according to $\mathbb{M}_{\text{comp}}$ yields the correct output of $\mathcal{C}$ by the correctness of the Raussendorf *et al.* protocol. ■

*Theorem 11.* Protocol 8 is blind while leaking at most $N$.

*Proof.* The proof is directly obtained from Theorem 4. ■

Protocol 7. Measurement pattern choice.

In what follows, choosing a measurement pattern means fixing the underlying graph state together with the appropriate angles of computation such that the resulting pattern implements the desired computation due to universality. Similarly choosing a partial measurement pattern means fixing the underlying graph state together with a partial set of angles of computation corresponding to a partial computation, where the rest of the angles will be fixed in Protocol 8 where this protocol is called as a subroutine. Here we assume that a standard labeling of the vertices of each dotted-complete graph state is known to both Alice and Bob.

(1) Alice chooses security parameter $d$ and then transforms the quantum circuit $\mathcal{C}$ corresponding to her desired computation into (or directly designs) a measurement pattern $\mathbb{M}_{\text{comp}}$ on a graph state $\mathcal{G}_{\mathcal{L}}$ that implements her computation using the encoding for topological fault-tolerant measurement-based quantum computation due to Raussendorf *et al.* [41], where $\mathcal{G}_{\mathcal{L}}$ is taken to correspond to the graph state of the three-dimensional lattice $\mathcal{L}$ introduced in [41] with sufficient dimensions $D_x$, $D_y$, and $D_z$ to implement her computation using an encoding with parameters as follows: defect thickness $d$, lattice scale parameter $\lambda = 5d$, distillation of resource states $|A\rangle$ and $|Y\rangle$ using $L = \lceil \log_3(d) \rceil$ levels, and for each concatenation level $1 < \ell < L$ the thickness parameter and scale parameter for that level are chosen as $d_\ell = 3d_{\ell-1}$ and $\lambda_\ell = \lambda_{\ell-1}$, with $d_1 = 1$, $\lambda_1 = 5$, $d_L = d$, and $\lambda_L = \lambda$.

(2) Alice chooses a partial measurement pattern $\mathbb{M}_{\text{reduce}}$ that reduces the graph state $\tilde{\mathcal{K}}_N$ to the graph state $\mathcal{G}_{\mathcal{L}}$ through Pauli measurements (Theorem 6), where $N$ is the total number of qubits in $\mathcal{L}$.

(3) Alice chooses a partial measurement pattern $\mathbb{M}_P$ on the graph state $\tilde{\mathcal{K}}_N$ such that every qubit corresponding to a vertex in $A(\tilde{K}_N)$ is set to be a dummy qubit. Hence all vertices in $P(\tilde{K}_N)$ are isolated traps.

(4) Alice chooses a partial measurement pattern $\mathbb{M}_A$ on the graph state $\tilde{\mathcal{K}}_N$ such that every qubit corresponding to a vertex in $P(\tilde{K}_N)$ is set to be a dummy qubit. Hence all vertices in $A(\tilde{K}_N)$ are isolated traps.

(5) For the graph $\tilde{K}_{3N}$, Alice chooses uniformly at random a partitioning $\mathbb{P}$ of the vertices into three equal-size sets of vertices $P_1$, $P_2$, and $P_3$.

(6) Alice takes $\mathbb{M}_{\mathbb{P}}$ to be the partial measurement pattern where the required vertices in $A(\tilde{K}_{3N})$ are set to be dummy qubits such that the resulting state is the tensor product of three graph states of the three disconnected graphs $\tilde{k}_1 = \tilde{K}_N$, $\tilde{k}_2 = \tilde{K}_N$, and $\tilde{k}_3 = \tilde{K}_N$ such that $P(\tilde{k}_i) = P_i$.

(7) Alice calculates $\mathbb{M}$, her overall measurement pattern on a graph state corresponding to $\tilde{K}_{3N}$, by combining the partial pattern $\mathbb{M}_{\mathbb{P}}$ with $\mathbb{M}_{\text{comp}}$ and $\mathbb{M}_{\text{reduce}}$ applied to subgraph $\tilde{k}_1$ and $\mathbb{M}_P$ and $\mathbb{M}_A$ applied to subgraphs $\tilde{k}_2$ and $\tilde{k}_3$, respectively, to obtain a full measurement pattern.

Protocol 8. Verifiable universal blind quantum computation.

(1) Alice's resources
(i) Alice chooses the pattern $\mathbb{M}$ and random partitioning $\mathcal{P}$ according to Protocol 7.
(ii) The dummy qubits position, set $D$, chosen according to Protocol 7.
(iii) A sequence of measurement angles $\phi = (\phi_i)_{1\leqslant i\leqslant 3N(3N+1)/2}$, with $\phi_i \in A$, according to the description of Protocol 7, where $\phi_i = 0$ for all the trap and dummy qubits. The ordering of the measurements on $P(\tilde{\mathcal{K}}_{3N})$ is chosen uniformly at random subject to the constraint that the partial ordering of measurements from $\mathbb{M}_{\text{comp}}$ determined by flow is preserved. Such a random ordering is required to hide the position of the trap qubits. The qubits in $A(\tilde{\mathcal{K}}_{3N})$ are measured first in the order that the relevant edge entry appears in the adjacency matrix of $\mathcal{K}_{3N}$ once this random ordering has been taken into account. That is, the site in $A(\tilde{\mathcal{K}}_{3N})$ that is joined by edges to $i$ and $j$ in $P(\tilde{\mathcal{K}}_{3N})$, with $i < j$ in the random ordering imposed on $P(\tilde{\mathcal{K}}_{3N})$, is measured in position $3N(i-1)+j-\frac{i(i+1)}{2}$. Note that the measurement order of the vertices in $A$ should be independent of the computation (and traps), so in the above we prescribe one such suitable sequence. This is followed by the measurements of $P(\tilde{\mathcal{K}}_{3N})$ in the randomly chosen order.
(iv) $3N(3N+1)/2$ random variables $\theta_i$ with value taken uniformly at random from $A$.
(v) $3N(3N+1)/2$ random variables $r_i$ and $|D|$ random variable $d_i$ with values taken uniformly at random from $\{0,1\}$.
(vi) A fixed function $C(i,\phi_i,\theta_i,r_i,\mathbf{s})$ that for each nonoutput qubit $i$ computes the angle of the measurement of qubit $i$ to be sent to Bob.

