



# Trading Locality for Time: Certifiable Randomness from Low-Depth Circuits

Matthew Coudron<sup>1</sup>, Jalex Stark<sup>2</sup>, Thomas Vidick<sup>3</sup> 

<sup>1</sup> National Institute of Standards and Technology/QuICS, University of Maryland, College Park, USA.  
E-mail: [mcoudron@umd.edu](mailto:mcoudron@umd.edu)

<sup>2</sup> University of California Berkeley, Berkeley, USA. E-mail: [jalex@cs.berkeley.edu](mailto:jalex@cs.berkeley.edu)

<sup>3</sup> California Institute of Technology, Pasadena, USA. E-mail: [vidick@caltech.edu](mailto:vidick@caltech.edu)

Received: 23 January 2019 / Accepted: 15 January 2021  
Published online: 9 February 2021 – © The Author(s) 2021

**Abstract:** The generation of certifiable randomness is the most fundamental information-theoretic task that meaningfully separates quantum devices from their classical counterparts. We propose a protocol for exponential certified randomness expansion using a single quantum device. The protocol calls for the device to implement a simple quantum circuit of constant depth on a 2D lattice of qubits. The output of the circuit can be verified classically in linear time, and is guaranteed to contain a polynomial number of certified random bits assuming that the device used to generate the output operated using a (classical or quantum) circuit of sub-logarithmic depth. This assumption contrasts with the locality assumption used for randomness certification based on Bell inequality violation and more recent proposals for randomness certification based on computational assumptions. Furthermore, to demonstrate randomness generation it is sufficient for a device to sample from the ideal output distribution within constant statistical distance. Our procedure is inspired by recent work of Bravyi et al. (Science 362(6412):308–311, 2018), who introduced a relational problem that can be solved by a constant-depth quantum circuit, but provably cannot be solved by any classical circuit of sub-logarithmic depth. We develop the discovery of Bravyi et al. into a framework for robust randomness expansion. Our results lead to a new proposal for a demonstrated quantum advantage that has some advantages compared to existing proposals. First, our proposal does not rest on any complexity-theoretic conjectures, but relies on the physical assumption that the adversarial device being tested implements a circuit of sub-logarithmic depth. Second, success on our task can be easily verified in classical linear time. Finally, our task is more noise-tolerant than most other existing proposals that can only tolerate multiplicative error, or require additional conjectures from complexity theory; in contrast, we are able to allow a small constant additive error in total variation distance between the sampled and ideal distributions.

## 1. Introduction

A fundamental point of departure between quantum mechanics and classical theory is that the former is non-deterministic: quantum mechanics, through the Born rule, posits the existence of experiments that generate *intrinsic randomness*. This observation leads to the simplest and most successful “test of quantumness” to have been designed and implemented: the Bell test [Bel64]. Far beyond its role as a test of the foundations of quantum mechanics, the Bell test has become a fundamental building block in quantum information, from protocols for quantum cryptography (e.g. device-independent quantum key distribution [Eke91, VV14]) to complexity theory (e.g. delegated quantum computation [RUV13], multiprover interactive proof systems [CHTW04]) and much more [BCP+14]. Yet, while a loophole-free implementation of a Bell test has been demonstrated [HBD+15, GVW+15, SMSC+15] it remains a challenging experimental feat, which unfortunately leaves its promising applications wanting (here “loophole-free” refers to a stringent set of experimental standards which ensure that all required assumptions have been verified “beyond reasonable doubt”). The increasingly powerful quantum devices that are being experimentally realized tend to be single-chip, and do not have the ability to implement loophole-free Bell tests. The task of devising convincing “tests of quantumness” for such devices is challenging.

Until recently the only proposal for such tests was the design of so-called “quantum supremacy experiments” [HM17], which specify classical sampling tasks that can in principle be implemented by a mid-scale quantum device, but cannot be simulated by any efficient classical randomized algorithm (under somewhat standard computational assumptions [AC17, HM17]). These proposals share a number of well-recognized limitations. Firstly, while the sampling part of the process can be done efficiently on a quantum computer, verifying that the quantum computer is sampling from a hard distribution requires a computational effort which scales exponentially in the number of qubits. Secondly, their experimental realization is hindered by a generally poor tolerance to errors in the implementation, which is compounded by the necessity to implement circuits with relatively large (say, at least  $\sqrt{N}$  for an  $N \times N$  grid) depth. Combined with the resort to complexity-theoretic assumptions for which there is little guidance in terms of concrete parameter settings (see however [DHKLP18]), this has led to an ongoing race in efficient simulations [CZX+18, HNS18, MFIB18]. Indeed, the proposals operate in a limited computational regime, requiring a machine with, say, at least 50 qubits (to prevent direct classical simulation) but at most 70 qubits (so that verification can be performed in a reasonable amount of time)—leaving open the question of what to do with a device with more than, say, 100 qubits. At a more conceptual level, the proposals are based on computational tasks that appear arbitrary (such as the implementation of a random quantum circuit from a certain class). In particular, they do not lead to any further characterization of the successful device, that could be used to e.g. build a secure delegation protocol or even simply certify a simple property such as the preparation of a specific quantum state or the implementation of a certain class of measurements.

