

# Nonadaptive fault-tolerant verification of quantum supremacy with noise

Theodoros Kapourniotis<sup>\*</sup> and Animesh Datta<sup>†</sup>

*Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom*

(Dated: March 29, 2017)

Quantum samplers are believed capable of sampling efficiently from distributions that are classically hard to sample from. We consider a sampler inspired by the Ising model. It is nonadaptive and therefore experimentally amenable. Under a plausible average-case hardness conjecture, classical sampling upto additive errors from this model is known to be hard. We present a trap-based verification scheme for quantum supremacy that only requires the verifier to prepare single-qubit states. The verification is done on the same model as the original sampler, a square lattice, with only a constant factor overhead. We next revamp our verification scheme to operate in the presence of noise by emulating a fault-tolerant procedure without correcting on-line for the errors, thus keeping the model non-adaptive, but verifying supremacy fault-tolerantly. We show that classically sampling upto additive errors is likely hard in our revamped scheme. Our results are applicable to more general sampling problems such as the Instantaneous Quantum Polynomial-time (IQP) computation model. It should also assist near-term attempts at experimentally demonstrating quantum supremacy and guide long-term ones.

## I. INTRODUCTION

Considerable experimental efforts are being directed towards the realisation of quantum information processing technologies, with the eventual aim of constructing universal quantum computers and simulators [1–4]. One of the motivations for this exercise is their expected ability to simulate physical systems that are believed to be intractable classically [5, 6]. An example of such a system is a lattice of interacting spins in the presence of a magnetic field, represented by the Ising model [7]. The Ising model is a workhorse in condensed matter physics, employed to study magnetic properties of spin glasses [8] and spin liquids [9] as well as phase transitions in many-body systems such as lattice gases [10], and numerous other problems. A quantity central to most of these studies is the partition function. Computing the partition function (and ground state) of the Ising model in an external magnetic field is, however, NP-hard even in two dimensions [11].

Although NP-hard problems are not expected to be solvable efficiently on a universal quantum computer, recent theoretical results [12–17] suggest the possibility of demonstrating ‘quantum supremacy’ [18] in devices that do not require the full complement of DiVincenzo’s criteria [19] for their implementation. Their scalable implementation is thus anticipated to be more achievable than a universal quantum computer’s, providing tangible demonstrations of quantum supremacy sooner [16, 20–23]. This expectation is purchased at the price of sacrificing the full power of universal quantum computers, promising only to efficiently sample from certain distributions instead. Among them is a non-universal quantum simulator based on a nonadaptive translationally-invariant Ising model that allows sampling from the dis-

tribution of Ising model partition functions [17]. It shows, in line with other models [12–16], that under an average-case hardness conjecture, a highly unexpected collapse of the polynomial hierarchy to the third level occurs if a classical sampler can sample the partition function distribution upto additive errors. Unlike these models, however, the hardness results of the Ising sampler [17] hold for a single fixed instance of the problem, relieving the burden of creating random instances.

We tackle two crucial challenges facing any experimental demonstration of quantum supremacy for the Ising sampler. The first is verifying that the output distribution is indeed from the correct quantum sampler, as opposed to a classical sampler emulating a similar distribution [24]. The second is ensuring that the sampling task remains classically hard to simulate even in the presence of experimental noise and decoherence [24].

In this paper, we present a nonadaptive verification scheme with exponentially low probability of failure and only linear complexity for the Ising sampler (prover). Our scheme applies to a totally untrusted prover with entangling and measuring capabilities, limited only by the laws of quantum mechanics, and requires the verifier to prepare random, single-qubit states with constant local errors. We first present a verification scheme that should aid demonstrating quantum supremacy with few qubits (Theorem 1). We then prove a fault-tolerant version (Theorem 2), which is achieved by an amalgamation of trap-based verification techniques for decision problems [25] with recent results on demonstrating quantum supremacy by emulating fault-tolerance via post-selection [23]. For the former, we develop a simpler construction performing verification and computation on different planar graphs. For the latter, we prove that our fault-tolerant verifiable model is classically hard subject to the same average-case hardness conjecture (Theorem 3). Our work thus opens the way for tailored trap-based, fault-tolerant verification techniques for quantum sampling problems.

<sup>\*</sup> T.Kapourniotis@warwick.ac.uk

<sup>†</sup> animesh.datta@warwick.ac.uk

Our fault-tolerant verification scheme applies to any non-adaptive sampler based on cluster states, however we will use the Ising sampler of [17] as a particular example, thus keeping the benefit of the single instance property and experimental feasibility of this model. Our work is however a significant improvement on the verification method of the original, which was based on certification techniques [26] and required local measurements with errors polynomially small in the total number of qubits and quadratically many exchanges between verifier and prover.

## II. QUANTUM SAMPLING IN THE VERIFIER-PROVER SETTING

In the verifier-prover setting, the verifier can prepare bounded-error, single-qubit states and the prover implements the rest of the computation including the measurements, and returns the output samples to the verifier. The role of the verifier is to ascertain if the prover is acting honestly and executing the correct operation. The prover is, in general, malicious, trying to pass any tests designed by the verifier while deviating from the correct implementation at the same time. This malice may be intentional if the prover is trying to convince the verifier of its quantum power when it has none, or incidental if the prover possesses an imperfect quantum device prone to noise and errors. We assume that the prover's deviations are governed by quantum mechanics.

We begin by adapting the Ising spin model to a blind verifier-prover cryptographic setting [27]. Blindness, which ensures that the prover remains ignorant of the actual computation, is a necessary ingredient in our verification scheme. Our Ising spin model consists of qubits in state  $|+\rangle$  subject to nearest neighbour controlled  $\pi$ -phase rotations, denoted by  $cZ$ . All the qubits are measured simultaneously in a basis in the  $xy$ -plane of the Bloch sphere. The measurement outcome of classical bits is the output sample. This model corresponds to the well-studied measurement-based quantum computing (MBQC) model [28] without the adaptive measurements. This last restriction, which relaxes one of DiVincenzo's criteria, makes this model non-universal for quantum computing.

In the particular Ising sampler presented in [17] (for more details see Appendix B), the structure of the graph state is fixed, but its size scales with the width  $m$  and depth  $n$ . The measurement angles are also fixed to specific values from the set  $\{-\pi/4, -\pi/8, 0, \pi/8, \pi/4\}$ . This choice of graph, which we call the 'extended' brickwork state, and a fixed angle for each physical qubit has the following benefit: Each possible combination of measurement outcomes 'chooses' a different angle for each qubit of the original brickwork state from the set  $\{k\pi/4\}$ ,  $k = \{0, \dots, 7\}$ . This effectively makes a single instance of the model a random quantum circuit generator, a property exploited to prove its hardness.

The correspondence to a quantum Ising model comes from the locality of spin interactions and decomposing each MBQC measurement into a unitary rotation around the  $z$ -axis corresponding to an external magnetic field, followed by a Pauli  $X$  measurement. The quantum state just before the Pauli  $X$  measurement is given by the unitary evolution due to the Hamiltonian

$$\mathcal{H} = - \sum_{i,j} J Z_i Z_j + \sum_i B_i Z_i \quad (1)$$

where  $J$  is the local interaction term,  $B_i$  the local field strength and  $Z_i$  the Pauli  $Z$  operator on qubit  $i$ .

The probability  $p(\mathbf{x})$  of measuring a bit string  $\mathbf{x}$  corresponds to the partition function  $\mathcal{Z}_{\mathbf{x}}$  of the Ising model with Hamiltonian  $\mathcal{H}' \equiv \mathcal{H} + \frac{\pi}{2} \sum_i x_i Z_i$  is given by

$$p(\mathbf{x}) = \frac{|\text{Tr}(e^{-i(\mathcal{H} + \frac{\pi}{2} \sum_i x_i Z_i)})|^2}{2^{2N}} \equiv \frac{|\mathcal{Z}_{\mathbf{x}}|^2}{2^{2N}}, \quad (2)$$

where  $N = mn$ . The second term in  $\mathcal{H}'$  comes from the measurement outcomes of the Pauli  $X$  measurements, and the partition function is evaluated at an imaginary temperature  $\beta = 1/k_B T = i$ . Although the partition function at imaginary temperatures may appear unphysical, it has deep connections to quantum complexity theory [29] as well as quantum statistical and condensed matter physics via analytic continuations. The model could thus be used to sample from the partition function of the classical Ising model with real temperatures [30], but the probability of observing a sample is exponentially low in the problem size.

Testing the honesty of the prover, in our case the Ising sampler, requires the 'blind' injection of certain 'trap' qubits. To keep the identity of these trap qubits from the prover, the verifier applies some encoding on the original translationally-invariant Ising spin model. This makes the model translationally variant - now both the participating qubits and the measurement angles on the graph state have a randomly chosen extra rotation according to the scheme described next. Incidentally, this relaxation away from translational invariance could arguably go a long way in easing its experimental implementation.

Specifically, each qubit  $i$  is individually prepared by the verifier in the state  $|+\theta_i\rangle$ , where  $\theta_i$  is chosen uniformly at random from the set  $A = \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \dots, \frac{15\pi}{8}\}$ . Instead of the prover measuring in fixed predetermined angles, as in the original Ising sampler, the verifier sends encrypted angles to the prover:  $\delta_i = \theta_i + (-1)^{r'_i} \phi_i + r_i \pi$  for  $r_i, r'_i \in_R \{0, 1\}$ . Rotations by  $\theta_i$  on the qubit and on the angles mutually cancel and the classical information that the prover receives (containing the actual measurement angles  $\phi_i$ ) is classically one time padded by  $\theta_i$ . The bits  $r_i, r'_i$  provide some extra randomness to restrict the information the prover gets from the quantum state and can be corrected by classical post-processing of the sample. Our innovation over [27] lies in the number of angles used in the set  $A$ , and comes from the fact that we use

a different decomposition of the computation. We next present the details of the construction of our verification protocol.

### III. VERIFICATION OF ISING SAMPLERS

This section contains our main results, in two stages, on the verification of the Ising sampler's output. We first present a non-fault-tolerant verification scheme (Theorem 1). Since it requires decreasing qubit preparation errors with increasing system size, this is only viable in small-sized experimental demonstrations of quantum supremacy. We next present a fault-tolerant version (Theorem 2) which is scalable but requires qubit preparation errors below certain thresholds.

Establishing verifiability consists of proving correctness and soundness. Correctness means that the protocol performs the correct computation when the prover is honest. A protocol is sound if, for any deviation of the prover, the verifier is able with a good confidence level, to accept or reject the hypothesis that the output distribution is close to the correct one. This deviation captures both a malevolent prover who tries to cheat and uncontrollable errors in the prover's device. The soundness parameter [25] is a bound on the probability that the hypothesis test fails for any deviation of the prover.

#### A. Non-fault-tolerant verification

The output of a quantum sampler must be classical for it to be comparable to that of a classical sampler, a prerequisite for demonstrating quantum supremacy. This allows us to simplify trap-based verification strategies for universal quantum computation [25, 31, 32] to having disjointed computational resource and trap states - an idea also used in circuit-based verification [33]. This permits an exponentially small error in our estimation of the fidelity of the output using a square lattice. Finally, a trap-based technique instead of certification ones [17, 26] enables us to reduce the resource complexity of the verification protocol from quadratic to linear.