(2) Initial step
(i) Alice's move: Alice sets all the value in $\mathbf{s}$ to be 0 and prepares the qubits in the form

$$|d_i\rangle \forall i \in D,$$
$$\prod_{j\in N_G(i)\cap D} Z^{d_j}|+_{\theta_i}\rangle \forall i \notin D$$

and sends Bob all the $3N(3N+1)/2$ qubits in the order of the labeling of the vertices of the graph.
(ii) Bob's move: Bob receives $3N(3N+1)/2$ single qubits and entangles them according to $\tilde{K}_{3N}$.

(3) Step $i$: $1 \leqslant i \leqslant 3N(3N+1)/2$
(i) Alice's move: Alice computes the angle $\delta_i = C(i,\phi_i,\theta_i,r_i,\mathbf{s})$ and sends it to Bob.
(ii) Bob's move: Bob measures qubit $i$ with angle $\delta_i$ and sends Alice the result $b_i$.
(iii) Alice's move: Alice sets the value of $s_i$ in $\mathbf{s}$ to be $s_i + r_i$.

(4) Verification
Alice accepts if $s_i = r_i$ for all the white and black trap qubits $i$.

In order to prove the verification property, as stated in Theorem 9, we require that the measurement pattern is encoded in such a way that any Pauli error of weight less than $d$ will be either corrected or detected. We now show that this is true for the Raussendorf *et al.* scheme; although this is already implicit in their paper [41], we make it explicit here for completeness. In what follows we take $\mathcal{L}$ to be the three-dimensional lattice corresponding to the resource state used in [41].

*Lemma 5*. Let $\mathbb{M}_{\mathcal{C}}$ be a measurement pattern that implements a computation $\mathcal{C}$ on $\mathcal{G}_{\mathcal{L}}$, the graph state corresponding to the lattice $\mathcal{L}$, using the Raussendorf *et al.* fault-tolerance scheme with the following parameters: defect thickness $d$, lattice scale parameter $\lambda = 5d$, distillation of resource states $|A\rangle$ and $|Y\rangle$ using $L = \lceil \log_3(d)\rceil$ levels, and for each concatenation level $1 < \ell < L$ the thickness parameter and scale parameter for that level are chosen as $d_\ell = 3d_{\ell-1}$ and $\lambda_\ell = 3\lambda_{\ell-1}$, with $d_1 = 1$, $\lambda_1 = 5$, $d_L = d$, and $\lambda_L = \lambda$. Take $\sigma = \{\sigma^i\}$ to be a set of Pauli operators such that each $\sigma^i \in \{I,X,Y,Z\}$ and acts on qubit $i$. Then for any $\sigma$, if $\mathbb{M}_{\mathcal{C}}$ is implemented on state $|G_{\mathcal{L}}\rangle$, but the output of each measurement result or unmeasured qubit $i$ is modified by applying $\sigma^i$, then either the computation is correct (corresponding to a run where all $\sigma^i = I$) or an error is detected when the output is decoded, unless $|B_{\mathcal{L}}| + |C_{\mathcal{L}}| + |D^O_{\mathcal{L}}| \geqslant 2d$, where $B_{\mathcal{L}} = \{\gamma : \sigma^\gamma = X\}$, $C_{\mathcal{L}} = \{\gamma : \sigma^\gamma = Y\}$, and $D^O_{\mathcal{L}} = \{\gamma : \sigma^\gamma = Z,\ \gamma \in O\}$ and where $O$ is the set of output (unmeasured) qubits.

*Proof*. In the Raussendorf *et al.* scheme, logical qubits are topologically protected against errors. The two lowest weight topological errors are error cycles around defects and error chains running between defects. As defects have thickness $d$, any cross section forms a rectangle of dimension at least $d\times d$ and thus perimeter at least $4(d+1)$. As an error cycle must fit around the remaining defect and the minimum error cycle is at least $4d$. As the centers of defects are separated by distance $\lambda$, the minimum distance between defects is $\lambda - d$ and hence for our parameters we have $\lambda - d = 4d$.