We propose a different kind of experiment, or “test of quantumness”, for large but noisy quantum devices, that is inspired from recent work of Bravyi et al. on the power of low-depth quantum circuits [BGK18]. Our test is applicable in a regime where the device has a large number of qubits, but may only have the ability to implement circuits of low (constant) depth, due e.g. to a limited gate fidelity. We argue that the test overcomes the main limitations outlined above: it generates useful outcomes (certifiably random bits), it is easily verifiable (in classical linear time), and it is robust to a small amount of error (it is sufficient to generate outputs within constant statistical distance from the

ideal distribution<sup>1</sup>). The test does not require any assumption from complexity theory, but instead considers a novel physical assumption (introduced in [BGK18]): that the device implements a circuit whose depth is at most a small constant times the logarithm of the number of qubits. Intuitively, this assumption trades off locality (as required by the Bell test) for time (as measured by circuit depth). It is particularly well-suited to quantum devices for which the number of qubits can be made quite large, but the gate fidelity remains low, limiting the depth of a circuit that can be implemented. Informally, we show the following.

**Theorem 1.1.** *There exists universal constants  $c, d > 0$ , a family of distributions  $\{\mathcal{D}_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $\mathcal{D}_N$  is an efficiently sampleable distribution on  $\{0, 1\}^{N^2}$ , and a family of efficiently verifiable relations  $\{\mathcal{R}_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $\mathcal{R}_N \subseteq \{0, 1\}^{N^2} \times \{0, 1\}^{N^2}$ , such that the following holds:*

- *(Completeness) There exists a family of depth-3 geometrically local (in 2D) quantum circuits  $\{C_N\}_{N \in \mathbb{N}}$  such that for any  $N \geq 1$  and any input  $x$  in the support of  $\mathcal{D}_N$  it holds that  $(x, C_N(x)) \in \mathcal{R}_N$  with probability 1.*
- *(Soundness) For any family of classical circuits  $\{C_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $C_N$  has constant fan-in and depth at most  $c \log N$ , the probability that  $(x, C_N(x)) \in \mathcal{R}_N$  for  $x \sim \mathcal{D}_N$  is  $O(N^{-1/5})$ .*
- *(Randomness certification) It is possible to efficiently generate a sample from  $\mathcal{D}_N$  using  $O(\log^2 N)$  uniformly random bits. Moreover, for any family  $\{C_N\}_{N \in \mathbb{N}}$  of (classical or quantum) circuits with gates of constant fan-in and such that  $C_N$  has depth at most  $c \log N$  and satisfies*

$$\Pr_{x \sim \mathcal{D}_N} [(x, C_N(x)) \in \mathcal{R}_N] = \Omega(N^{-1/5}),$$

*it holds that the distribution of  $C_N(x)$  for  $x \sim \mathcal{D}_N$  has  $\Omega(N^d)$  bits of (smooth) min-entropy.*

We refer to Theorem 6.7 for a precise statement. In particular, the output entropy is quantified using the quantum conditional min-entropy, conditioned on the inputs to the circuit and quantum side information that may be correlated with the initial state of the circuit. In this sense it is guaranteed that outputs generated by the circuit contain “genuine” randomness, that is independent from the circuit inputs and uncorrelated from an eavesdropper that may hold a quantum state entangled with the state of any ancilla qubits on which the circuit operates. Thus our construction achieves both randomness *certification* (outputs of the circuit have entropy as long as a simple test is passed with sufficiently large probability) and *expansion* (outputs of the circuit have entropy even conditioned on the inputs).

Theorem 6.7 provides a “test of quantumness” in the sense that given a (family of) circuit  $C_N$  that maps  $N^2$  bits to  $N^2$  bits and satisfies  $(x, C_N(x)) \in \mathcal{R}_N$  with sufficiently large probability when  $x \sim \mathcal{D}_N$ , it must be that either the circuit is non-classical, or it has at least logarithmic depth. The requirement for succeeding in our test is more robust than most existing proposals, that require the output distribution to be multiplicatively close to the target distribution. In contrast, in our case it is sufficient to hit a certain target (the relation  $\mathcal{R}$ , that itself is very permissive) an inverse polynomial fraction of the time! The downside is that the condition that the circuit has small depth may be difficult to certify in practice; we discuss the use of timing assumptions below.

<sup>1</sup> In fact, even less is needed; see the description of the protocol.