Our verification protocol relies on judiciously selecting the measurement angles and placing special *dummy* qubits prepared in the state  $|0\rangle$ , which, together with  $xy$ -plane measurements, allows us to *carve* different types of graphs from a square lattice graph, as shown in Fig. (1). Placing one dummy qubit between any two other qubits prevents the prover's entangling operators to have any entangling effect between the participating qubits, so the prover can apply exactly the same operations that produce the square lattice but create a different graph state. The graphs carved out are the 'extended' brickwork state (Appendix B) and two other graphs that contain special 'trap' qubits in the state  $|+\rangle$ . The extended brickwork state is used to run the Ising sampler. The traps in

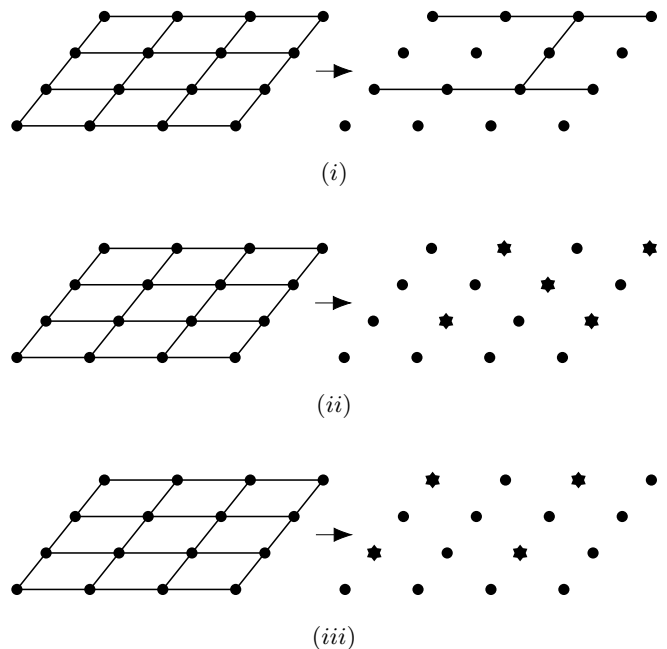


FIG. 1. Verifier chooses a random ordering of  $2\kappa + 1$  graph states - the computational graph on the right of figure (i);  $\kappa$  identical test graphs on the right of figure (ii) which have traps (starred nodes) on the even parity positions of the sub-graph that corresponds to the computational graph; and  $\kappa$  identical graphs on the right of figure (iii) which have traps on the odd parity positions of the sub-graph that corresponds to the computational graph. All of these graphs can be generated from a square lattice (on the left) by replacing  $|+\rangle$  qubits with  $|0\rangle$  at the positions (isolated dot nodes) we do not want entangled with their neighbours when  $cZ$  is applied. Further detail on the carving procedure, which can be made blind (Section II), is provided in Appendix A.

the test graphs are measured in the same basis as prepared, yielding a deterministic test for the prover. Two different types of the trap graphs are needed to enable placing a trap at any position in the graph with equal probability. The order of the graphs is chosen at random and the whole procedure is implemented blindly to thwart the prover from distinguishing test and computation qubits.

A sketch of the protocol appears in Fig. (2), and the details in Appendix B. Time complexity of the quantum operations in the protocol is constant and the number of qubits needed is  $O(N)$ , where  $N$  is the number of the qubits of the Ising sampler. We prove the correctness of our protocol in Appendix B 1. For the soundness of our protocol the following theorem holds.

**Theorem 1** (Soundness). *Suppose the qubit preparation error by the verifier is local and bounded by  $\epsilon'$ . After  $\alpha/(2N^2\epsilon'^2)$  repetitions of the non-fault tolerant protocol in Fig. (2) with  $\kappa = 2$  (thus 5 graphs for each repetition), the verifier can decide with confidence level  $2^{-\alpha}$ , whether or not the total variation distance between the actual  $q'(\mathbf{x})$  and the correct (honest) distribution  $q(\mathbf{x})$  of*

1. Verifier selects a random ordering of  $2\kappa + 1$  graphs, 1 computational and  $\kappa$  from each type of test graphs.
2. Verifier prepares, one by one, the qubits needed for the blind implementation of the  $2\kappa + 1$  cluster states and sends them to the prover.
3. Verifier sends the encrypted measurement angles to the prover.
4. Prover entangles all received qubits in the  $2\kappa + 1$  cluster states.
5. Prover measures all qubits simultaneously in the instructed angles and returns the results.
6. Verifier decrypts the outputs and accepts if all trap results are correct, otherwise rejects.

FIG. 2. Nonadaptive verification protocol

the Ising sampler

$$\sum_{\mathbf{x}} |q(\mathbf{x}) - q'(\mathbf{x})| \leq \epsilon \quad (3)$$

is bounded by  $\epsilon = \sqrt{8N\epsilon' + \frac{1}{5}}$ , where the sum is over all binary strings  $\mathbf{x}$  of size  $N$  that represent the output samples of all measurements in the computational graph.

The value of  $1/5$  above is a consequence of choosing  $\kappa = 2$  and comes from an explicit arithmetic calculation. While it is sufficient for our arguments, it can be reduced by allowing more rounds of testing in the protocol. A sketch of our proof appears in Appendix B 2 and a full proof in Appendix D.

Using our verifiable quantum sampler to demonstrate quantum supremacy is underwritten by results which show that approximating the Ising sampler upto constant total variation distance is hard classically, subject to the same average case hardness conjecture as in the original model [17].  $N\epsilon'$  must be constant for the approximation to be constant in Theorem 1, requiring the local error  $\epsilon'$  decrease linearly with the number of qubits. This is only realistic in quantum supremacy experiments involving few qubits.

To overcome this restriction, we next consider a fault-tolerant version of our verification protocol.

## B. Fault-tolerant verification

Ensuring  $N\epsilon'$  in Theorem 1 to be constant will get harder experimentally for increasing  $N$ . Therefore, we present a new verification scheme where the total noise scales linearly with the size, and prove that it provides a distribution that is hard to sample classically from upto constant additive error. We then prove that noise scaling with system size does not affect our strategy of proving the soundness of verifying the prover's distribution.

Quantum error correction can overcome noise and fault-tolerance ensuring scalability, but requires adaptive operations. Our new scheme allows for fault-tolerant computation, but does not apply any corrections, since they require adaptivity which is prohibited within the Ising sampler. Leveraging arguments due to Fujii [23] and his improved fault-tolerant thresholds, we prove the hardness of the scheme by relying only on conditional probabilities that post-select the null syndromes in the syndrome measurements. Our verification scheme performs actual fault-tolerance on the trap computations. Since these trap computations only involve Clifford operations, their fault tolerance can be performed nonadaptively, and corrected by classical post-processing.

An additional intricacy needs resolving. Since blindness is an ingredient in our verification scheme, its straightforward application (on the logical level) risks leaking the logical measurement angles in the distillation procedure, where many copies of the same magic state need to be sent. Also, for the distillation procedure to be effective, we need to reveal information about the state distilled. Our stratagem for circumventing this is to apply blindness on the lowest level of MBQC, on which the fault-tolerant construction is based. On the other hand, traps are applied at the logical MBQC level, since those are the qubits needing protection from noise, as outlined in Fig. (3).

The fault-tolerant verification protocol has the same structure as in Fig. (2) but based on a different cluster state. We use the lattice by Raussendorf-Harrington-Goyal (RHG) [34] and their topological fault-tolerant procedure implemented on MBQC. Time complexity of the quantum operations in the protocol is constant and the number of qubits needed is  $O(N\text{PolyLog}(N))$  [34],

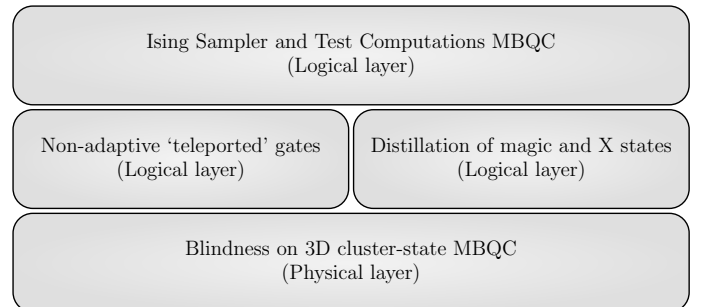


FIG. 3. Layered structure of verifiable FT computation.



the polylogarithmic overhead coming from the properties of the topological code and the use of concatenation in the distillation procedure. The full protocol and explanation is given in Appendix C.

Our proof of the classical hardness of Ising sampling in this case (Theorem 3) relies on proving soundness on the verification of conditional probabilities (Theorem 2). They are proved in Appendices G and E respectively.

**Theorem 2** (Fault-tolerant soundness). *Suppose the qubit preparation error by the verifier is local and bounded by  $\epsilon' < \epsilon_{\text{thres}}$ , the threshold. After  $\alpha/2\epsilon'^2$  repetitions of the fault-tolerant protocol with  $\kappa = 2$ , the verifier can decide with confidence level  $2^{-\alpha}$ , whether or not the total variation distance (when they are conditioned on the syndrome measurement outcome  $y$  giving the null result) between the actual distribution  $q'(\mathbf{x}|y=0)$  and the correct (honest) distribution  $q(\mathbf{x}|y=0)$  of the Ising sampler*

$$\sum_{\mathbf{x}} |q(\mathbf{x}|y=0) - q'(\mathbf{x}|y=0)| \leq \epsilon, \quad (4)$$

is bounded by  $\epsilon = \sqrt{8\epsilon'' + \frac{1}{5}}$ . Here  $\epsilon''$  is the error rate of the code and the sum is over all binary strings  $\mathbf{x}$  of size  $N$  that represent the output samples of all measurements. The threshold for the topological code running on MBQC is  $\epsilon_{\text{thres}} = 0.75\%$  [34].

Assuming the following conjecture, we prove in Appendix G the quantum supremacy theorem for the noisy case in the following theorem.

**Conjecture 1.** *Approximating  $q(\mathbf{x}|y=0)$  by*

$$|p(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)| \leq \frac{q(\mathbf{x}|y=0)}{\text{poly}(n)} + \frac{\epsilon}{2^{mn}\gamma}, \quad (5)$$

where  $2\epsilon \leq \gamma$ , is as hard as the worst case ( $\#P$ -hard) for more than  $1 - \gamma$  instances of  $\mathbf{x}$ .

**Theorem 3** (Fault-tolerant Hardness). *Assume that Conjecture 1 is true. Then weakly sampling from the distribution  $q'(\mathbf{x}, y)$  of the actual noisy Ising sampler with a classical machine, assuming the hypothesis test passes, implies a collapse in the polynomial hierarchy to the third level.*

## IV. DISCUSSIONS

We have taken the first steps towards tackling the two crucial challenges facing experimental demonstrations of quantum supremacy for the Ising sampler - verifying the output distribution of a noisy quantum sampler and proving that the sampling task remains classically hard to simulate even in the presence of experimental noise. Our work should direct near-term demonstrations of supremacy of quantum sampling and illuminate the design of long-term quantum simulators.

Our verification scheme could be further simplified by generating random universal measurement angles using

only multiples of  $\pi/4$  in each vertex of a different graph state, which we leave as an open problem, thus making the fault-tolerant translation easier. Other universal constructions with  $xy$ -plane measurements can also be considered [35].

The trap-based techniques used here could be applied in efficient implementations of the Boson Sampling model [13] in a fault-tolerant quantum computer [36]. Our methods should also apply to the random circuit IQP model [37] which, in the ‘graph program’ implementation [38] requires a smaller but non-planar resource state.

Trap-based techniques require blindness, which is not believed possible with a classical verifier [39]. Even verification protocols that do not require blindness, such as [40], still need some level of quantum encryption. This is true for any protocol based on quantum authentication schemes [41], made possible by a quantum verifier. Verification protocols with classical verifiers exist [38, 42], but require extra assumptions such as additional computational hardness conjectures or non-communicating provers respectively.

Our scheme is complementary to work on IQP verification [43], which provides a composable verification scheme based on a classical hypothesis test and verifier’s ability to prepare perfect stabilizer states. Our work allows for a simpler underlying graph due to the nonadaptive Ising sampler being a special case of the IQP model. Also, we require the verifier to prepare non-stabilizer states but, on the other hand, we allow for constant local errors. Recently, a fault-tolerant verification technique for MBQC has appeared that requires the verifier to perform measurements [44], as opposed to preparations as in our scheme.

Finally, a composable proof of security is needed to guarantee that verifiability holds in any setting where information is leaking to the environment. The goal is to show that the protocol is indistinguishable from a protocol where the prover does not receive any information from the verifier and therefore cannot use this information to his aid. An outline of this proof technique for trap-based protocols is given in [45] and a formal proof of compositability is an important next step.

## ACKNOWLEDGEMENTS

We thank Zhang Jiang, Ashely Montanaro, Michael Bremner for early discussions and Elham Kashefi, Petros Wallden and Andru Gheorghiu for discussions on the trap-based verification technique. We also thank Dominic Branford for helping with the figures. This work was supported, in part, by the UK EPSRC (EP/K04057X/2), and the UK National Quantum Technologies Programme (EP/M013243/1).