The only region where this topological protection breaks down is within the regions used to distill the resource states $|A\rangle$ and $|Y\rangle$. This distillation is performed using a concatenation of $L$ levels of the Reed-Muller ($|A\rangle$) or Steane ($|Y\rangle$) codes. Each level $\ell$ of distillation is topologically protected with parameters $d_\ell$ and $\lambda_\ell$. As the Reed-Muller and Steane codes are both distance 3, an error at level $\ell$ can be caused by either a topological error at that level or not less than three errors at the previous level. However, since at each level $\ell < L$ we have $\lambda_\ell - d_\ell = 4d_\ell$ and $d_\ell = 3d_{\ell-1}$, the minimum weight $w_\ell$ to create an error at level $\ell$ is $\min(4d_\ell, 8d_{\ell-1}, 4d_{\ell-1} + w_{\ell-1}, 3w_{\ell-1})$. The four terms in this last expression account, respectively, for the minimum weight errors in each of the four possible cases: (1) The error is entirely topological at level $\ell$, (2) the error is entirely topological at level $\ell-1$, (3) the error includes both topological errors at level $\ell-1$ (which in the worst case affects two qubits with a single weight $4d_\ell$ error chain) and inherited errors from level $\ell-2$, and (4) the case where all errors are inherited from level $\ell-2$.

We then prove that $w_\ell > 2d_\ell$ by induction, as follows. Assume that at level $i$ we have $w_i > 2d_i$. In that case we have $w_{i+1} = \min(4d_{i+1}, 6d_i)$, since by assumption $4d_i + w_i > 6d_i$ and $3w_i > 6d_i$, and clearly $8d_i > 6d_i$. However, we have $d_{i+1} = 3d_i$ for all levels except the top level, where $d_L \leqslant 3d_{L-1}$. Thus, in general, $2d_{i+1} \leqslant 6d_i$ and hence $w_{i+1} > 2d_{i+1}$. At the lowest level the error distillation uses unencoded qubits measured in non-Pauli bases and so $w_0 = 1$ and $w_1 = 3 > 2d_1 = 2$ and thus by induction on $i$ we obtain the result that $w_L > 2d$ as required.

Note, however, that any operation on a measured qubit that is diagonal in the computational basis ($\sigma^i \in \{I, Z\}$) does not alter the computation. Hence an undetectable logical error is not created unless the total number of measured sites for which $\sigma^i \in \{X, Y\}$ plus the total number of output qubits for which $\sigma^i \in \{X, Y, Z\}$ is equal to or greater than $2d$. Thus the outcome is either correct or when decoded results in a detected error, unless $|B_\mathcal{L}| + |C_\mathcal{L}| + |D^O_\mathcal{L}| \geqslant 2d$. ■

Now we link the above general property of the Raussendorf *et al.* scheme to our specific protocol. To do so, we first introduce the notion of independently detectable errors.

*Definition 11.* Given a dotted-complete graph state $\tilde{K}_N$, a set of output qubits $O$, a measurement pattern $\mathbb{M}_{\text{target}}$ containing only $X$-$Y$ plane measurements and $Z$ basis measurements, and a set of single-qubit Pauli operators $\sigma = \{\sigma^i\}_{i=1}^N$, with $\sigma^i \in \{I, X, Y, Z\}$, which represent errors that modify each measurement result or unmeasured output qubit $i$ by the application of $\sigma^i$, for each location $i$ we define the set $\epsilon_i = \{i\}$ for $i \in P(\tilde{K}_N)$ and $\epsilon_i = N_{\tilde{K}_N}(i)$ for $i \in A(\tilde{K}_N)$. We say that $\sigma$ contains $k$ *independently detectable errors* if and only if there exists a set $\mathcal{E}$ of $k$ locations such that for all $i \in \mathcal{E}$, $\sigma^i \in \{X, Y\}$ if $i \notin O$ or $\sigma^i \in \{X, Y, Z\}$ if $i \in O$, and $\epsilon_i \cap \epsilon_j = 0$ for all pairs $i, j \in \mathcal{E}$.

The intuition behind this definition is that in Protocol 8 the qubits in $P(\tilde{K}_{3N})$ are independently randomly distributed between the two trap graphs and the computation graph and whether or not a qubit in $A(\tilde{K}_{3N})$ coincides with a trap depends only on the placement of the neighboring qubits [which are both in $P(\tilde{K}_{3N})$]. The first condition ensures that the error anticommutes with some possible measurement of the system and is hence truly an error, while the second condition ensures that we are considering only qubits associated with disjoint subsets of $P(\tilde{K}_{3N})$ and hence whether or not they coincide with a trap is uncorrelated. With this definition in place, we can proceed with proving a corollary to Lemma 5 that links that result with Protocol 8.

*Corollary 2.* Let $\mathbb{M}_\mathcal{C}$ be a measurement pattern that implements a computation $\mathcal{C}$ on graph state $\mathcal{G}_\mathcal{L}$ of $N$ vertices using the Raussendorf *et al.* scheme with parameters defect thickness $d$, lattice scale parameter $\lambda = 5d$, distillation of resource states $|A\rangle$ and $|Y\rangle$ using $L = \lceil \log_3(d) \rceil$ levels, and for each concatenation level $1 < \ell < L$ the thickness parameter and scale parameter for that level are chosen as $d_\ell = 3d_{\ell-1}$ and $\lambda_\ell = \lambda_{\ell-1}$, with $d_1 = 1$, $\lambda_1 = 5$, $d_L = d$, and $\lambda_L = \lambda$. Further, let $\mathbb{M}_{\text{reduce}}$ be a partial measurement pattern consisting of Pauli $Z$ and Pauli $Y$ measurements on qubits corresponding to the vertices in $A(\tilde{K}_N)$ that reduces $\tilde{\mathcal{K}}_N$ to $\mathcal{G}_\mathcal{L}$ up to local $Z$ rotations. Let $\mathbb{M}$ be the measurement pattern for graph state $\tilde{\mathcal{K}}_N$ produced by applying the partial pattern $\mathbb{M}_{\text{reduce}}$ to the qubits corresponding to vertices in $A(\tilde{K}_N)$ and $\mathbb{M}_\mathcal{C}$ (with appropriate local $Z$ rotations applied) to the qubits corresponding to vertices in $P(\tilde{K}_N)$.