Aside from the application to randomness expansion, Theorem 1.1 strengthens the main result of Bravyi et al. [BGK18] in multiple ways. In the first version of their work that was publicized [BGK17] Bravyi et al. provide a relation such that for any classical circuit of sufficiently low depth, there *exists* an input such that the circuit must return an output that satisfies the relation with probability bounded away from 1. In contrast, we point out the existence of an efficiently sampleable distribution on inputs such that, for any classical low-depth circuit, we know that *on average* over the choice of an input the circuit returns an output that satisfies the relation with at most small probability. This improvement follows using a simple extension of the arguments in [BGK17], and in fact a similar improvement was independently derived by Bravyi et al. in the final version of their paper [BGK18]. In addition to this improvement, which is key to the practical relevance of the scheme, we address the following question left open in [BGK18]: how small can one drive the maximum success probability of any *classical* low-depth circuit (i.e. the “soundness”)?

**Theorem 1.2.** (Exponential soundness) *There exists universal constants  $c, c' > 0$ , a family of distributions  $\{\mathcal{D}_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $\mathcal{D}_N$  is an efficiently sampleable distribution on  $\{0, 1\}^{N^2}$ , and a family of efficiently verifiable relations  $\{\mathcal{R}_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $\mathcal{R}_N \subseteq \{0, 1\}^{N^2} \times \{0, 1\}^{N^2}$ , such that the following holds:*

- (Completeness) *As in Theorem 1.1.*
- (Soundness) *For any family of classical circuits  $\{\mathcal{C}_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $\mathcal{D}_N$  has depth at most  $c \log N$ , the probability that  $(x, \mathcal{C}_N(x)) \in \mathcal{R}$  for  $x \sim \mathcal{D}_N$  is  $O(\exp(-N^{c'}))$ .*

Note that the strengthened statement on the possibility for low-depth classical circuits to generate outputs that satisfy the right relation comes at the cost that in Theorem 1.2 it is no longer the case that it is possible to sample from  $\mathcal{D}$  using poly-logarithmically many bits. Nevertheless, it remains possible to sample from  $\mathcal{D}$  in classical randomized polynomial-time, which is crucial for an experimental demonstration. In this case the improvement in the soundness guarantee is significant since it allows to use the test provided by the relation  $\mathcal{R}$  as a means to distinguish relatively noisy quantum circuits of depth 3—where here by “noisy” we mean that the circuits would satisfy  $(x, \mathcal{D}(x)) \in \mathcal{R}$  with probability that is e.g. inverse-polynomial, instead of 1 for a perfect implementation—from classical circuits of logarithmic depth, that may achieve  $(x, \mathcal{D}(x)) \in \mathcal{R}$  with probability that is exponentially small at best.

**Discussion.** We comment on the depth assumption that underlies our results, and their potential for a practical demonstration of a quantum advantage (a.k.a. “quantum supremacy experiment”). The quantum circuit required for a successful implementation of our task is relatively straightforward to implement. It can be realized in three phases. A first, offline phase initializes EPR pairs (or three-qubit GHZ states) between nearest-neighbor qubits on a 2D grid. In a second phase, each qubit is provided an input, according to which either the qubit should be measured according to a single-qubit Pauli observable, or the qubit and one of its neighbors should be measured in the Bell basis. Finally, in the third phase the measurement outcomes are aggregated and verified using a simple classical linear-time computation.

In order to demonstrate a quantum advantage, the crucial requirement is that the second phase should be implemented using a procedure that is “certified” to have low depth. Since this is a physical assumption, it can never be rigorously proven. Nevertheless, it is possible to imagine experiments under which the assumption would hold “beyond reasonable doubt”. We describe two such experiments.

In a first scenario, the verification of the depth constraint could be based on a calculation that takes into account state-of-the-art clock speeds. The fastest classical processors operate at speeds of order 1GHz, so that for an integer  $N$ , a circuit of depth  $d = \log(N)$  takes time of order  $10^{-9} \log(N)$  seconds to implement. In contrast, current gate times for, say, ion-trap quantum computers are of order 100 nanoseconds [SBT+18], meaning that the quantum circuit realizing our task could be implemented in time roughly  $10^{-7}$  seconds. To observe a quantum advantage it is thus necessary to ensure  $\log(N) \gg 10^2$ , leading to an impractical circuit size. However, a reasonable factor 10 improvement in the gate time for quantum gates could enable a demonstration based on a grid of order  $2^{10} \times 2^{10}$  qubits. Although far beyond current capabilities, the number is not beyond reach. Keeping in mind the extreme simplicity of the task to be implemented, it is not unreasonable to hope that such circuits may exist within 5–10 years.