- 
- [1] I. M. Georgescu, S. Ashhab, and F. Nori, *Rev. Mod. Phys.* **86**, 153 (2014).
- [2] T. Fukuhara, A. Kantian, M. Endres, M. Cheneau, P. Schausz, S. Hild, D. Bellem, U. Schollwöck, T. Giamarchi, C. Gross, I. Bloch, and S. Kuhr, *Nat Phys* **9**, 235 (2013).
- [3] T. Lanting, A. J. Przybysz, A. Y. Smirnov, F. M. Spedalieri, M. H. Amin, A. J. Berkley, R. Harris, F. Altomare, S. Boixo, P. Bunyk, N. Dickson, C. Enderud, J. P. Hilton, E. Hoskinson, M. W. Johnson, E. Ladizinsky, N. Ladizinsky, R. Neufeld, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, S. Uchaikin, A. B. Wilson, and G. Rose, *Phys. Rev. X* **4**, 021041 (2014).
- [4] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, *Nature* **519**, 66 (2015).
- [5] R. P. Feynman, *International Journal of Theoretical Physics* **21**, 467 (1982).
- [6] I. Buluta and F. Nori, *Science* **326**, 108 (2009).
- [7] E. Ising, *Zeitschrift für Physik A Hadrons and Nuclei* **31**, 253 (1925).
- [8] K. Binder and A. P. Young, *Rev. Mod. Phys.* **58**, 801 (1986).
- [9] L. Savary and L. Balents, *Reports on Progress in Physics* **80**, 016502 (2017).
- [10] T. D. Lee and C. N. Yang, *Phys. Rev.* **87**, 410 (1952).
- [11] F. Barahona, *Journal of Physics A: Mathematical and General* **15**, 3241 (1982).
- [12] M. J. Bremner, R. Jozsa, and D. J. Shepherd, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **467**, 459 (2011).
- [13] S. Aaronson and A. Arkhipov, *Theory of Computing* **9**, 143 (2013).
- [14] K. Fujii and T. Morimae, *New Journal of Physics* **19**, 033003 (2017).
- [15] E. Farhi and A. W. Harrow, (2016), [arXiv:1602.07674](#).
- [16] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, J. M. Martinis, and H. Neven, (2016), [arXiv:1608.00263](#).
- [17] X. Gao, S.-T. Wang, and L.-M. Duan, *Physical Review Letters* **118**, 040502 (2017), [arXiv:1607.04947](#).
- [18] J. Preskill, (2012), [arXiv:1203.5813](#).
- [19] D. P. DiVincenzo, *Fortschritte der Physik* **48**, 771 (2000).
- [20] J. B. Spring, B. J. Metcalf, P. C. Humphreys, W. S. Kolthammer, X.-M. Jin, M. Barbieri, A. Datta, N. Thomas-Peter, N. K. Langford, D. Kundys, J. C. Gates, B. J. Smith, P. G. R. Smith, and I. A. Walsley, *Science (New York, N.Y.)* **339**, 798 (2013).
- [21] L. Latmiral, N. Spagnolo, and F. Sciarrino, *New Journal of Physics* **18**, 113008 (2016).
- [22] M. J. Bremner, A. Montanaro, and D. J. Shepherd, (2016), [arXiv:1610.01808](#).
- [23] K. Fujii, (2016), [arXiv:1610.03632](#).
- [24] S. Aaronson and L. Chen, (2016), [arXiv:1612.05903](#).
- [25] J. F. Fitzsimons and E. Kashefi, (2012), [arXiv:1203.5217](#).
- [26] D. Hangleiter, M. Kliesch, M. Schwarz, and J. Eisert, *Quantum Science and Technology* **2**, 015004 (2017).
- [27] A. Broadbent, J. Fitzsimons, and E. Kashefi, in *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on* (IEEE, 2009) pp. 517–526.
- [28] R. Raussendorf and H. J. Briegel, *Phys. Rev. Lett.* **86**, 5188 (2001).
- [29] L. A. Goldberg and H. Guo, (2014), [arXiv:1409.5627](#).
- [30] M. Van den Nest, W. Dür, and H. J. Briegel, *Phys. Rev. Lett.* **100**, 110501 (2008).
- [31] T. Kapourniotis, V. Dunjko, and E. Kashefi, *arXiv preprint arXiv:1506.06943* (2015).
- [32] E. Kashefi and P. Wallden, (2015), [arXiv:1510.07408](#).
- [33] A. Broadbent, (2015), [arXiv:1509.09180](#).
- [34] R. Raussendorf, J. Harrington, and K. Goyal, *New Journal of Physics* **9**, 199 (2007).
- [35] A. Mantri, T. F. Demarie, and J. F. Fitzsimons, (2016), [arXiv:1607.00758](#).
- [36] B. Peropadre, A. Aspuru-Guzik, and J. J. Garcia-Ripoll, (2015), [arXiv:1509.02703](#).
- [37] M. J. Bremner, A. Montanaro, and D. J. Shepherd, *Phys. Rev. Lett.* **117**, 080501 (2016).
- [38] D. Shepherd and M. J. Bremner, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **465**, 1413 (2009).
- [39] T. Morimae and T. Koshiba, *arXiv preprint arXiv:1407.1636* (2014).
- [40] D. Aharonov, M. Ben-Or, and E. Eban, in *Proceedings of Innovations in Computer Science 2010*, ICS2010 (2010) pp. 453–.
- [41] H. Barnum, C. Crepeau, D. Gottesman, A. Smith, and A. Tapp, in *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on* (IEEE, 2002) pp. 449–458.
- [42] B. W. Reichardt, F. Unger, and U. Vazirani, *Nature* **496**, 456 (2013).
- [43] D. Mills, A. Pappa, T. Kapourniotis, and E. Kashefi, Ongoing work (2017).
- [44] K. Fujii and M. Hayashi, (2016), [arXiv:1610.05216](#).
- [45] V. Dunjko, J. F. Fitzsimons, C. Portmann, and R. Renner, in *Advances in Cryptology—ASIACRYPT 2014* (Springer, 2014) pp. 406–425.
- [46] M. Hein, J. Eisert, and H. J. Briegel, *Phys. Rev. A* **69**, 062311 (2004).
- [47] S. Bravyi and A. Kitaev, *Phys. Rev. A* **71**, 022316 (2005).
- [48] A. G. Fowler and K. Goyal, (2008), [arXiv:0805.3202](#).
- [49] K. Fujii, *Quantum Computation with Topological Codes: from qubit to topological fault-tolerance*, Vol. 8 (Springer, 2015).
- [50] C. Dankert, R. Cleve, J. Emerson, and E. Livine, *Phys. Rev. A* **80**, 012304 (2009).
- [51] L. Stockmeyer, *SIAM Journal on Computing* **14**, 849 (1985).

## Appendix A: Bridge and break operators and the Ising Sampler

To understand the procedure of carving a specific graph out of a square lattice state by only  $xy$ -plane measurements we need to explain two types of operations originally introduced in [25, 46].

The first type is *break* operators. Let  $i$  be a vertex we want to remove from the original graph, together with the connection to its neighbours. One can achieve this by performing a Pauli  $Z$  measurement on the qubit that corresponds to  $i$  and discard the outcome. However, we do not have the power to perform measurements out of the  $xy$ -plane in our protocol (otherwise we risk revealing the position of the traps by asking the prover to change the basis). Pauli  $Z$ -measurements can be simulated by preparing  $|0\rangle$  qubits called *dummy* qubits. Then, when the prover applies  $cZ$  the operation does not have any effect in entangling it with its neighbours. The measurement can have any arbitrary rotation since the qubit is isolated and does not participate in the computation. In order to keep the whole procedure blind we instead prepare the qubit in state  $|d_i\rangle$  for  $d_i$  chosen independently and uniformly at random from  $\{0, 1\}$  (so that the state looks identical to the maximally mixed state as it is the case for the other qubits in the blind protocol). Also, we apply on each of its neighbours Pauli operation  $Z^{d_i}$ , before sending them to the prover, so that we cancel the effect that the prover's entangling will have on that neighbour.

The second type is called *bridge* operators. Let  $i$  be a vertex of degree 2 that we want to remove in the original graph and join its neighbours by a new edge. To achieve this we apply a Pauli  $Y$  measurement (measurement angle  $\pi/2$ ) on the qubit that corresponds to vertex  $i$  and add  $\pi/2$  on the angles of each the two neighbours. Conditioned on this measurement giving 0, the resulting graph when we trace out the measured qubit is the desired one. If the measurement outcome is 1 then, in order to get the correct graph, we need to apply a  $Z$  correction on the neighbours. Since our Ising sampler model is non-adaptive and all our measurements are in the  $xy$  plane we can achieve this by flipping the measurement outcomes of the corresponding qubits. Notice that this is not an issue in the test rounds that we explain in Appendix B since there are no bridge operations in this case.

## Appendix B: Non-fault-tolerant verification protocol

Here follows the full description of the protocol introduced in Section III A:

### Protocol 1:

1. Verifier selects a random ordering of  $2\kappa + 1$  graphs, one for computation and  $2\kappa$  for testing' as in Fig. (1). This fixes the position of computational basis qubits called the *dummy* qubits (see Appendix A) and the measurement angles  $\{\phi_i\}_{i=1}^N$ ,

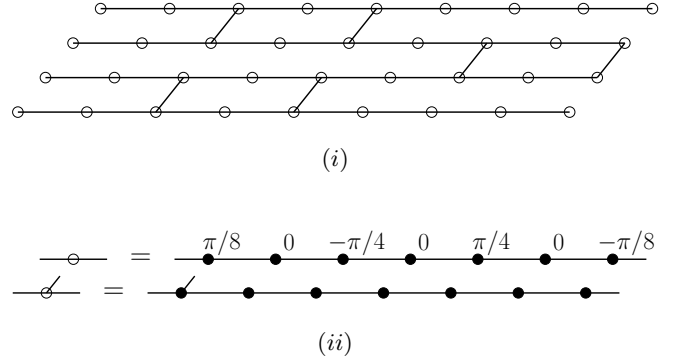


FIG. 4. The original brickwork state (i) is a universal resource for MBQC under  $xy$ -plane measurements, where white vertices represent qubits and the edges represent  $cZ$  operations. The 'extended' brickwork state (ii) is used in the original Ising sampler [17], where each white vertex is replaced by 7 physical qubits (black vertices). The measurement angle for each qubit is fixed to the value written above each vertex. There is no adaptation of the angles on previous measurement outcomes as in normal MBQC.

where  $N = m \times n$  is the total number of qubits, so that:

- (a) in the computational graph we carve from the square lattice a universal resource state, the 'extended' brickwork state of Figure (4) and fix the rest of the measurement angles according to the Ising sampler model.
  - (b) in the test graphs the dummy qubits are used to isolate the traps, which are placed at fixed positions: for half of the graphs in positions with odd parity that correspond to non-dummy qubits in the computational graph and in the other half in positions with even parity that correspond to non-dummies in the computational graph. The traps are measured with angles  $\phi = 0$  so that the measurement is deterministic. Crucially, the trap graphs do not contain any 'bridge' operations so there is no need for adaptive corrections.
2. Verifier prepares the qubits that will compose the cluster state one by one and sends them to the prover.
    - (a) The dummy qubits are prepared in  $\{|d_i\rangle : d_i \in_R \{0, 1\}\}$ .
    - (b) The rest of the qubits are prepared in  $Z^{d_{k \sim j}} |+\theta_j\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i(\theta_j + d_{k \sim j}\pi)} |1\rangle)$ , where  $\theta_j$  is chosen uniformly at random from the set  $A = \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \dots, \frac{15\pi}{8}\}$  and  $d_{k \sim j}$  is the parity of the  $d$ 's of all neighbours of  $j$ . Notice that the set  $A$  is different from the original trap-based protocol of Ref. [25].
  3. Verifier sends the encrypted computational measurement angles to the prover:  $\delta_i = \theta_i + (-1)^{r'_i} \phi_i + r_i \pi$  for  $r_i, r'_i \in_R \{0, 1\}$ . Parameters  $r_i, r'_i$  create a classical one-time pad on measurement outputs.
  4. Prover entangles all received qubits according to

- the  $2\kappa + 1$  cluster states, each of dimension  $n \times m$ , by applying  $cZ$  gates for every edge of each cluster.
5. Prover measures all qubits simultaneously in angles  $\delta_i$  and returns the measurement results to the verifier.
  6. The verifier applies a bit flip to the output bit  $i$  when  $r_i = 1$  and to its (non-dummy) neighbours when  $r'_i = 1$  to undo the classical one time pad. The output string of the measurements of the computation round is the output of the protocol. The verifier sets an extra bit to accept if all the traps give the correct result (decoded measurement result 0).