Take $\sigma = \{\sigma^i\}$ to be a set of single-qubit Pauli operators such that each $\sigma^i \in \{I, X, Y, Z\}$ acts on qubit $i$. Then for any $\sigma$, if $\mathbb{M}_\mathcal{C}$ is implemented on state $\tilde{K}_N$, but the output of each measurement result or unmeasured qubit is modified by applying $\sigma^i$, then either the computation is correct (corresponding to a run where all $\sigma^i = I$) or an error is detected when the output is decoded, unless $\sigma$ contains at least $\lceil \frac{2d}{5} \rceil$ independently detectable errors.

*Proof.* First we note that only qubits in $P(\tilde{K}_{3N})$ are contained in $O$, since all qubits in $A(\tilde{K}_{3N})$ will be measured to make the required resource states. All measurements on qubits associated with vertices $A(\tilde{K}_N)$ are in either the $Y$ or $Z$ basis, allowing any error in the measurement outcome to be associated with an $X$ error on the underlying qubit. As the generators for the stabilizer of $\tilde{\mathcal{K}}_N$ are simply the operators $X_i \prod_{j \in N_{\tilde{\mathcal{K}}_N}(i)} Z_j$ and each vertex in $A(\tilde{K}_N)$ has only two neighbors, both of which lie in $P(\tilde{K}_N)$, an $X$ error on a qubit associated with a vertex in $A(\tilde{K}_N)$ is equivalent to a local error on each of two qubits in $P(\tilde{K}_N)$. Thus any local Pauli operator in $\sigma^i$ associated with a vertex in $A(\tilde{K}_N)$ either can be replaced by at most two local operators acting on qubits associated with vertices in $P(\tilde{K}_N)$ without altering the outcome of the computation or has no effect on the computation. Note that since Pauli $Z$ operators always commute with $Z$ basis measurements and anticommute with any measurement in the $X$-$Y$ plane, these local operators are always Pauli operators due to the corresponding restriction on $\mathbb{M}_{\text{target}}$.

The only Pauli terms that can affect the outcome of the computation are those that either flip a measurement outcome ($X$ or $Y$) or act nontrivially upon an unmeasured qubit (as either $X$, $Y$, or $Z$). By Lemma 5, the outcome of the computation is unaltered unless $\sigma$ produces such errors on at least $2d$ sites. To show that this implies the existence of at least $\lceil \frac{2d}{5} \rceil$ independently detectable errors we will consider the effects of errors on $A(\tilde{\mathcal{K}}_N)$ and $P(\tilde{\mathcal{K}}_N)$ in relation to the resource state for the Raussendorf *et al.* scheme $\mathcal{G}_\mathcal{L}$. Errors on $A(\tilde{\mathcal{K}}_N)$ only occur when the qubit in question is measured in the $Y$ basis, since for $Z$ basis measurements dummy qubits are used and the outcome of Bob's measurement is ignored. Thus, as we have shown above, such errors correspond to local Pauli errors at either end of an edge in the $\mathcal{G}_\mathcal{L}$. Errors in $P(\tilde{\mathcal{K}}_N)$, however, correspond simply to errors on single vertices in $\mathcal{G}_\mathcal{L}$. Therefore, we can consider any error introduced by $\sigma$ as corresponding to a subgraph $g_\sigma$ of $\mathcal{G}_\mathcal{L}$, where $i \in A(\tilde{\mathcal{K}}_N)$ introduces the vertices in $N_{\tilde{\mathcal{K}}_N}(i)$ together with a connecting edge, while $i \in P(\tilde{\mathcal{K}}_N)$ simply introduces the vertex $i$. Such a subgraph contains all of the qubits in $\mathcal{G}_\mathcal{L}$ that can possibly be affected by local errors after the measurement of qubits according to $\mathbb{M}_{\text{reduce}}$ are taken into account [propagating errors from $A(\tilde{\mathcal{K}}_N)$ to $P(\tilde{\mathcal{K}}_N)$].

We note that any connected subgraph $g^\gamma_\sigma$ of $g_\sigma$ containing $n_\gamma$ vertices necessarily contains at least $n_\gamma - 1$ edges. Note also that $\mathcal{G}_\mathcal{L}$ is four-edge colorable (see Fig. 10). Thus, by the pigeonhole principle, there is at least one color for that subgraph that corresponds to at least $\lceil \frac{n_\gamma - 1}{4} \rceil$ edges. As the various subgraphs $g^\gamma_\sigma$ are disconnected, we are free to choose the coloring independently for each and hence can choose a single four-edge coloring for $g_\sigma$ such that it includes at