In the previous scenario we allowed both the classical and quantum procedures solving our task to do so in a highly localized, single-chip fashion. The distributed nature of the task lends itself well to another type of implementation, that would be more demanding for a classical adversarial behavior, and may thus lead to a more practical demonstration of quantum advantage. Consider a network of constant-qubit devices arranged in a  $N \times N$  grid, such that devices may be separated by large (say, kilometric) distances. In the first, offline phase the devices use nearest-neighbor quantum communication channels to distribute EPR pairs. In the second phase, each device receives a classical input, performs a simple local measurement, and returns a classical output (no communication is required). Our result implies that, to even approximately reproduce the output distribution implemented by this procedure, a classical network would need to operate in at least  $\Omega(\log N)$  rounds, where in each round a device can communicate with a constant number of devices located at arbitrary locations in the network (the network need not be 2D: at each step, a device is allowed to broadcast arbitrarily but can only receive information from a constant number of devices, whose identity must be fixed ahead of time). Taking into account inevitable latency delays incurred in any such network, this second scenario suggests that our task may lead to an interesting test for a future quantum internet [WEH18].

Finally we comment on the fidelity requirement for the gates of a quantum circuit implementing our task. Even though the circuit is only of constant depth, it is important that, along a typical path of length  $O(N)$  between two qubits in the  $N \times N$  grid, none of the gates leads to an error. This means that per-gate fidelity is required to be of order  $1 - O(1/N)$ . For  $N$  of order  $2^{10}$ , as suggested in the first scenario described above, such fidelities are within reach. We also note that by changing the architecture of the circuit from a 2D grid to a 3D grid it may be possible to leverage existing protocols for entanglement distribution using noisy resources [RBH05]. Unfortunately, this comes with the drawback of a challenging 3D architecture for which there is no current implementation.

**Proof idea.** Our starting point is the key observation, made by Bravyi et al. [BGK18], that a sub-logarithmic depth circuit made of gates with constant fan-in has a form of implied locality, where the “forward lightcone” of most input vertices only includes a vanishing fraction of output vertices. In particular, two randomly chosen input locations are unlikely to have overlapping lightcones. If the input to the circuit is non-trivial in those two locations only, then the outputs in each input location’s forward lightcone are obtained by a computation that depends on that input only. In other words, we have a reduction from classical, low-depth circuits to two-party local computation that exactly preserves properties of the output. While the same lightcone argument holds true for a quantum circuit, the quantum circuit has the ability of distributing entanglement

across any two locations in depth 2, by executing a sequence of entanglement swapping procedures in parallel. Thus the same reduction maps a quantum, low-depth circuit to a two-party local computation, where the parties may perform their local computation on a shared entangled state. Since there are well-known separations between the kinds of distributions that can be generated by performing local operations on an entangled state, as opposed to no entanglement at all—this is precisely the scope of Bell inequalities—Bravyi et al. have obtained a separation between the power of low-depth classical and quantum circuits.

We build on this argument in the following way. Our first contribution is to boost the argument in [BGK18] from a worst-case to a “high probability” statement. Instead of showing that (i) for every classical circuit, there is some choice of input on which the classical circuit will fail, and (ii) there is a quantum circuit that succeeds on every input, we show that there exists a suitable distribution on inputs that is such that, (i) any classical circuit fails with high probability given an input from the distribution, and (ii) there is a quantum circuit that succeeds with high probability (in fact, probability 1) on the distribution. Second, we observe that the construction in [BGK18] imposes constraints not only on classical low-depth circuits, but also on quantum low-depth circuits; this observation enables the reduction to nonlocal games hinted at above. Finally, we amplify the argument to show how a polynomial number of Bell experiments can be simultaneously “planted” into the input to the circuit. This allows us to perform a reduction to a nonlocal game in which there is a large number of players divided into pairs which each perform their own distinct Bell experiment. By adapting techniques from the area of randomness expansion from nonlocal games [AFDF+18] we are then able to conclude that any sub-logarithmic-depth circuit, classical or quantum, that succeeds on our input distribution, must generate large amounts of entropy. Moreover, this guarantee holds even if the circuit only correctly computes a sufficiently large but constant fraction of outputs for the games.

**Related work.** Two recent works investigate the question of certified randomness generation outside of the traditional framework of Bell inequalities. In [BCM+18] randomness is guaranteed based on the computational assumption that the device does not have sufficient power to break the security of post-quantum cryptography. The main advantages of this proposal are that the assumption is a standard cryptographic assumption, and that verification is very efficient. A drawback is the interactive nature of the protocol, where only a fraction of a bit of randomness is extracted in each round. In [Aar18], Aaronson announced a randomness certification proposal based on the Boson Sampling task. The main advantage of the proposal is that it can potentially be implemented on a device with fewer than 100 qubits. Drawbacks are the difficulty of verification, that scales exponentially, and the resort to somewhat non-standard complexity conjectures, for which there is little evidence of practical hardness (e.g. it may not be clear how to set parameters for the scheme so that an adversarial attack would require time  $2^{80}$ ). In comparison, we would say that an advantage of our proposal is its simplicity to implement (on an axis different from Aaronson’s: we require many more qubits, but a much simpler circuit, of constant depth and with classically controlled Clifford gates only), its robustness to errors, and its ease of verification. A possible drawback is the physical assumption of bounded depth, that may or may not be reasonable depending on the scenario (in contrast to cryptographic or even complexity-theoretic assumptions, that operate at a higher level of generality).