A variation of the protocol will be to ask to the prover to entangle all the graphs, instead of the square lattice state, directly in the ‘extended brickwork state’. This will leak no extra more information to the prover from what is publicly known. However, one may want to have a more generic prover and keep the protocol as presented.

Now we count the resources required in more detail. The number of qubits prepared by the verifier and sent to the prover one-by-one is  $(2\kappa + 1)N$  where  $N$  is the original size of the computation. The classical information exchanged is linear on  $N$  and can be sent in one round. Similarly the classical outcomes of the measurements can be sent at once. The prover’s requirement is to entangle all neighbouring qubits in a square lattice and apply single qubit measurements in the  $xy$ -plane.

### 1. Correctness

We have to show that the protocol performs the correct operation if the prover honestly follows the prescribed steps. We begin with a circuit diagram of the operations on the prover’s side in Fig. (5). Any measurement by angle  $\{\delta_i\}$  for the prover is here mathematically decomposed into a  $z$ -rotation ( $R_z$ ) controlled by  $\delta_i$  and a Pauli  $X$  measurement. Without loss of generality, since everything before the measurements is unitary we can assume that even a dishonest prover will apply the correct unitary operators and then chose his deviation  $U_B$  on all systems, including his private qubits  $|0\rangle^{\otimes |B|}$ . Since we are proving correctness in this section we assume  $U_B = I$ . Also the measurement angles  $\delta_i$  received by the prover are here represented as (computational basis) multi-qubit states  $|\delta_i\rangle$ .

This circuit can be simplified in a number of ways, resulting in the circuit of Fig. (6). The  $cZ$  gates between the dummy qubits and their neighbours cancel the Pauli  $Z$  pre-rotation on the neighbours. Also, we can write explicitly the rotation angles on each of the controlled  $R_z$  gates and remove the control lines.

Further simplification is done when the  $z$ -rotations by  $\theta_i$  which are part of the  $R_z$  gates and the  $z$ -rotations by  $\theta_i$  applied by the verifier to the qubits before sending them to the prover mutually cancel after commuting with the  $cZ$  gates. Notice that the dummy qubits are

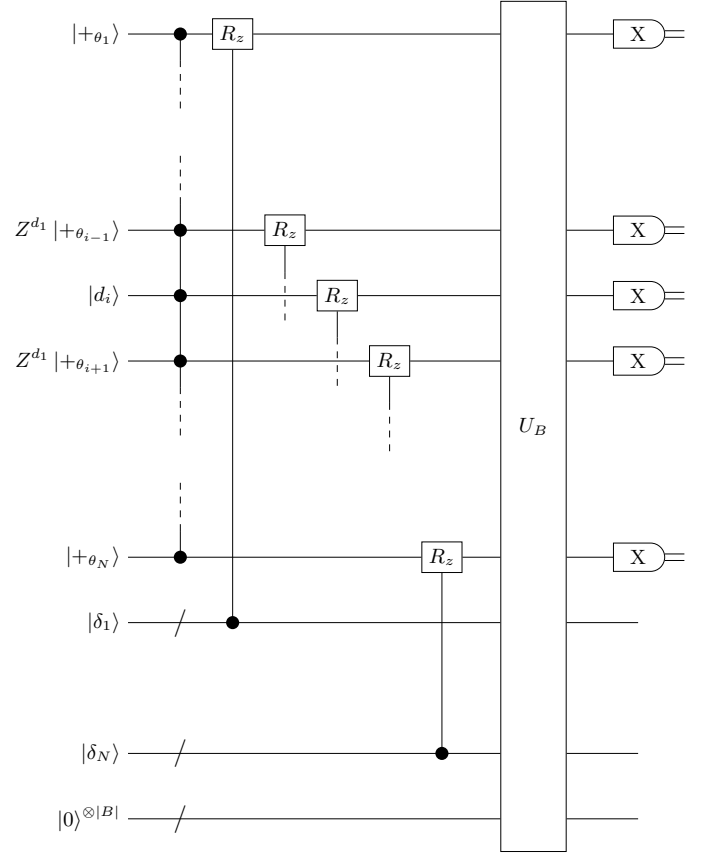


FIG. 5. Schematic representation of the prover’s operation for any arbitrary unitary deviation  $U_B$  on prover’s system. When prover is honest  $U_B = I$ .

an exception since  $\theta_i$  rotations remain but have no effect other than a global phase. Moreover, we can extract the Pauli operators from the  $R_z$  by applying identities:  $R_z(-\chi) = X R_z(\chi) X$  and  $R_z(\chi + \pi) = Z R_z(\chi)$ . The Pauli  $X$  operators from the left hand side of  $R_z$  can be rewritten as  $Z$  rotations on their entangled neighbours. This results in the circuit diagram depicted in Fig. (7). Notice that the remaining Pauli  $X$  operators do not have any effect on the Pauli  $X$  measurements (we remind the reader that in this proof  $U_B = I$ ) and the Pauli  $Z$  operators flip the measurement results.

Let us denote all the measurement outcomes of the protocol except the dummy qubit measurements by the binary vector  $\mathbf{x}$  and  $q(\mathbf{x})$  the probability of obtaining it. Let  $p(\mathbf{x})$  denote the probability of obtaining  $\mathbf{x}$  in an ideal MBQC implementation using the same measurement pattern  $\{\phi\}_i^N$  as input. The only difference between the actual and the ideal case are the Pauli  $Z$  operators before the measurements, which flip the outcomes. Therefore, by relabelling the probabilities  $q(\mathbf{x})$  to  $q(\mathbf{x}')$ , where  $x'_i = x_i \oplus r_i \oplus \sum_j r'_{j \sim i}$ , we get  $p(\mathbf{x}) = q(\mathbf{x}')$ . In other words, we can sample from the correct distribution by simply correcting the bit flips caused by the random Pauli  $Z$ , which are known to the verifier.

In MBQC, we can also write the distribution in terms



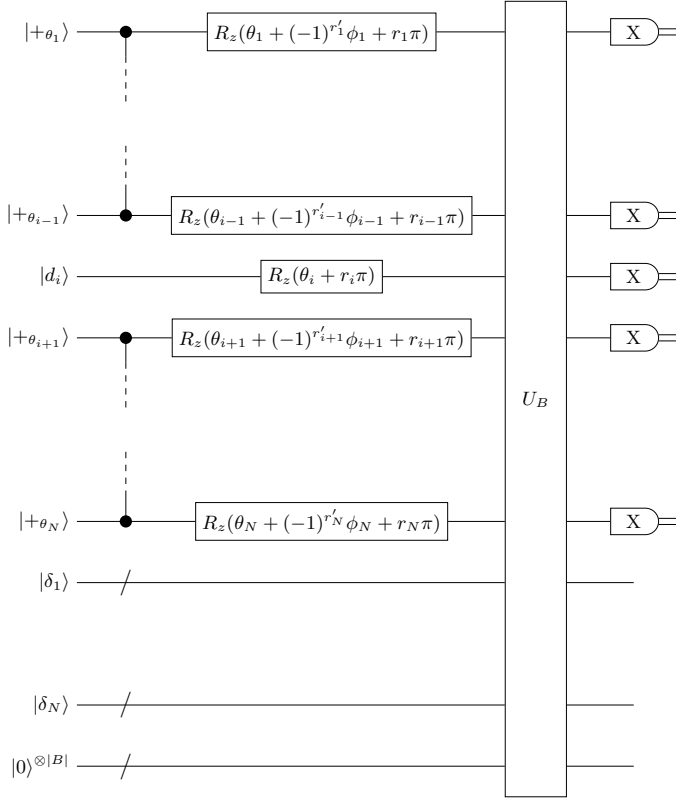


FIG. 6. Dummy qubits have the effect of isolating their neighbours and cancelling their pre-rotations. Also, unitary rotations are written explicitly.

of unitaries, labelled by the measurement outcomes (the so-called branches of the MBQC computation) of all the layers except the last. For dimension  $m \times n$ , we have (up to global phases)

$$p(\mathbf{x}) = \frac{|\langle x_{(n-1)m+1}, \dots, x_{nm} | U_{x_1, \dots, x_{(n-1)m}} | +1, \dots, +m \rangle|^2}{2^{(n-1)m}}, \quad (\text{B1})$$

since all the computational branches  $(x_1, \dots, x_{(n-1)m})$  are equiprobable and they define a unitary operation on the input.

## 2. Soundness

The proof of Theorem 1 is based on the fact that the acceptance rate of the protocol is indeed a good estimator for the average fidelity of the computational output. Thus, looking at the acceptance rate gives us with high confidence a lower bound on the fidelity, or similarly an upper bound on the trace distance between the actual and the ideal computation.

We outline the main arguments employed to prove this theorem in stages here, and provide the explicit algebraic derivations in Appendix (D).

Firstly, a unitary deviation  $U_B$  applied before the measurements, depicted in Fig. (7), captures in all generality

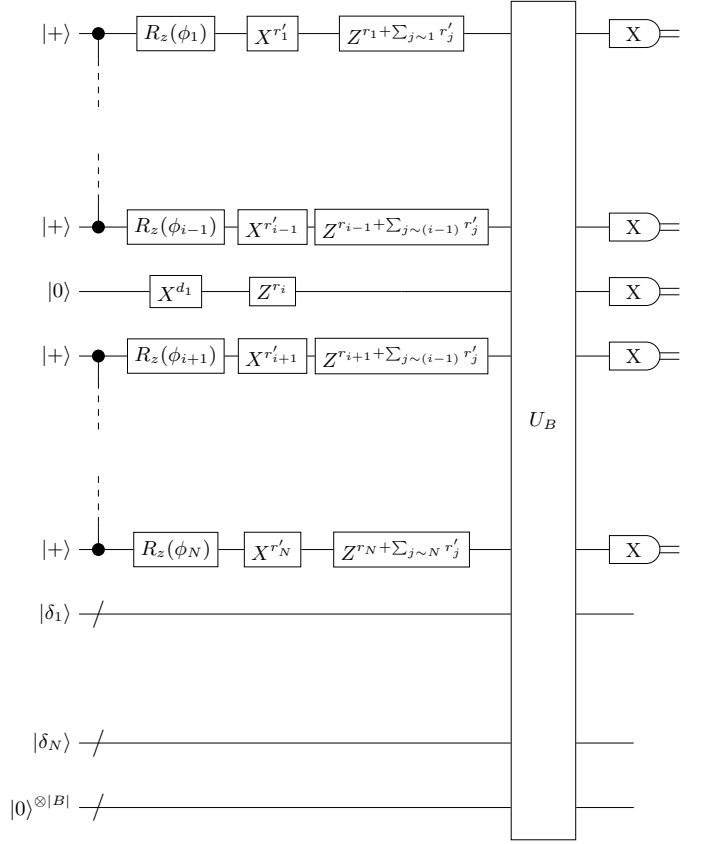


FIG. 7. Each  $z$ -rotation by  $\theta_i$  commutes with the entanglement operators and undoes the pre-rotation of the qubits participating in the graph. Also, Pauli operators are written separately.

the prover's dishonesty. To see this, consider the case when the prover performs measurements different from the honest ones. This corresponds to applying a unitary basis rotation followed by Pauli  $X$  measurements. Then,  $U_B$  applies also on the prover's private subsystem so he can use this power to replace the qubits he receives with any other qubits he chooses to prepare privately. In any case, he has to report some classical measurement results so we always keep the final Pauli  $X$  measurements in the picture. Our proof should therefore apply to any choice of  $U_B$ .