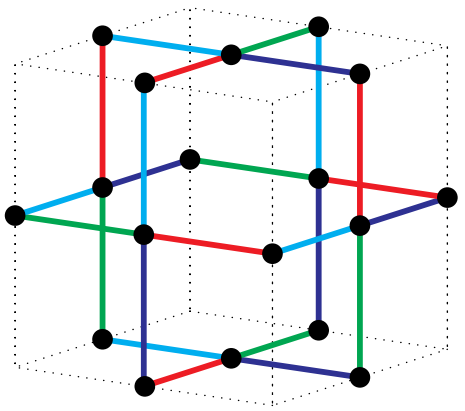

FIG. 10. Unit cell for the lattice corresponding to the Raussendorf *et al.* scheme $\mathcal{G}_{\mathcal{L}}$ complete with one choice of four-edge coloring.

least $\lceil \frac{n_\gamma - 1}{4} \rceil$ edges from each subgraph. We then take the set $\mathcal{E}$ to correspond to qubits in $A(\tilde{\mathcal{K}}_N)$ corresponding to edges of this color, as well as to the single vertex in any $g_\sigma^\gamma$ for which $n_\gamma = 1$, hence $\epsilon_i \cap \epsilon_j = 0$. By Lemma 5, this ensures that either the outcome of the computation is correct or an error is detected upon decoding, or $\sigma$ contains at least $\sum_{\gamma:n_\gamma \geqslant 2} \lceil \frac{n_\gamma - 1}{4} \rceil + \sum_{\gamma:n_\gamma = 1} 1$ independently detectable errors, where $\sum_\gamma n_\gamma \geqslant 2d$. Note that

$$\sum_{\gamma:n_\gamma \geqslant 2} \left\lceil \frac{n_\gamma - 1}{4} \right\rceil + \sum_{\gamma:n_\gamma = 1} 1 \geqslant \frac{2d}{5}$$

and hence either the computation is correct or an error is detected upon decoding, or $\sigma$ contains at least $\lceil \frac{2d}{5} \rceil$ independently detectable errors. ■

The above corollary guarantees that one of the conditions of Theorem 9 for the verification with the amplified security is satisfied. However, we cannot yet directly use that theorem since, as stated before, the positions of the traps are not completely random as the positions of the black traps are fixed once we choose the random position assignment of qubits in $P(\tilde{\mathcal{K}}_{3N})$ to each of the three subgraphs. This is why we have introduced the notion of independently detectable errors. Here we give a direct proof of verification for Protocol 8 following the same steps as the proof of Theorem 9.

*Theorem 12*. Protocol 8 is in general $(5/6)^{\lceil 2d/5 \rceil}$ verifiable and in the case of only classical output is $(2/3)^{\lceil 2d/5 \rceil}$ verifiable, where $d$ is the security parameter as described in Protocol 7.

*Proof*. The proof of this theorem follows the same strategy as Theorem 8, first taking the most general strategy for Bob, expanding this in terms of Pauli operators, and finally showing that any Pauli term that leads to an incorrect outcome is detected with high probability. We note that any deviation by Bob from Protocol 8 can be rewritten in the form shown in Fig. 7. The proof of this is identical to the corresponding step in the proof of Theorem 8: Without loss of generality, any deviation by Bob from the protocol can be written in the form of Fig. 6. We can treat $\{\delta_i\}$ as inputs to the circuit without violating causality, as they do not interact with any other part of the computation until after $b_j$ has been measured, for all $j < i$. Then simply by reordering the operators via their commutation relations we obtain the form in Fig. 7 as required.

As a result, any deviation by Bob can be written as a single deviation operator $\Omega$ that acts upon the quantum states Bob receives from Alice as well as $\delta_i$ and some private register held by Bob. Similar to the proof of Theorem 8 the probability of Alice accepting an incorrect outcome density operator is then

$$\begin{aligned} p_{\text{incorrect}} &= \sum_\nu p(\nu) \text{Tr}\left[ P^\nu_{\text{incorrect}} B_j(\nu) \right] \\ &= \sum_{b,\nu} p(\nu) \text{Tr}\Big( P_{\text{incorrect}} |b + c_r\rangle\langle b| C_{\nu_C,b} \Omega(\mathcal{P}|\Psi^{\nu,b}\rangle \\ &\quad \times \langle \Psi^{\nu,b}|\mathcal{P}^\dagger) C^\dagger_{\nu_C,b} |b\rangle\langle b + c_r| \Big) \\ &= \sum_{k,b,i,j,\nu} p(\nu) \alpha_{ki} \alpha^*_{kj} \text{Tr}\Bigg[ P_\perp \Bigg( \bigotimes_{t\in T} |\eta_t^{\nu_T}\rangle\langle \eta_t^{\nu_T}| \Bigg) y \\ &\quad \times |b + c_r\rangle\langle b| C_{\nu_C,b} \sigma_i \mathcal{P} |\Psi^{\nu,b}\rangle \\ &\quad \times \langle \Psi^{\nu,b}| \mathcal{P}^\dagger \sigma_j C^\dagger_{\nu_C,b} |b\rangle\langle b + c_r| \Bigg], \end{aligned}$$

where as in previous proofs we take the Kraus operators associated with the $\Omega$, once Bob's private system has been removed, to be $\chi_k = \sum_i \alpha_{ki}\sigma_i$, with $\sum_k \sum_i \alpha_{ki}\alpha^*_{ki} = 1$.