Two other works obtained concurrently and independently from ours establish directly related, but strictly incomparable, results. In [Gal18] Le Gall obtains an average-

case hardness result that is very similar to our Theorem 1.2, with a concrete constant  $c' = 1/2$  that is likely better than the one that we achieve here. Le Gall's proof is based on an ingenious construction using the framework of graph states; although some aspects are similar in spirit to ours (such as the use of parallel repetition to amplify the soundness guarantees) the proof rests on rather different intuition. In independent work, Bene Watts et al. [BWKST19] extend the results of [BGK18] to obtain a result analogous to our Theorem 1.2, with a strengthened soundness property which holds even against so-called  $AC^0$  circuits.  $AC^0$  circuits are still required to have constant depth but may contain AND and OR gates of arbitrary fan-in (instead of constant fan-in for [BGK18] and our results). Their proof applies to the same relation as [BGK18] but uses more involved techniques from classical complexity theory to obtain the strengthened lower bound. Neither of these results obtains an application to randomness expansion as in our Theorem 1.1.

## 2. Preliminaries

*2.1. Notation.* Finite-dimensional Hilbert spaces are designated using calligraphic letters, such as  $\mathcal{H}$ . A register  $A, B, R$ , represents a physical subsystem, whose associated Hilbert space is denoted  $\mathcal{H}_A, \mathcal{H}_B$ , etc. We say a register is *classical* if there is a fixed basis  $|i\rangle$  in which the only allowed states of the subsystem are states of the form  $\sum_i p_i |i\rangle\langle i|$ , which are diagonal in the designated basis. We write  $\text{Id}_R$  for the identity operator on  $\mathcal{H}_R$ . A POVM  $\{M^a\}$  on  $\mathcal{H}$  is a collection of positive semidefinite operators on  $\mathcal{H}$  such that  $\sum_a M^a = \text{Id}$ . For  $X$  a linear operator on  $\mathcal{H}$ , we write  $\text{Tr}(X)$  for the trace and  $\|X\|_1 = \text{Tr}\sqrt{X^\dagger X}$  for the Schatten-1 norm.

For an integer  $d \geq 1$  an *observable over  $\mathbb{Z}_d$*  is a unitary operator  $A$  such that  $A^d = \text{Id}$ . For  $\omega = e^{\frac{2i\pi}{d}}$  and taking addition modulo  $d$  we write

$$X = \sum_{i=0}^{d-1} |i+1\rangle\langle i| \quad \text{and} \quad Z = \sum_{i=0}^{d-1} \omega^i |i\rangle\langle i|$$

for the generalized qudit Pauli  $X$  and  $Z$  operators, which are observables acting on  $\mathcal{H} = \mathbb{C}^d$ . Given an integer  $d \geq 1$  and a tuple  $s \in \mathbb{Z}_d^2$ , we write  $\sigma_s = X^{s_0} Z^{s_1}$  for a one-qudit Pauli acting on  $\mathbb{C}^d$ . Given an integer  $n \geq 1$  and a string  $r \in (\mathbb{Z}_d^2)^n$ , we write  $\sigma_r = \otimes_i \sigma_{r_i}$  for an  $n$ -qudit Pauli acting on  $(\mathbb{C}^d)^{\otimes n}$ .

*2.2. Nonlocal games.* We consider two types of games: multiplayer nonlocal games, and circuit games. Circuit games are nonstandard, and we introduce them in Sect. 5. Nonlocal games are defined as follows.

**Definition 2.1.** (*Nonlocal game*) Let  $\ell \geq 1$  be an integer. An  $\ell$ -player nonlocal game  $\mathcal{G}$  consists of finite question and answer sets  $X = X_1 \times \dots \times X_\ell$  and  $A = A_1 \times \dots \times A_\ell$  respectively, a distribution  $\pi$  on  $X$ , and a family of coefficients  $V(a_1, \dots, a_\ell | x_1, \dots, x_\ell) \in [0, 1]$ , for  $(x_1, \dots, x_\ell) \in X$  and  $(a_1, \dots, a_\ell) \in A$ . We call an element  $x \in X$  in the support of  $\pi$  a *query*, and for  $i \in \{1, \dots, \ell\}$  the  $i$ th entry  $x_i$  of  $x$  a *question* to the  $i$ th player. We refer to the function  $V(\cdot | \cdot)$  as the win condition for the game, and for any query  $x$ , to a tuple  $a$  such that  $V(a | x) = 1$  as a valid (or winning) tuple of answers (to query  $x$ ). When players return valid answers we say that they win the game.