Secondly, we bound the total variation distance of the output distribution via the trace distance  $D(\rho_c, \rho'_c)$ , where  $\rho_c$  represents the state of the computational system just *after* the Pauli  $X$  measurements if the prover is honest and  $\rho'_c$  the same state if the prover is dishonest. Thus,

$$\begin{aligned} \sum_{\mathbf{x}} |q(\mathbf{x}) - q'(\mathbf{x})| &\leq D(\rho_c, \rho'_c) \leq \sqrt{1 - F^2(\rho_c, \rho'_c)} \\ &= \sqrt{1 - \text{Tr}^2(\rho_c \rho'_c)} \quad (\text{B2}) \end{aligned}$$

The main idea leading to the statement of the theorem is that the fidelities of the computational system  $F(\rho_c, \rho'_c)$  and the trap system  $F(\rho_t, \rho'_t)$  are close in absolute dif-

ference, when averaged over the random parameters and considering the preparation noise. Therefore, by estimating the fidelity of the trap system (by counting the number of acceptances over many repetitions of the protocol), we get a good estimate of the fidelity of the computational system and therefore of the total variation distance between the honest and the dishonest distribution. We begin our analysis for the case of perfect preparations and will incorporate the effect of noise after.

Averaged over the random parameters, the fidelity of the trap system just after the trap measurements, which is equal to the probability of getting all trap outcomes 1 is

$$\bar{F}_t \equiv \sum_{\mathbf{r}, \boldsymbol{\theta}, \mathbf{t}} p(\mathbf{r}, \boldsymbol{\theta}, \mathbf{t}) F(\rho_t, \rho'_t) = \sum_{\mathbf{r}, \boldsymbol{\theta}, \mathbf{t}} p(\mathbf{r}, \boldsymbol{\theta}, \mathbf{t}) \text{Tr}(\rho_t \rho'_t),$$

where  $\rho_t$  and  $\rho'_t$  represent the honest and dishonest state of the trap system just after the Pauli  $X$  measurements. Tracing over the prover's private system and dummy system, and summing over the random variables  $r_i, r'_i, d_i$  and  $\theta_i$ , we get (Appendix D)

$$\bar{F}_t = \sum_{\mathbf{t}} p(\mathbf{t}) \sum_k |a_k|^2 \prod_{i \in \mathbf{t}} |\langle +|_i P_{k|i} |+\rangle_i|^2, \quad (\text{B3})$$

where  $\mathbf{t}$  is the vector of the indices of the positions of the traps in all  $2\kappa$  test systems and  $\sum_{\mathbf{t}}$  takes all possible values allowed by the construction with equal probability  $p(\mathbf{t})$ . The summation over the random parameters results in the attack on the trap system to be transformed into a convex combination of Pauli operators  $P_k$ , each with probability  $|a_k|^2$ . By  $P_{k|i}$  we represent the Pauli operator that applies on qubit  $i$ .

A similar calculation for the average fidelity  $\bar{F}_c$  of the computational system leads to

$$\bar{F}_c = \sum_{\mathbf{t}} p(\mathbf{t}) \sum_k |a_k|^2 \prod_{i \in \mathbf{c}(\mathbf{t})} |\langle +|_i P_{k|i} |+\rangle_i|^2. \quad (\text{B4})$$

where  $\mathbf{c}(\mathbf{t})$  denotes the positions of the qubits that participate in the computation and depends on the random ordering of the  $2\kappa + 1$  rounds and therefore is a function of the position of the traps.

Choosing  $\kappa = 2$  we obtain (Appendix D)

$$|\bar{F}_t - \bar{F}_c| \leq \frac{1}{10}. \quad (\text{B5})$$

If the initial preparations are erroneous, the initial trap and computational fidelities are no more than  $N\epsilon'$  apart from their perfect counterparts. As all the prover's operations are contractive, these distances are not increased and  $|\bar{F}_t - \bar{F}_c| \leq \frac{1}{10} + 2N\epsilon'$ . By averaging over  $\alpha$  repetitions of the protocol, Hoeffding's inequality leads to high confidence in the estimation of the average trap fidelity  $\bar{F}_t$ , which leads to the statement of the theorem. See Appendix D for details.

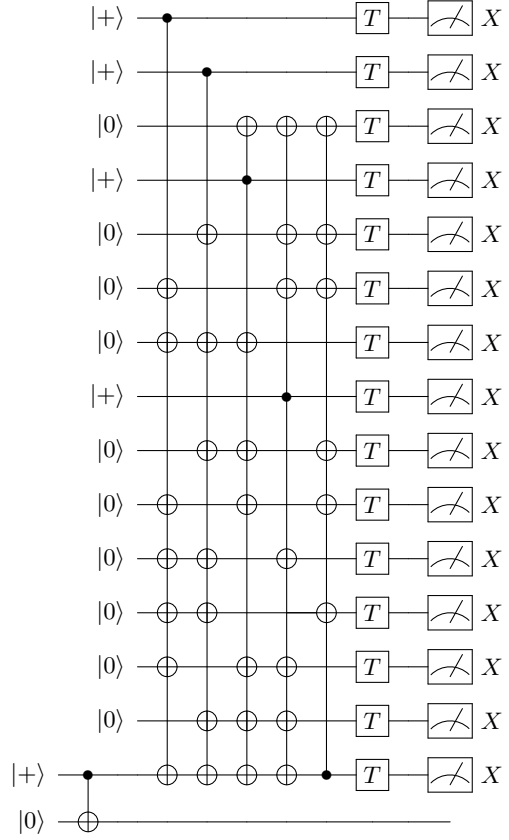


FIG. 8. Distillation step [47]. A logical  $|+\rangle$  is produced using the quantum  $(15, 1, 3)$  Reed-Muller code. Then a transversal  $T$  gate applied using the technique of Figure (10), stabilizer measurements and teleportation to an auxiliary qubit gives a 'cleaner' magic state (up to Pauli  $Z$  correction on the teleported qubit) when the error syndromes are correct. Everything is topologically protected. Picture adapted from Ref. [14].

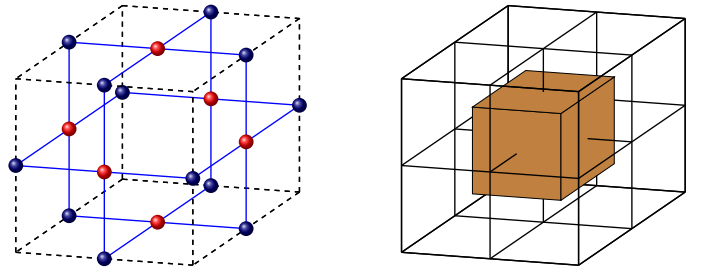


FIG. 9. 3D cluster state used in the RHG code using MBQC. Blue dots are the qubits that represent the primal cubic lattice edges (or equivalently the dual cubic lattice faces) and red dots are the qubits that represent the primal cubic lattice faces (or equivalently the dual cubic lattice edges). Entangling operations ( $cZ$ ) are represented by blue lines. On the right hand side you can see the primal and dual cubes. Pictures adapted from Refs. [34, 48].

### Appendix C: Fault-tolerant verification protocol

The fault-tolerant computation scheme used is the one proposed by Raussendorf, Harrington and Goyal (for short here the RHG code) [34]. Single qubit preparation/distillation, entangling gates ( $cX$ ) and Pauli  $X$  measurements are topologically protected using the three dimensional lattice shown in Fig. (9) and measurement-based implementation of the topological operations (more details on the MBQC implementation of the RHG code see [48], [49]). Universality comes from topologically protected concatenated distillation (Fig. (8)) of magic states:

$$|Y\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle), |A\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{4}}|1\rangle)$$

which are generated by single physical qubit measurements. Using the logical distilled magic states and the gate teleportation model one can implement a universal set of gates (Fig. 10). One can simulate an MBQC computation by using these gates and Pauli measurements and consequently add a forth dimension to the system, which comes from the flow of the logical MBQC operations (notice that this layer is distinct from the physical MBQC layer on which the topological code is implemented).

Our fault-tolerant protocol follows the protocol for the non-fault-tolerant verification of quantum supremacy, introduced in Sec. (III B).

#### Protocol 2:

1. Verifier selects a random ordering of  $2\kappa + 1$  sufficiently large 3D graphs of Fig. (9), one for computation and  $2\kappa$  for testing. In the computation run the logical computation is the same as the one in the non-fault-tolerant protocol (see Fig. 4). In the test rounds, the logical graph contains isolated traps in the same configuration as in the non-fault-tolerant version (see Fig. 1). The logical MBQC uses distilled logical  $|+\phi\rangle$  states as described before (Fig. 10) followed by Pauli  $X$  measurements to implement the  $|\pm\phi\rangle$  measurements. In order to keep the same structure in the physical operations between the computational and the test runs - so that the physical attacks are oblivious to the difference - even the  $|\pm\rangle$  deterministic measurements of the traps use logical states that go through the distillation and teleportation procedure (Fig. 10 (iii)). In the case of  $|+\rangle$  ‘distillation’ we also have the power to apply correction on the distillation step. Also, the logical Pauli  $Z$  measurements of the test graphs are implemented using *logical dummies* ( $|0\rangle$ ). For the defect regions no qubits are sent to the prover (this is equivalent to making Pauli  $Z$  measurements and there is not need of hiding since the structure of the defects is the same in each of the  $2\kappa + 1$  graphs).
2. Verifier prepares, one by one, the pre-rotated physical qubits  $|+\theta\rangle$ ,  $\theta \in_R \{0, \frac{\pi}{4}, \frac{2\pi}{4}, \dots, \frac{7\pi}{4}\}$ , needed for

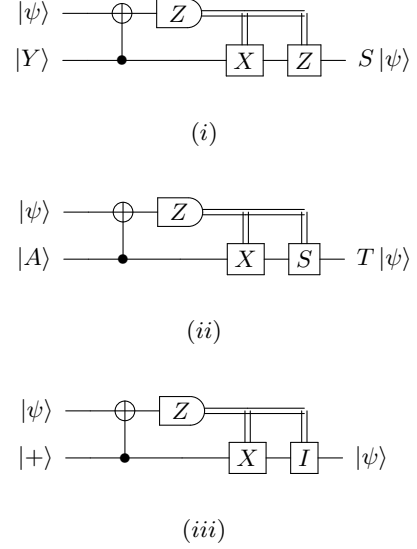


FIG. 10. (i), (ii) Gate teleportation (up to global phase) using magic states  $|Y\rangle, |A\rangle$ , (iii) State teleportation using auxiliary state  $|+\rangle$  that mimics gate teleportation (via blindness). These operations are applied in a topologically protected way, both during state distillation using ‘impure’ states and to implement the corresponding logical operations during computation using distilled states.

the blind implementation of the topological protected computation on the three dimensional cluster states and sends them to the prover. Blindness, induced by the random rotations, hides from the server the physical operations applied and therefore the logical structure of the computation in the topologically protected (vacuum) and isolated qubit region. In particular, the prover is not able to distinguish between implementing distillation and teleportation of a magic state or a  $|+\rangle$  state used for computation and testing respectively.

3. Verifier sends all the encrypted measurement angles  $\delta_i = \theta_i + (-1)^{r'_i} \phi_i + r_i \pi$  for  $r_i, r'_i \in_R \{0, 1\}$ . Parameters  $r_i, r'_i$  are classical one-time pads for the measurement outputs.

Since the model is nonadaptive (instantaneous) we do not allow for any dependencies on the measurement angles  $\phi_i$  and measurement outcomes. This means two distinct things. First, we do not implement the  $S$  correction after any  $|A\rangle$  state gate teleportation in the computational run. Note that this is not a problem for the other two cases of teleportation of Fig. 10 since the verifier can just keep track of the Pauli frame. Second, it means that we always fix which magic states we keep for the next level of distillation independently of the syndrome measurements outcomes. These elements lead to a notion of fault tolerance different and weaker than the usual one in quantum computation, but is enough to get quantum supremacy

under post-selection type arguments. For the trap runs, however, we do not have these two problems (we have only  $|+\rangle$  ‘distillation’ with error correction) and therefore verification is fault-tolerant. All other topological corrections are implemented by the verifier keeping track of the Pauli frame.

4. Prover runs the computation, by entangling, measuring all at once and returning the results.
5. The verifier classically corrects the returned outcomes using the correction procedure of the quantum error correcting code used in distillation and the topological code and undoes the  $r, r'$  pad.