By Corollary 2, $P_\perp$ projects out the terms in the above sum where $\sigma_i$ does not contain at least $\lceil \frac{2d}{5} \rceil$ independently detectable errors on the computation graph. This is a somewhat stronger condition than we actually need and so we will consider terms corresponding to any $\sigma_i$ that produces at least $\lceil \frac{2d}{5} \rceil$ independently detectable errors in total across all three subgraphs (the computation graph and the two trap graphs). We will denote by $\mathcal{I}$ the set of all $i$ for which $\sigma_i$ does not satisfy this condition. Similar to the proof of Theorem 8, all terms for which $i \neq j$ average to zero. Thus, as in the proof of Theorem 9, we obtain

$$\begin{aligned} p_{\text{incorrect}} \leqslant \sum_k \sum_{i\notin\mathcal{I}} \sum_T p(T) |\alpha_{ki}|^2 \prod_{t\in T} \\ \times \left( \sum_{\theta_t, r_t} p(\theta_t) p(r_t) \langle \eta_t^{\nu_T} | \sigma_i | \eta_t^{\nu_T} \rangle^2 \right). \end{aligned}$$

As before, we introduce notional sets $S_\gamma$ of three qubits each such that exactly one qubit from each set is on each of the three subgraphs (the two trap graphs and the computation graph) and where either all of the qubits are in $P(\tilde{\mathcal{K}}_{3N})$ or all of the qubits are in $A(\tilde{\mathcal{K}}_{3N})$ (ensuring exactly one trap and at least one dummy qubit per set). As every $\sigma_i$ in the above sum corresponds to at least $\lceil \frac{2d}{5} \rceil$ independently detectable (and hence uncorrelated) errors across these sets $S_\gamma$, we have

$$\begin{aligned} p_{\text{incorrect}} &\leqslant \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \prod_\gamma \\ &\quad \times \left( \sum_{t_\gamma, r_{t_\gamma}, \theta_{t_\gamma}} p(t_\gamma) p\left(r_{t_\gamma}\right) p\left(\theta_{t_\gamma}\right) \langle \eta_t^{\nu_T} | \sigma_i | \eta_t^{\nu_T} \rangle^2 \right) \\ &= \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \prod_\gamma \left( \sum_{t_\gamma, r_{t_\gamma}, \theta_{t_\gamma}} \frac{1}{48} \langle \eta_t^{\nu_T} | \sigma_i | \eta_t^{\nu_T} \rangle^2 \right), \end{aligned}$$

where as before $t_\gamma$ denotes the location of the trap qubit in set $S_\gamma$. Averaging over all values of $t_\gamma$, $r_{t_\gamma}$, and $\theta_{t_\gamma}$, we obtain

$$\begin{aligned} p_{\text{incorrect}} &\leqslant \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \prod_\gamma \left(1-\frac{w_\gamma}{6}\right) \\ &\leqslant \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \prod_\gamma \left(1-\frac{1}{6}\right)^{w_\gamma} \\ &= \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \left(\frac{5}{6}\right)^{\sum_\gamma w_\gamma} \\ &\leqslant \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \left(\frac{5}{6}\right)^{\lceil 2d/5\rceil} \\ &\leqslant \left(\frac{5}{6}\right)^{\lceil 2d/5\rceil}, \end{aligned}$$

where $w_\gamma$ denotes the number of independently detectable errors that fall within set $S_\gamma$. In the special case of all classical output, however, the bound can be made tighter, since $|\eta^v_{t_\gamma}\rangle = |r^v_{t_\gamma}\rangle$, and hence

$$\begin{aligned} p_{\text{incorrect}} &\leqslant \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \prod_\gamma \mathrm{Tr}\left(\sum_{t_\gamma, r_{t_\gamma}} \frac{1}{6}\langle r^v_{t_\gamma}|\sigma_{i|t}|r^v_{t_\gamma}\rangle^2\right) \\ &\leqslant \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \prod_\gamma \left(1-\frac{w_\gamma}{3}\right) \\ &\leqslant \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \prod_\gamma \left(1-\frac{1}{3}\right)^{w_\gamma} \\ &= \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \left(\frac{2}{3}\right)^{\sum_\gamma w_\gamma} \\ &\leqslant \sum_k \sum_{i\notin\mathcal{I}} |\alpha_{ki}|^2 \left(\frac{2}{3}\right)^{\lceil 2d/5\rceil} \\ &\leqslant \left(\frac{2}{3}\right)^{\lceil 2d/5\rceil}. \end{aligned}$$

■

## VIII. CONCLUSION

We have extended the original universal blind quantum computing protocol presented in [3] with different concepts of blind preparation of isolated dummy qubits (a qubit prepared randomly in the set $\{|0\rangle,|1\rangle\}$) and isolated trap qubits (a qubit prepared randomly in the set $\{|+\rangle_\theta\}$). These two modifications lead to a different construction for unconditionally verifiable blind quantum computation. However, in this way only polynomially bounded security could be achieved. Building upon these ideas, combined with fault-tolerant computation, we presented a UBQC protocol that achieves exponentially bounded security for the verification scheme using a different resource state, the dotted-complete graph state. This protocol extends the topological fault-tolerant measurement-based quantum computation scheme due to Raussendorf *et al.* [41] to a blind setting. We note that while consideration of fault tolerance in the blind computation itself is beyond the scope of the present work, if Protocol 8 is modified so as to allow Alice to accept a finite error rate on the trap qubits, the probability of Bob successfully cheating is exponentially suppressed in the gap between the expected error weight inferred from trap measurements and our threshold of $\lceil \frac{2d}{5} \rceil$ and so a fault-tolerant adaptation of this protocol should be possible.