**Definition 2.2.** (*Strategy*) Let  $\ell \geq 1$  be an integer, and  $\mathcal{G}$  an  $\ell$ -player nonlocal game. An  $\ell$ -player strategy  $\tau = (\rho, \{M_{x_i}\})$  for  $\mathcal{G}$  consists of an  $\ell$ -partite density matrix  $\rho \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_\ell$ , and for each  $i \in \{1, \dots, \ell\}$  a collection of measurement operators  $\{M_{x_i}^{a_i}\}_{a_i \in A_i}$  on  $\mathcal{H}_i$  indexed by  $x_i \in X_i$  and with outcomes  $a_i \in A_i$ .

**Definition 2.3.** (*Game value*) Let  $\mathcal{G}$  be an  $\ell$ -player nonlocal game, and  $\tau = (\rho, \{M_{x_i}^{a_i}\})$  a strategy for the players in  $\mathcal{G}$ . The *value* of  $\tau$  in  $\mathcal{G}$  is

$$\omega_\tau^*(\mathcal{G}) = \sum_{x_1, \dots, x_\ell} \pi(x_1, \dots, x_\ell) \sum_{a_1, \dots, a_\ell} V(a_1, \dots, a_\ell | x_1, \dots, x_\ell) \text{Tr}((M_{x_1}^{a_1} \otimes \cdots \otimes M_{x_\ell}^{a_\ell}) \rho).$$

A strategy  $\tau$  is called *perfect* if  $\omega_\tau^*(\mathcal{G}) = 1$ . The *entangled value* (or simply *value*) of  $\mathcal{G}$ ,  $\omega^*(\mathcal{G})$ , is defined as the supremum over all strategies  $\tau$  of  $\omega_\tau^*(\mathcal{G})$ .

To compare strategies we first introduce a notion of distance between measurements, with respect to an underlying state. (This is a standard definition in the area of self-testing.)

**Definition 2.4.** (*State-dependent distance*) Let  $\rho$  be a density matrix in  $\mathcal{H}$  and let  $M = \{M^a\}_a, N = \{N^a\}_a$  be two POVM on  $\mathcal{H}$  that have the same set of outcomes. The *state-dependent distance* between  $M$  and  $N$  is

$$d_\rho(M, N) = \left( \sum_a \text{Tr}((M^a - N^a)^2 \rho) \right)^{1/2}. \quad (1)$$

**Definition 2.5.** (*Closeness of strategies*) Let  $\tau = (\rho, \{M_{x_i}^{a_i}\})$ ,  $\tilde{\tau} = (\tilde{\rho}, \{\tilde{M}_{x_i}^{a_i}\})$  be strategies for an  $\ell$ -player nonlocal game  $\mathcal{G}$ . We say that  $\tau$  is  $\varepsilon$ -close to  $\tilde{\tau}$  if and only if  $\|\rho - \tilde{\rho}\|_1 \leq \varepsilon$  and for all  $i \in \{1, \dots, \ell\}$  it holds that  $\mathbb{E}_x d_\rho(M_{x_i}, \tilde{M}_{x_i}) \leq \varepsilon$ , where the expectation is over  $x = (x_1, \dots, x_\ell)$  drawn from  $\pi$ .

**Definition 2.6.** (*Isometric strategies*) Let  $\tau = (\rho, \{M_{x_i}^{a_i}\})$  and  $\tau' = (\rho', \{(M')_{x_i}^{a_i}\})$  be strategies for an  $\ell$ -player nonlocal game  $\mathcal{G}$ , and  $\varepsilon > 0$ . We say that  $\tau$  is  $\varepsilon$ -isometric to  $\tau'$  if and only if there exist isometries  $V_i : \mathcal{H}_i \rightarrow \mathcal{H}'_i$  for each  $i \in \{1, \dots, \ell\}$  such that  $\tau'$  is  $\varepsilon$ -close to the strategy  $\tilde{\tau} = (\tilde{\rho}, \{\tilde{M}_{x_i}^{a_i}\})$ , where  $\tilde{\rho} = (V_1 \otimes \cdots \otimes V_\ell) \rho (V_1 \otimes \cdots \otimes V_\ell)^\dagger$  and for all  $i \in \{1, \dots, \ell\}$ ,  $x_i \in X_i$  and  $a_i \in A_i$ ,  $\tilde{M}_{x_i}^{a_i} = V_i M_{x_i}^{a_i} V_i^\dagger$ .