6. Verifier accepts if all the results of the logical trap computations are correct, otherwise rejects.

Correctness of the protocol comes by applying the same analysis as in the non-fault-tolerant case. We can eliminate the pre-rotations by the  $\theta$ 's of the computation in the low level MBQC due to  $\theta$  being in  $\delta$  and, then, the computation is the correct up to Pauli  $Z$  corrections before the measurements. Local noise in the trap runs is taken care of by the error correction if is lower than the threshold of the RHG code. Thus we avoid scaling issues that we had in the non-fault tolerant protocol. The proof of soundness is given in Appendix E.

#### Appendix D: Proof of Theorem 1

*Proof.* Let  $U_P(\mathbf{r}, \mathbf{d})$  denote the correct unitary operation of the protocol. It includes everything preceeding  $U_B$  in Figure (7), and we have only included in the arguments the random parameters that will be averaged over later. In the honest case, after  $U_P(\mathbf{r}, \mathbf{d})$  is applied, the state of the trap system becomes  $\rho_t = \bigotimes_{i \in \mathbf{t}} Z^{r_i} |+\rangle_i \langle +|_i Z^{r_i}$ , where the index  $i$  takes values from the elements of  $\mathbf{t}$  that represent the positions of the traps. In the dishonest case (again based on Figure 7), tracing out the prover's private system, the deviation  $U_B$  becomes an arbitrary CPTP map denoted by  $\mathcal{E}$ . The fidelity of the trap system  $\rho'_t$ , right after the measurements, can be written as:

$$\bar{F}_t \equiv \sum_{\mathbf{r}, \boldsymbol{\theta}, \mathbf{t}, \mathbf{d}} p(\mathbf{r}, \boldsymbol{\theta}, \mathbf{t}, \mathbf{d}) \text{Tr} \left( \bigotimes_{i \in \mathbf{t}} Z^{r_i} |+\rangle_i \langle +|_i Z^{r_i} \rho'_t \right) \quad (\text{D1})$$

$$= \sum_{\mathbf{r}, \boldsymbol{\theta}, \mathbf{t}, \mathbf{d}, \mathbf{b}} p \text{Tr} \left( \bigotimes_{i \in \mathbf{t}} Z^{r_i} |+\rangle_i \langle +|_i Z^{r_i} \text{Tr}_{\{i: i \notin \mathbf{t}\}} (Z^{b_i} |+\rangle \langle +| Z^{b_i} \mathcal{E}(U_P(\mathbf{r}, \mathbf{d}) \bigotimes_i |+\rangle_i \langle +|_i \bigotimes_i |0\rangle_i \langle 0|_i U_P(\mathbf{r}, \mathbf{d})^\dagger \right. \\ \left. \bigotimes_i |\delta_i(\theta_i, r_i)\rangle \langle \delta_i(\theta_i, r_i)|) Z^{b_i} |+\rangle \langle +| Z^{b_i} \right) \quad (\text{D2})$$

where the second trace in the formula is taken over all the systems except the trap system.

Summing over  $\theta$ 's creates the maximally mixed state for the  $\delta$ 's and summing over the  $r$ 's and  $d$ 's of the computational system and the dummy system creates the maximally mixed state for those systems and therefore we can trace them out and update the CPTP map  $\mathcal{E}$  to a new CPTP map  $\mathcal{E}'$  that applies on the remaining system (of dimension  $2^{N'}$ ) and does not depend on the secret parameters. The CPTP map  $\mathcal{E}'$  can be decomposed in the Pauli basis. Also, we remind the reader that the  $\phi$  parameters of the trap qubits are all zero and therefore the remaining honest operation consists only of the rotations by the  $r$  parameters.

$$\bar{F}_t = \sum_{\mathbf{r}_t, \mathbf{t}, \mathbf{b}_t} p(\mathbf{r}_t, \mathbf{t}) \text{Tr} \left( \sum_{k=1}^{4^{N'}} \sum_{l=1}^{4^{N'}} a_k a_l^* \bigotimes_{i \in \mathbf{t}} Z^{r_i} |+\rangle_i \langle +|_i Z^{r_i} \bigotimes_{i \in \mathbf{t}} Z^{b_i} |+\rangle \langle +| Z^{b_i} P_{k|i} Z^{r_i} |+\rangle_i \langle +|_i Z^{r_i} P_{l|i} Z^{b_i} |+\rangle \langle +| Z^{b_i} \right) \quad (\text{D3})$$

where  $P_{k|i}$  denotes the Pauli operator that applies on qubit with index  $i$  from the generalized Pauli basis operator  $P_k$ . Because of the structure of the state we can add for free some Pauli  $X$  operators randomized by new parameters  $r'$  taken uniform at random.

$$\bar{F}_t = \sum_{\mathbf{r}_t, \mathbf{r}'_t, \mathbf{t}, \mathbf{b}_t} p(\mathbf{r}_t, \mathbf{r}'_t, \mathbf{t}) \text{Tr} \left( \sum_{k,l} a_k a_l^* \bigotimes_{i \in \mathbf{t}} Z^{r_i} |+\rangle_i \langle +|_i Z^{r_i} \bigotimes_{i \in \mathbf{t}} Z^{b_i} |+\rangle \langle +| X^{r'_i} Z^{b_i} P_{k|i} Z^{r_i} X^{r'_i} |+\rangle_i \langle +|_i X^{r'_i} Z^{r_i} P_{l|i} Z^{b_i} |+\rangle \langle +| Z^{b_i} \right) \quad (\text{D4})$$

By changing variables  $b'_i = b_i + r_i$  and applying the cyclic property of the trace to move  $Z^{r_i} X^{r'_i}$  around:

$$\bar{F}_t = \sum_{\mathbf{r}_t, \mathbf{r}'_t, \mathbf{t}, \mathbf{b}'_t} p(\mathbf{r}_t, \mathbf{r}'_t, \mathbf{t}) \text{Tr} \left( \sum_{k,l} a_k a_l^* \bigotimes_{i \in \mathbf{t}} |+\rangle_i \langle +|_i \bigotimes_{i \in \mathbf{t}} Z^{b'_i} |+\rangle \langle +| X^{r'_i} Z^{b'_i + r_i} P_{k|i} Z^{r_i} X^{r'_i} |+\rangle_i \langle +|_i X^{r'_i} Z^{r_i} P_{l|i} Z^{b'_i + r_i} X^{r'_i} |+\rangle \langle +| Z^{b'_i} \right) \quad (\text{D5})$$

By applying the Pauli twirl lemma [50] (proven in Appendix (F)) when averaging over  $\mathbf{r}_t, \mathbf{r}'_t$ , we get



$$\begin{aligned}
\bar{F}_t &= \sum_{\mathbf{t}, \mathbf{b}'_t} p(\mathbf{t}) \sum_k |a_k|^2 \prod_{i \in \mathbf{t}} |\langle + |_i Z^{b'_i} | + \rangle \langle + | Z^{b'_i} P_{k|i} | + \rangle_i|^2 \\
&= \sum_{\mathbf{t}} p(\mathbf{t}) \sum_k |a_k|^2 \prod_{i \in \mathbf{t}} |\langle + |_i P_{k|i} | + \rangle_i|^2,
\end{aligned} \tag{D6}$$

A similar analysis is applied to calculate the average fidelity  $\bar{F}_c = F(\rho_c, \rho'_c)$  of the computational state after the measurements. In the honest case the computational state  $\rho_c$  just before the measurement will be disentangled from the rest of the system:  $\bigotimes_{i \in \mathbf{c}} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle \langle G| \bigotimes_{j \in \mathbf{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j}$ , where  $\mathbf{c}(\mathbf{t})$  are the positions of the computational qubits for a choice of trap positions  $\mathbf{t}$  and  $|G\rangle \langle G|$  is the computational graph state ( $E_G \bigotimes_{i \in \mathbf{c}} |+\rangle_i \langle + |_i E_G^\dagger$ ). In the dishonest case, for an attack  $\mathcal{E}$  the fidelity  $\bar{F}_c$  is

$$\begin{aligned}
\bar{F}_c &\equiv \sum_{\mathbf{r}, \mathbf{r}', \boldsymbol{\theta}, \mathbf{t}, \mathbf{d}} p(\mathbf{r}, \mathbf{r}', \boldsymbol{\theta}, \mathbf{t}, \mathbf{d}) F(\rho_c, \rho'_c) = \sum_{\mathbf{r}, \mathbf{r}', \boldsymbol{\theta}, \mathbf{t}, \mathbf{d}} p(\mathbf{r}, \mathbf{r}', \boldsymbol{\theta}, \mathbf{t}, \mathbf{d}) \text{Tr}(\rho_c \rho'_c) \\
&= \sum_{\mathbf{r}, \mathbf{r}', \boldsymbol{\theta}, \mathbf{t}, \mathbf{d}, \mathbf{b}} p \text{Tr} \left( \bigotimes_{i \in \mathbf{c}} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle \langle G| \bigotimes_{j \in \mathbf{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j} \text{Tr}_{\{i: i \notin \mathbf{c}\}} (Z^{b_i} |+\rangle \langle + | Z^{b_i} \mathcal{E}(U_P(\mathbf{r}, \mathbf{r}', \mathbf{d})) \right. \\
&\quad \left. \bigotimes_i |+\rangle_i \langle + |_i \bigotimes_i |0\rangle_i \langle 0|_i U_P(\mathbf{r}, \mathbf{r}', \mathbf{d})^\dagger \bigotimes_i |\delta_i(\theta_i, r_i, r'_i)\rangle \langle \delta_i(\theta_i, r_i, r'_i)| Z^{b_i} |+\rangle \langle + | Z^{b_i}) \right).
\end{aligned} \tag{D7}$$

Summing over the  $\theta$ 's of the  $\delta$ 's and the  $r$ 's and the  $d$ 's of the trap and the dummy system creates the maximally mixed system for these systems which can be traced over. Expressing the attack on the remaining system (of dimension  $2^{N'}$ ) in the Pauli basis, we get

$$\begin{aligned}
\bar{F}_c &= \sum_{\mathbf{t}, \mathbf{r}_{\mathbf{c}(\mathbf{t})}, \mathbf{r}'_{\mathbf{c}(\mathbf{t})}, \mathbf{b}_{\mathbf{c}(\mathbf{t})}} p \text{Tr} \left( \sum_{k=1}^{4^{N'}} \sum_{l=1}^{4^{N'}} a_k a_l^* \bigotimes_{i \in \mathbf{c}} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle \langle G| \bigotimes_{j \in \mathbf{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j} \bigotimes_{i \in \mathbf{c}} Z^{b_i} |+\rangle \langle + | Z^{b_i} \right. \\
&\quad \left. P_{k|i} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle \langle G| \bigotimes_{j \in \mathbf{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j} P_{l|j} Z^{b_i} |+\rangle \langle + | Z^{b_i} \right)
\end{aligned} \tag{D8}$$

By changing variables  $b' = b + r$  and applying the cyclic property of the trace

$$\begin{aligned}
\bar{F}_c &= \sum_{\mathbf{t}, \mathbf{r}_{\mathbf{c}(\mathbf{t})}, \mathbf{r}'_{\mathbf{c}(\mathbf{t})}, \mathbf{b}'_{\mathbf{c}(\mathbf{t})}} p \text{Tr} \left( \sum_{k=1}^{4^{N'}} \sum_{l=1}^{4^{N'}} a_k a_l^* \bigotimes_{i \in \mathbf{c}} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle \langle G| \bigotimes_{j \in \mathbf{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} \bigotimes_{i \in \mathbf{c}} Z^{b'_i} |+\rangle \langle + | Z^{b'_i + r_i} \right. \\
&\quad \left. P_{k|i} Z^{r_i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle \langle G| \bigotimes_{j \in \mathbf{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} Z^{r_j} P_{l|j} Z^{b'_i + r_i} |+\rangle \langle + | Z^{b'_i} \right)
\end{aligned} \tag{D9}$$