As mentioned before, a verifiable UBQC protocol can be viewed as an interactive proof system where Alice acts as the verifier and Bob as the prover [3–5]. This link to complexity theory suggests a different approach to questions such as the open problem of finding an interactive proof for any problem in BQP with a BQP prover, but with a purely classical verifier. The conceptual link between blindness and interactive proof systems is the key ingredient for verifying the high-complexity quantum-theoretic models with low-complexity classical ones.

## ACKNOWLEDGMENTS

We thank Anne Broadbent for many insightful discussions throughout the writing of this paper. We would also like to acknowledge Robert Raussendorf and Earl Campbell for their help on the properties of the topological fault-tolerance scheme. We also thank Vedran Dunjko, Iordanis Kerenedis, Urmila Mahadev, and Tomoyuki Morimae for helpful discussions on the proof of Theorem 8 and for pointing out to us an error in the first draft. J.F.F. acknowledges support from the National Research Foundation and Ministery of Education, Singapore. This material is based on research supported in part by the Singapore National Research Foundation under National Research Foundation (Singapore). E.K. acknowledges support from Engineering and Physical Sciences Research Council Grant No. EP/E059600/1.

[1] A. M. Childs, Secure assisted quantum computation, Quantum Inf. Comput. **5**, 456 (2005).

[2] P. Arrighi and L. Salvail, Blind quantum computation, Int. J. Quantum Inf. **4**, 883 (2006).

[3] A. Broadbent, J. Fitzsimons, and E. Kashefi, in *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science, Atlanta, 2009* (IEEE, Piscataway, NJ, 2009), p. 517.

[4] D. Aharonov, M. Ben-Or, and E. Eban, in *Proceedings of Innovations in Computer Science, Beijing, 2010* (Tsinghua University Press, Beijing, 2010), p. 453.

[5] B. Reichardt, F. Unger, and U. Vazirani, Classical command of quantum systems, Nature (London) **496**, 456 (2013).

[6] A. Broadbent, G. Gutoski, and D. Stebila, in *Advances in Cryptology—CRYPTO 2013, Proceedings of the 33rd Cryptology Conference, Santa Barbara, 2013*, edited by R. Canetti and J. A. Garay, Lecture Notes in Computer Science (Springer, Berlin, 2013), pp. 344–360.

[7] J. Feigenbaum, in *Advances in Cryptology, Proceedings of the CRYPTO '85*, edited by H. C. Williams, Lecture Notes in Computer Science Vol. 218 (Springer, Berlin, 1986), p. 477.

[8] M. Abadi, J. Feigenbaum, and J. Kilian, On hiding information from an oracle, J. Comput. Syst. Sci. **39**, 21 (1989).

[9] R. Rivest, L. Adleman, and M. Dertouzos, On data banks and privacy homomorphisms, Found. Secure Comput. **32**, 169 (1978).

[10] R. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Commun. ACM **21**, 120 (1978).

[11] C. Gentry, in *Proceedings of the 41st Annual ACM Symposium on Theory of Computing* (ACM Press, New York, 2009), p. 169.

[12] V. Vaikuntanathan, in *Progress in Cryptology—INDOCRYPT 2012, Proceedings of the 12th International Conference on Cryptology in India, Chennai, 2011*, edited by S. Galbraith and M. Nandi, Security and Cryptology Vol. 7668 (Springer, Berlin, 2012), pp. 1–15.

[13] P. P. Rohde, J. F. Fitzsimons, and A. Gilchrist, Quantum Walks with Encrypted Data, Phys. Rev. Lett. **109**, 150501 (2012).

[14] S.-H. Tan, J. A. Kettlewell, Y. Ouyang, L. Chen, and J. F. Fitzsimons, A quantum approach to homomorphic encryption, Sci. Rep. **6**, 33467 (2016).

[15] A. Broadbent and S. Jeffery, in *Advances in Cryptology—CRYPTO 2015, Proceedings of the 35th Annual Cryptology Conference, Santa Barbara, 2015*, edited by R. Gennaro and M. Robshaw, Security and Cryptology Vol. 9215 (Springer, Berlin, 2015), Pt. 1, pp. 609–629.

[16] Y. Ouyang, S.-H. Tan, and J. Fitzsimons, Quantum homomorphic encryption from quantum codes, arXiv:1508.00938.

[17] Y. Dulek, C. Schaffner, and F. Speelman, in *Advances in Cryptology—CRYPTO 2016, Proceedings of the 35th Annual Cryptology Conference, Santa Barbara, 2016*, edited by M. Robshaw and J. Katz, Security and Cryptology Vol. 9814 (Springer, Berlin, 2016), Pt. 1, pp. 3–32.

[18] L. Yu, C. A. Pérez-Delgado, and J. F. Fitzsimons, Limitations on information-theoretically-secure quantum homomorphic encryption, Phys. Rev. A **90**, 050303 (2014).

[19] A. Ambainis, M. Mosca, A. Tapp, and R. Wolf, in *Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS 2000)* (IEEE, Piscataway, NJ, 2000), pp. 547–553.

[20] P. O. Boykin and V. Roychowdhury, Optimal encryption of quantum bits, Phys. Rev. A **67**, 042317 (2003).

[21] R. Raussendorf and H. J. Briegel, A One-Way Quantum Computer, Phys. Rev. Lett. **86**, 5188 (2001).