**Definition 2.7.** We say that a game  $\mathcal{G}$  is *robustly rigid* if the following holds. There is a continuous function  $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $f(0) = 0$  and a strategy  $\tau$  for  $\mathcal{G}$  such that for any  $\delta \geq 0$ , any strategy  $\tau'$  with value at least  $\omega_\tau^*(\mathcal{G}) - \delta$  is  $f(\delta)$ -isometric to  $\tau$ . We refer to  $f$  as the robustness of the game.

Note that for a game to be robustly rigid it is necessary that there exists a unique strategy  $\tau$  such that  $\omega_\tau^*(\mathcal{G}) = \omega^*(\mathcal{G})$ , up to isometry.

**2.3. Circuits.** We refer to [NC02] for an introduction to the quantum circuit model. We consider layered circuits over an arbitrary gate set. The choice of a specific gate set may affect the depth of a circuit; for concreteness, the reader may consider the standard gate set  $\{X, Z, H, T, CNOT\}$ , where here  $X, Z$  are the Pauli observables over  $\mathbb{C}^2$ ,

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix},$$



and *CNOT* is the controlled-NOT gate. In general, gates in the gate set used to specify the circuit may have arbitrary fan-out, but are restricted to fan-in at most  $K$ , where  $K \geq 2$  is a parameter that is considered a constant (in contrast to the depth  $D$  of the circuit, that is allowed to grow with the number of input wires to the circuit). Note that if  $\mathcal{C}$  is a quantum circuit, “fan-in” is the same as locality, i.e. the number of qubits that a gate acts on nontrivially. In particular, for quantum circuits bounded fan-in automatically implies bounded fan-out.

It is convenient to generalize the usual notion of Boolean circuit to allow circuits that act on inputs taken from a larger domain, e.g.  $\mathcal{C} : \Sigma^n \rightarrow \Sigma^m$ , where  $\Sigma$  is a finite alphabet. Similar to the fan-in, whenever using the  $O(\cdot)$  notation we consider the cardinality of  $\Sigma$  a constant. A circuit of depth  $D$  and fan-in  $K$  over  $\Sigma$  can be converted in a straightforward way in a circuit of depth  $D$  and fan-in  $K \cdot \lceil \log_2 |\Sigma| \rceil$  over  $\{0, 1\}$ . For the case of quantum circuits, allowing a non-Boolean  $\Sigma$  amounts to considering a circuit that operates on  $d$ -dimensional qudits, for  $d = |\Sigma|$ , instead of 2-dimensional qubits.

**2.4. Entropies.** Given a bipartite density matrix  $\rho_{AB}$  we write  $H(A|B)_\rho$ , or simply  $H(A|B)$  when  $\rho_{AB}$  is clear from context, for the conditional von Neumann entropy,  $H(A|B) = H(AB) - H(B)$ , with  $H(X)_\sigma = -\text{Tr}(\sigma \ln \sigma)$  for any density  $\sigma$  on  $\mathcal{H}_X$ . We recall the definition of (smooth) min-entropy.

**Definition 2.8.** (*Min-entropy*) Let  $\rho_{XE}$  be a density matrix on two registers  $X$  and  $E$ , such that the register  $X$  is classical. The *min-entropy* of  $X$  conditioned on  $E$  is defined via the following optimization problem over the space  $\text{Pos}(\mathcal{H}_E)$  of positive semidefinite operators on  $\mathcal{H}$ .

$$H_{\min}(X|E)_\rho = \max\{\lambda \geq 0 : \exists \sigma_E \in \text{Pos}(\mathcal{H}_E), \text{Tr}(\sigma_E) \leq 1, \text{ s.t. } 2^{-\lambda} \text{Id}_X \otimes \sigma_E \geq \rho_{XE}\}.$$

When the state  $\rho$  with respect to which the entropy is measured is clear from context we simply write  $H_{\min}(X|E)$  for  $H_{\min}(X|E)_\rho$ . For  $\varepsilon \geq 0$  the  $\varepsilon$ -smooth min-entropy of  $X$  conditioned on  $E$  is defined as

$$H_{\min}^\varepsilon(X|E)_\rho = \max_{\sigma_{XE} \in \mathcal{B}(\rho_{XE}, \varepsilon)} H_{\min}(X|E)_\sigma,$$

where  $\mathcal{B}(\rho_{XE}, \varepsilon)$  is the ball of radius  $\varepsilon$  around  $\rho_{XE}$ , taken with respect to the purified distance.<sup>2</sup>

The following theorem justifies the use of the smooth min-entropy as the appropriate notion of entropy for randomness extraction.