Using Corollary 1 in Appendix (F) and the cyclic property of the trace and sum over  $\mathbf{r}_{\mathbf{c}(\mathbf{t})}$ , we can eliminate all Pauli  $X$  operators of the attack that differ in the two sides (we denote this by replacing the summation over  $l$  with a summation over  $l_x$  where the element  $P_l$  agrees with  $P_k$  on all the Pauli  $X$  components). We also use the cyclic property of the trace to get

$$\begin{aligned}
\bar{F}_c &= \sum_{\mathbf{t}, \mathbf{r}_{\mathbf{c}(\mathbf{t})}, \mathbf{r}'_{\mathbf{c}(\mathbf{t})}, \mathbf{b}'_{\mathbf{c}(\mathbf{t})}} p \text{Tr} \left( \sum_k \sum_{l_x} a_k a_l^* \langle G | \bigotimes_{j \in \mathbf{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} \bigotimes_{i \in \mathbf{c}} Z^{b'_i} |+\rangle \langle + | Z^{b'_i} \right. \\
&\quad \left. P_{k|i} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle \langle G| \bigotimes_{j \in \mathbf{c}} X^{r'_j} R_z(-\phi_j) X^{r'_j} P_{l_x|j} Z^{b'_i} |+\rangle \langle + | Z^{b'_i} \bigotimes_{i \in \mathbf{c}} X^{r'_i} R_z(\phi_i) X^{r'_i} |G\rangle \right)
\end{aligned} \tag{D10}$$

Then, the  $X^{r'_i}$  ( $X^{r'_j}$ ) operators that are next to  $|G\rangle$  ( $\langle G|$ ) can be rewritten as Pauli  $Z$  operators on their neighbours, which then commute with  $z$  rotations and the attack (since the Pauli  $X$  operators of the attack are the same from both sides) and with the projectors  $Z^{b'_i} |+\rangle \langle + | Z^{b'_i}$ , by changing variable  $b''_i = b'_i + r'_i$ , and cancel each other. The  $X^{r'_i}$  operators that are next to projectors  $Z^{b'_i} |+\rangle \langle + | Z^{b'_i}$  commute with them trivially.

$$\begin{aligned}
\bar{F}_c &= \sum_{\mathbf{t}, \mathbf{r}'_{\mathbf{c}(\mathbf{t})}, \mathbf{b}''_{\mathbf{c}(\mathbf{t})}} p \text{Tr} \left( \sum_k \sum_{l_x} a_k a_l^* \langle G | \bigotimes_{j \in \mathbf{c}} R_z(-\phi_j) \bigotimes_{i \in \mathbf{c}} Z^{b''_i} |+\rangle \langle + | Z^{b''_i} \right. \\
&\quad \left. X^{r'_j} P_{k|i} X^{r'_i} R_z(\phi_i) |G\rangle \langle G| \bigotimes_{j \in \mathbf{c}} R_z(-\phi_j) X^{r'_j} P_{l_x|j} X^{r'_i} Z^{b''_i} |+\rangle \langle + | Z^{b''_i} \bigotimes_{i \in \mathbf{c}} R_z(\phi_i) |G\rangle \right)
\end{aligned} \tag{D11}$$

Then we can use again Corollary 1, but with  $Q, Q'$  being Pauli  $Z$ +identity operators and  $\{P_i\}$  all tensor products of Pauli  $X$ +identity operators, and sum over  $\mathbf{r}'_{c(t)}$  to eliminate the Pauli  $Z$  components of the attack the differ in the two sides. Thus the attack becomes a convex combination of Pauli operators:

$$\bar{F}_c = \sum_{\mathbf{t}, \mathbf{b}''_{c(\mathbf{t})}} p(\mathbf{t}) \sum_k |a_k|^2 \bigotimes_{i \in c(\mathbf{t})} \langle + |_i E_G^\dagger \bigotimes_{i \in c(\mathbf{t})} R_z(-\phi_i) Z^{b''_i} | + \rangle \langle + | Z^{b''_i} P_{k|i} R_z(\phi_i) E_G \bigotimes_{i \in c(\mathbf{t})} | + \rangle_i |^2 \quad (\text{D12})$$

The Pauli  $X$  component of  $P_{k|i}$  can be replaced by  $I$  since the only effect is, depending on  $b''_i$ , to change the sign of the quantity inside the absolute and the sign is eliminated. Then, we can sum over the  $b$ 's to get identity and since  $E_G^\dagger R_z(-\phi_i)$  now commutes with the attack operators, it cancels with  $R_z(\phi_i) E_G$ .

$$\bar{F}_c = \sum_{\mathbf{t}} p(\mathbf{t}) \sum_k |a_k|^2 \prod_{i \in c(\mathbf{t})} |\langle + |_i P_{k|i} | + \rangle_i|^2 \quad (\text{D13})$$

It is easy to see, by the symmetry of the trap construction that for attacks that are the same on any of the  $2\kappa + 1$  graphs Equation D6 and Equation D13 give the same result when averaged over  $\mathbf{t}$ . However, one needs to deal with more clever attacks which attack a different qubit at any graph trying to coincide with dummies instead of traps with non-zero probability, i.e. for some of the possible permutations of the  $2\kappa + 1$  graphs. For the case of  $\kappa = 2$ , we calculate the gap between the fidelities when averaged over the permutations to be  $\frac{1}{10}$ .

In the case that there is error in the preparation there is an extra  $2N\epsilon'$  distance between the fidelities. This comes from the following facts: the initial state of the system the prover receives now is  $N\epsilon'$ -close to the correct state (from subadditivity of the trace distance). The operations applied by the prover are contractive so they cannot increase the trace distance. Therefore, the final fidelity is  $N\epsilon'$  close to the one calculated above in both the case of the actual fidelity and the fidelity that comes from the traps. By simple calculation the absolute difference between the two fidelities is bounded by  $2N\epsilon' + \frac{1}{10}$ .

Our estimate for the fidelity of the trap comes  $s$  repetitions of the protocol. By Hoeffding's inequality, repeating  $s = \alpha/(2N^2\epsilon'^2)$  times gets us  $N\epsilon'$ -close in our estimation with confidence  $2^{-\alpha}$ . Let us set the limit for accepting the estimated fidelity to  $(1 - N\epsilon')$ , we can be sure with the aforementioned confidence that the actual fidelity of the computation is lower bounded by  $1 - 4N\epsilon' - \frac{1}{10}$ . This means that for the total variation distance we have an upper bound  $\epsilon = \sqrt{8N\epsilon' + \frac{1}{5}}$ .

□

## Appendix E: Proof of Theorem 2

*Proof.* To show soundness we need again to show that the fidelity on a computational round and the fidelity of a test round are the same averaged over the random choice of the ordering of computation/test. Crucially, we assume that the physical noise is symmetric and does not depend on this choice.

The total variation distance between the actual (noisy and potentially dishonest) distribution of the Ising sampler  $q'(\mathbf{x}|y=0)$ , where  $y=0$  implies conditioning on the null syndrome, and the correct (noiseless and honest) one  $q(\mathbf{x}|y=0) = q(\mathbf{x})$  after the measurements is

$$\begin{aligned} \sum_{\mathbf{x}} |q(\mathbf{x}|y=0) - q'(\mathbf{x}|y=0)| &= \sum_{\mathbf{x}} \left| q(\mathbf{x}) - \frac{q'(\mathbf{x}, y=0)}{q'(y=0)} \right| \\ &= D\left(\rho_c, \frac{\rho_c'^{\text{post}}}{q'(y=0)}\right) \\ &\leq \sqrt{1 - F^2\left(\rho_c, \frac{\rho_c'^{\text{post}}}{q'(y=0)}\right)} \\ &= \sqrt{1 - \text{Tr}^2\left(\frac{\rho_c \rho_c'^{\text{post}}}{q'(y=0)}\right)}, \end{aligned} \quad (\text{E1})$$

where  $\rho_c$  is the correct state and  $\rho_c'^{\text{post}}$  the actual state, post-selected on the null syndrome measurements, after all measurements. For the rest of this section we denote  $q'(y=0)$  as  $q'_0$  for simplicity.

For the computation round, the average fidelity  $\bar{F}_c$  is calculated in the physical level of the computation and similarly to the non-fault-tolerant case. The qubits are pre-rotated by  $\theta_i$  or flipped by  $d_i$  in the case of dummies. Since the

computation is applied in a single time-step we represent the noise as a single CPTP-map  $\mathcal{N}$  applied on the whole of the system, after the (possibly dishonest) operations and before the measurements. The noise operator is such that the noise strength  $\|\mathcal{Z} - \mathcal{N}\|_\diamond$  is smaller than the size of the system times the threshold value.

We apply the same twirling steps as in the proof of Theorem 1 to twirl the CPTP map that is the composition of the attack and the noise. Notice that the twirl on the post-selected qubits is trivial since there is no sum over  $b'_i$ . Thus,

$$\tilde{F}_c \left( \rho_c^{\text{post}}, \frac{\rho_c'^{\text{post}}}{q'_0} \right) = \frac{1}{q'_0} \sum_{\mathbf{t}, \mathbf{b}'_{c(\mathbf{t})}} p(\mathbf{t}) \sum_k |a_k|^2 \bigotimes_{i \in c(\mathbf{t})} \langle 0|_i \bigotimes_{i \in c(\mathbf{t})} \langle +|_i E_G^\dagger \bigotimes_{i \in c(\mathbf{t})} R_z(-\phi_i) Z^{b'_i} |+\rangle_i \langle +|_i Z^{b'_i} \mathcal{N}' P_{k|i} R_z(\phi_i) E_G \bigotimes_{i \in c(\mathbf{t})} |+\rangle_i \bigotimes_{i \in c(\mathbf{t})} |0\rangle_i^L \big|^2, \quad (\text{E2})$$

where  $b'_i$ 's take fixed values in the sum for the syndrome measurements such that the syndrome indicates null errors. Operator  $\mathcal{N}'$  is the noise operator after the twirl: a convex combination of Pauli operators,  $\sum_i \beta_i P_i$ , with  $\sum_i \beta_i = 1$ . The only noise and attack operators that have an effect on the above quantity are tensor products of identity and Pauli  $Z$ . These operators flip the measurement outcome of the particular qubit.

By linearity of the classical decoding procedure, the effect of the noise operator  $\mathcal{N}'$  and the attack operators  $P_{k|i}$  can be considered separately. Detectable attacks disappear because of the projector to null syndromes. The undetected attacks that come from operators  $P_{k|i}$  can be written as logical bit flips on the subsequent measurements - since it will affect the classical post-processing. Using Theorem 1 from [23], noise in this post-selected setting is reduced to exponentially small additive error  $\epsilon''$  (independent of the size of the computation), assuming that the local noise  $\epsilon'$  is below the code threshold. Also, the normalization factor vanishes when we trace over the syndrome systems.

Therefore, at the logical MBQC level

$$\bar{F}_c \approx_{\epsilon''} \sum_{\mathbf{t}, \mathbf{b}''_{c(\mathbf{t})}} p(\mathbf{t}) \sum_k |a_k|^2 \bigotimes_{i \in c(\mathbf{t})} \langle 0|_i^L \bigotimes_{i \in c(\mathbf{t})} \langle +|_i^L E_G^{L\dagger} \bigotimes_{i \in c(\mathbf{t})} R_z(-\phi_i^L) Z^{b''_{c(\mathbf{t})}} |+\rangle_i^L \langle +|_i^L Z^{b''_{c(\mathbf{t})}} P_{k|i}^L R_z(\phi_i^L) E_G^L \bigotimes_{i \in c(\mathbf{t})} |+\rangle_i^L \bigotimes_{i \in c(\mathbf{t})} |0\rangle_i^L \big|^2, \quad (\text{E3})$$

where  $\approx_{\epsilon''}$  means  $\epsilon''$ -close in absolute difference. We can now sum over the index  $b''_{c(\mathbf{t})}$  to simplify the expression, by cancelling also the rotation and entangling operators. On the logical dummy system the logical Pauli  $Z$  attacks have no effect, therefore it has trace 1 and can be simplified to

$$\bar{F}_c \approx_\epsilon \sum_{\mathbf{t}} p(\mathbf{t}) \sum_k |a_k|^2 \prod_{i \in c(\mathbf{t})} |\langle +|_i^L P_{k|i}^L |+\rangle_i^L|^2. \quad (\text{E4})$$

The same technique can be employed for the test rounds, resulting in the same argument as in the non-fault-tolerance case, but now the confidence does not depend in the size of the computation.  $\square$

## Appendix F: Channel Twirl Lemma

The following lemma is used in the verifiability proofs.