[22] V. Danos, E. Kashefi, and P. Panangaden, The measurement calculus, J. ACM **54**, 8 (2007).

[23] S. Barz *et al.*, Demonstration of blind quantum computing, Science **335**, 303 (2012).

[24] A. Gheorghiu, E. Kashefi, and P. Wallden, Robustness and device independence of verifiable blind quantum computing, New J. Phys. **17**, 083040 (2015).

[25] M. Hajdušek, C. A. Pérez-Delgado, and J. F. Fitzsimons, Device-independent verifiable blind quantum computation, arXiv:1502.02563.

[26] T. Morimae, Continuous-Variable Blind Quantum Computation, Phys. Rev. Lett. **109**, 230502 (2012).

[27] T. Sueki, T. Koshiba, and T. Morimae, Ancilla-driven universal blind quantum computation, Phys. Rev. A **87**, 060301(R) (2013).

[28] T. Morimae, V. Dunjko, and E. Kashefi, Ground state blind quantum computation on AKLT state, Quantum Inf. Comput. **15**, 200 (2015).

[29] T. Morimae and K. Fujii, Blind quantum computation for alice who does only measurements, Phys. Rev. A **87**, 050301(R) (2013).

[30] Q. Li, W. H. Chan, C. Wu, and Z. Wen, Triple-server blind quantum computation using entanglement swapping, Phys. Rev. A **89**, 040302(R) (2014).

[31] T. Morimae and K. Fujii, Blind topological measurement-based quantum computation, Nat. Commun. **3**, 1036 (2012).

[32] C. H. Chien, R. V. Meter, and S. Y. Kuo, Fault-tolerant operations for universal blind quantum computation, ACM J. Emerg. Technol. Comput. Sys. **12**, 9 (2015).

[33] V. Giovannetti, L. Maccone, T. Morimae, and T. G. Rudolph, Efficient Universal Blind Computation, Phys. Rev. Lett. **111**, 230501 (2013).

[34] A. Mantri, C. A. Pérez-Delgado, and J. F. Fitzsimons, Optimal Blind Quantum Computation, Phys. Rev. Lett. **111**, 230502 (2013).

[35] C. A. Pérez-Delgado and J. F. Fitzsimons, Iterated Gate Teleportation and Blind Quantum Computation, Phys. Rev. Lett. **114**, 220502 (2015).

[36] V. Dunjko, E. Kashefi, and A. Leverrier, Blind Quantum Computing with Weak Coherent Pulses, Phys. Rev. Lett. **108**, 200502 (2012).

[37] K. Fisher *et al.*, Quantum computing on encrypted data, Nat. Commun. **5**, 3074 (2014).

[38] S. Barz, J. Fitzsimons, E. Kashefi, and P. Walther, Experimental verification of quantum computation, Nat. Phys. **9**, 727 (2013).

[39] M. Mosca and D. Stebila, Quantum coins, Contemp. Math. **523**, 35 (2010).

[40] I. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, Rev. Mod. Phys. **86**, 153 (2014).

[41] R. Raussendorf, J. Harrington, and K. Goyal, Topological fault-tolerance in cluster state quantum computation, New J. Phys. **9**, 199 (2007).

[42] V. Dunjko, J. F. Fitzsimons, C. Portmann, and R. Renner, in *Advances in Cryptology—ASIACRYPT 2014, Proceedings of the 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, 2014*, edited by P. Sarkar and T. Iwata, Lecture Notes in Computer Science Vol. 8874 (Springer, Berlin, 2014), pp. 406–425.

[43] D. Markham and B. C. Sanders, Graph states for quantum secret sharing, Phys. Rev. A **78**, 042309 (2008).

[44] V. Danos and E. Kashefi, Determinism in the one-way model, Phys. Rev. A **74**, 052310 (2006).

[45] D. Browne, E. Kashefi, M. Mhalla, and S. Perdrix, Generalized flow and determinism in measurement-based quantum computation, New J. Phys. **9**, 250 (2007).

[46] A. Broadbent and E. Kashefi, Parallelizing quantum circuits, Theor. Comput. Sci. **410**, 2489 (2009).

[47] D. E. Browne, E. Kashefi, and S. Perdrix, in *Proceedings of the Fifth Conference on Theory of Quantum Computation, Communication, and Cryptography, Leeds, 2010* (ACM Press, New York, 2010), p. 35.

[48] M. V. den Nest, W. Dur, A. Miyake, and H. J. Briegel, Fundamentals of universality in one-way quantum computation, New J. Phys. **9**, 204 (2007).

[49] V. Danos, E. Kashefi, and P. Panangaden, Parsimonious and robust realizations of unitary maps in the one-way model, Phys. Rev. A **72**, 064301 (2005).

[50] A. M. Childs, D. W. Leung, and M. A. Nielsen, Unified derivations of measurement-based schemes for quantum computation, Phys. Rev. A **71**, 032318 (2005).

[51] M. Hein, J. Eisert, and H. J. Briegel, Multi-party entanglement in graph states, Phys. Rev. A **69**, 062311 (2004).

[52] H. Barnum, C. Crépeau, D. Gottesman, A. Smith, and A. Tapp, in *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2002)* (IEEE, Piscataway, NJ, 2002), p. 449.

[53] R. Raussendorf, J. Harrington, and K. Goyal, A fault-tolerant one-way quantum computer, Ann. Phys. (NY) **321**, 2242 (2006).