**Lemma 1.**

$$\sum_{i=1}^{4^n} P_i Q P_i \rho P_i Q' P_i = 0, \text{ if } Q \neq Q' \quad (\text{F1})$$

where  $\rho$  is a matrix of dimension  $2^n \times 2^n$ ,  $Q, Q'$  are two arbitrary  $n$ -fold tensor products of Pauli+identity operators  $\{I, X, Y, Z\}$ , and  $\{P_i\}$  is the set of all  $n$ -fold tensor products of Pauli operators and the identity  $\{I, X, Y, Z\}$ .

A proof of this lemma is also provided in Ref. [50].

*Proof.* We can write  $Q$  as  $Z_{\mathbf{a}} X_{\mathbf{a}'} = Z^{a_1} \otimes \dots \otimes Z^{a_n} X^{a'_1} \otimes \dots \otimes X^{a'_n}$ , for arbitrary binary vectors  $\mathbf{a} = (a_1, \dots, a_n)$ ,  $\mathbf{a}' = (a'_1, \dots, a'_n)$ , and similarly  $Q' = Z_{\mathbf{b}} X_{\mathbf{b}'}$ . Assuming

$Q \neq Q'$ , either  $\mathbf{a} \neq \mathbf{b}$  or  $\mathbf{a}' \neq \mathbf{b}'$ . Summing over all  $P_{\mathbf{k}, \mathbf{k}'}$ 's which are the  $n$ -fold tensor products of the form  $Z_{\mathbf{k}} X_{\mathbf{k}'} = Z^{k_1} \otimes \dots \otimes Z^{k_n} X^{k'_1} \otimes \dots \otimes X^{k'_n}$  for binary vectors  $\mathbf{k} = (k_1, \dots, k_n)$ ,  $\mathbf{k}' = (k'_1, \dots, k'_n)$ , we get

$$\begin{aligned} & \sum_{\mathbf{k}, \mathbf{k}'} P_{\mathbf{k}, \mathbf{k}'} Q P_{\mathbf{k}, \mathbf{k}'} \rho P_{\mathbf{k}, \mathbf{k}'} Q' P_{\mathbf{k}, \mathbf{k}'} \\ &= \sum_{\mathbf{k}, \mathbf{k}'} Z_{\mathbf{k}} X_{\mathbf{k}'} Z_{\mathbf{a}} X_{\mathbf{a}'} Z_{\mathbf{k}} X_{\mathbf{k}'} \rho Z_{\mathbf{k}} X_{\mathbf{k}'} Z_{\mathbf{b}} X_{\mathbf{b}'} Z_{\mathbf{k}} X_{\mathbf{k}'} \\ &= \sum_{\mathbf{k}, \mathbf{k}'} Z_{\mathbf{k}} (X_{\mathbf{k}'} Z_{\mathbf{a}} X_{\mathbf{k}'}) X_{\mathbf{a}'} Z_{\mathbf{k}} \rho Z_{\mathbf{k}} (X_{\mathbf{k}'} Z_{\mathbf{b}} X_{\mathbf{k}'}) X_{\mathbf{b}'} Z_{\mathbf{k}} \\ &= \sum_{\mathbf{k}, \mathbf{k}'} Z_{\mathbf{k}} ((-1)^{\mathbf{k}' \cdot \mathbf{a}} Z_{\mathbf{a}}) X_{\mathbf{a}'} Z_{\mathbf{k}} \rho Z_{\mathbf{k}} ((-1)^{\mathbf{k}' \cdot \mathbf{b}} Z_{\mathbf{b}}) X_{\mathbf{b}'} Z_{\mathbf{k}} \end{aligned}$$

$$\begin{aligned}
&= \sum_{\mathbf{k}, \mathbf{k}'} (-1)^{\mathbf{k}' \cdot (\mathbf{a} \oplus \mathbf{b})} Z_{\mathbf{a}}(Z_{\mathbf{k}} X_{\mathbf{a}'} Z_{\mathbf{k}}) \rho Z_{\mathbf{b}}(Z_{\mathbf{k}} X_{\mathbf{b}'} Z_{\mathbf{k}}) \\
&= \sum_{\mathbf{k}'} (-1)^{\mathbf{k}' \cdot (\mathbf{a} \oplus \mathbf{b})} \sum_{\mathbf{k}} (-1)^{\mathbf{k} \cdot (\mathbf{a}' \oplus \mathbf{b}')} Z_{\mathbf{a}} X_{\mathbf{a}'} \rho Z_{\mathbf{b}} X_{\mathbf{b}'}
\end{aligned}$$

If either  $\mathbf{a} \neq \mathbf{b}$  or  $\mathbf{a}' \neq \mathbf{b}'$  the summation  $\sum_{\mathbf{k}'} ((-1)^{\mathbf{k}' \cdot (\mathbf{a} \oplus \mathbf{b})})$  or  $\sum_{\mathbf{k}} ((-1)^{\mathbf{k} \cdot (\mathbf{a}' \oplus \mathbf{b}')} )$  is equal to zero respectively, because (in either case) exactly half of the elements of the summation will be  $-1$  and half will be  $1$ . Therefore, since our assumption was that either  $\mathbf{a} \neq \mathbf{b}$  or  $\mathbf{a}' \neq \mathbf{b}'$  or both, the whole expression equals zero.  $\square$

**Corollary 1.**

$$\sum_{i=1}^{2^n} P_i Q P_i \rho P_i Q' P_i = 0, \text{ if } Q \neq Q' \quad (\text{F2})$$

where  $\rho$  is a matrix of dimension  $2^n \times 2^n$ ,  $Q, Q'$  are two arbitrary  $n$ -fold tensor products of Pauli  $X$  and identity operators  $\{I, X\}$ , and  $\{P_i\}$  is the set of all  $n$ -fold tensor products of Pauli  $Z$  and identity operators  $\{I, Z\}$ .

*Proof.* Since  $Q \neq Q'$ , Lemma 1 gives

$$\sum_{i=1}^{4^n} P_i Q P_i \rho P_i Q' P_i = 0 \quad (\text{F3})$$

where  $\{P_i\}$  is the set of all  $n$ -fold tensor products of the Pauli operators and the identity  $\{I, X, Y, Z\}$ . But since  $Q$  and  $Q'$  have only identity and Pauli  $X$  tensor elements the Pauli  $X$  operators of  $\{P_i\}$  commute with  $Q$  and  $Q'$  on each side and give identity.  $\square$

## Appendix G: Proof of Theorem 3

Our hardness proof follows a similar line of reasoning as the proof of the original translation invariant Ising sampler [17] - proof by contradiction.

If we can classically weakly sample from  $q'(\mathbf{x}, y)$ , then estimating the probabilities of the distribution with fixed accuracy is in  $\#P$ . We can construct a polynomial time non-deterministic Turing machine that uses the sampler as an oracle and accepts when a specific event happens, so that the probability can be estimated by counting the accepting branches. We can also estimate the marginal probabilities  $q'(y)$  in such a manner. From Stockmeyer's theorem [51], there exists an  $FBPP^{NP}$  machine that can compute explicitly the values  $p(\mathbf{x}, y)$ , such that for every  $\mathbf{x}, y$

$$|p(\mathbf{x}, y) - q'(\mathbf{x}, y)| \leq \frac{q'(\mathbf{x}, y)}{\text{poly}(N)}. \quad (\text{G1})$$

And the same holds for marginals. Then, it is easy to prove that for conditional probabilities (for a different polynomial of  $N$ ),

$$|p(\mathbf{x}|y=0) - q'(\mathbf{x}|y=0)| \leq \frac{q'(\mathbf{x}|y=0)}{\text{poly}(N)}. \quad (\text{G2})$$

Applying the triangle inequality, for every  $\mathbf{x}$  the distance between the values  $p(\mathbf{x}|y=0)$  and the correct (honest) conditional probability  $q(\mathbf{x}|y=0)$  of the Ising sampler is

$$\begin{aligned}
&|p(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)| \\
&\leq |p(\mathbf{x}|y=0) - q'(\mathbf{x}|y=0)| + |q'(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)| \\
&\leq \frac{q'(\mathbf{x}|y=0)}{\text{poly}(n)} + |q'(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)|. \quad (\text{G3})
\end{aligned}$$

From Eqn. (4), assuming that the verification test has passed, it follows that the expectation of  $|q'(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)|$  is  $\leq \frac{\epsilon}{2^{mn}}$ , where  $m, n$  are the dimensions of the logical 'extended' brickwork state.

Markov inequality relates a probability with its expectation. For a random variable  $X$  and  $\gamma > 0$ ,

$$p(X \geq \gamma) \leq \frac{E(X)}{\gamma}. \quad (\text{G4})$$

Applying the Markov inequality to the second term in Eqn. (G3)

$$p(|q'(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)| \geq \gamma) \leq \frac{\epsilon}{2^{mn}\gamma}.$$

By changing variables

$$p\left(|q'(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)| \geq \frac{\epsilon}{2^{mn}\gamma}\right) \leq \gamma$$

or

$$p\left(|q'(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)| \leq \frac{\epsilon}{2^{mn}\gamma}\right) \geq 1 - \gamma \quad (\text{G5})$$

From Eqn. (G3), for more than  $1 - \gamma$  instances of  $\mathbf{x}$

$$|p(\mathbf{x}|y=0) - q(\mathbf{x}|y=0)| \leq \frac{q(\mathbf{x}|y=0)}{\text{poly}(n)} + \frac{\epsilon(1 + o(1))}{2^{mn}\gamma},$$

which means that strong sampling from the Ising distribution with the above mixture of additive and multiplicative accuracy for more than  $1 - \gamma$  instances of  $\mathbf{x}$  is in the third level of the polynomial hierarchy.

But we already know [17] that for  $\frac{\epsilon}{\gamma} < \frac{1}{2}$ , there exists at least one choice of  $\mathbf{x}$  for which calculating of  $p(\mathbf{x}|y=0)$  is  $\#P$ -hard (due to the correspondence of a particular quantum computation, achievable by the Ising sampler, to a Gap-P complete problem [37]). Note that in our case  $\mathbf{x}$  also includes the measurement outcomes of the teleportation of the  $T$  gates which we set to 0 to get the correspondence to the same problem.

Assume now that Conjecture 1 is true. Notice that our conjecture, while it looks the same as in [17], applies to a slightly different distribution. The difference is that, in our case, we can have extra random rotations by  $\pi/2$  on some of the measurement angles, which come as a consequence of not applying the  $S$  gates conditioned on the



measurement outcome of the teleportation step of the FT  $T$  gates (Fig. 10 (ii)). This issue will not affect the implementation of the measurements with angles 0 in the extended brickwork state (Figure 4 (ii)), since they do not use magic states. It will affect the implementation of the measurements with non-zero angles. After calculation, the result is a  $k_0\pi/4$  rotation on the measurements of the brickwork state vertices (Figure 4 (i)), for some  $k_0$  which depends on the measurement outcomes of the magic state teleportation steps. The original argument made in [17] to support Conjecture 1 is that for random measurement outcomes of the vertices of the extended brickwork state, a uniform rotation over the 8 different  $\{k\pi/4\}$  angles is produced on the brickwork state. In our case it is the same, because the extra rotations cannot change the uniformity of these angles. Thus, the compu-

tation applied is based on a random brickwork MBQC pattern and connections to the random circuit model, such as in [16], can be made. In the latter paper, the average hardness of sampling from a random circuit is supported by drawing connections to quantum chaos and some numerical evidence.

Because of our assumption that there exists a  $F B P P^{NP}$  machine that can compute explicitly the values  $p(\mathbf{x}|y_0)$  for more than  $1 - \gamma$  instances of  $\mathbf{x}$  and at the same time (assuming the average case hardness) the probabilities of more than  $1 - \gamma$  instances of  $\mathbf{x}$  correspond to  $\#P$ -hard problems, we have  $PH \subseteq P^{\#P} \subseteq P^{F B P P^{NP}} \subseteq \Sigma_3^P$  (where the first inclusion is given by Toda's theorem) and the polynomial hierarchy collapses to the third level, an event expected to be highly unlikely.