**Theorem 2.9.** [DPVR12] *For any integers  $n, m$  and for any  $\varepsilon > 0$  there exists a  $d = O(\log^2(n/\varepsilon) \cdot \log m)$  and an efficient classical procedure  $\text{EXT} : \{0, 1\}^n \times \{0, 1\}^d \rightarrow \{0, 1\}^m$  such that for any density matrix  $\rho_{XE} = \sum_x |x\rangle\langle x|_X \otimes \rho_E^x$  such that the register  $X$  is an  $n$ -bit classical register and  $H_{\min}(X|E) \geq 2m$ , letting  $\rho_{ZYE} = 2^{-d} \sum_{x,y} |\text{EXT}(x, y)\rangle\langle \text{EXT}(x, y)|_Z \otimes |y\rangle\langle y|_Y \otimes \rho_E^x$  it holds that*

$$\|\rho_{ZYE} - U_m \otimes U_d \otimes \rho_E\|_1 \leq \varepsilon,$$

where for an integer  $\ell \geq 1$ ,  $U_\ell = 2^{-\ell} \text{Id}$  is the totally mixed state on  $\ell$  qubits and  $\rho_E = \sum_x \rho_E^x$ .

<sup>2</sup> The definition of the purified distance is not important for us, and we defer to [Tom15] for a precise definition.

### 3. Stabilizer Games

In this section we introduce a restricted class of nonlocal games that we will be concerned with throughout the paper. We call the games *stabilizer games*. They have the property that the game always has a perfect quantum strategy  $\tau = (\rho, \{M_{x_i}\})$  that uses an entangled state  $\rho = |\psi\rangle\langle\psi|$  that is a graph state, on which the players make measurements that are specified by tensor products of Pauli observables. It is important for our results that there is a perfect strategy such that the entangled state can be prepared by a quantum circuit of low depth (in fact, constant depth) starting on a  $|0\rangle$  state. It will also be convenient that the same perfect strategy only requires the measurement of Pauli operators, and that the win condition in the game is a linear function of the players' answers.

We proceed with a formal definition. The games we consider have  $\ell$  players. In the intended strategy for the players, each player  $j \in \{1, \dots, \ell\}$  holds  $k_j$  qudits, measures  $m_j$  commuting Pauli observables over  $\mathbb{Z}_d$  (depending on its question), and reports the  $m_j$  outcomes.

**Definition 3.1.** (*Stabilizer game*) An  $(\ell, k, m)$  stabilizer game  $\mathcal{G} = (X_i, \{w_x, b_x\})$  is an  $\ell$ -player nonlocal game defined from the following data.

- a number of players  $\ell$ ,
- a parameter  $d$  for the dimension of the qudits (in the honest strategy),
- for  $j \in \{1, \dots, \ell\}$ , a parameter  $k_j$  for the number of qudits held by the  $j$ th player (ibid),
- for  $j \in \{1, \dots, \ell\}$ , a parameter  $m_j$  for the number of simultaneous measurements made by the  $j$ th player (ibid),
- for  $j \in \{1, \dots, \ell\}$ , a set  $X_j$ , each element of which is identified with the label  $x \in (\mathbb{Z}_d^2)^{k_j}$  of a  $k_j$ -qudit Pauli,
- a distribution  $\pi$  on queries  $x \in \prod_{j=1}^{\ell} X_j^{m_j}$ , such that any  $(x_1, \dots, x_\ell)$  in the support of  $\pi$  is such that for each  $j$ ,  $x_j$  designates an  $m_j$ -tuple of commuting  $k_j$ -qudit Pauli observables,
- for each query  $x$  in the support of  $\pi$ , a vector  $w_x \in \prod_{j=1}^{\ell} (\mathbb{Z}_d)^{m_j}$  and a coefficient  $b_x \in \mathbb{Z}_d$  that are used to specify the win condition in the game.

To play, the verifier samples a question  $x_j \in X_j^{m_j}$  for each player. Each player responds with a string  $a_j \in \mathbb{Z}_d^{m_j}$ . Let  $x = (x_1, \dots, x_\ell)$  and  $a = (a_1, \dots, a_\ell)$ . The players win if

$$w_x \cdot a = b_x, \quad (2)$$

where the inner product is over vectors in  $\mathbb{Z}_d^{\sum m_j}$ . Using the notation from Definition 2.1,  $V(a|x) = 1$  if  $w_x \cdot a = b_x$ , and 0 otherwise.

In a stabilizer game each player is tasked with reporting  $m$  values in  $\mathbb{Z}_d$ . It is then natural to use a representation of strategies in terms of observables over  $\mathbb{Z}_d$ . We adapt Definition 2.2 as follows.

**Definition 3.2.** Let  $\mathcal{G} = (X_i, \{w_x, b_x\})$  be a stabilizer game. A *strategy*  $\tau = (\rho, \{M_{x_j}\})$  for  $\mathcal{G}$  is specified by an  $\ell$ -partite density matrix  $\rho$  and for each  $j \in \{1, \dots, \ell\}$  and  $x_j = (x_{j,1}, \dots, x_{j,m_j}) \in X_j^{m_j}$  a family of  $m_j$ -tuples of commuting observables  $M_{x_j} = (M_{x_{j,1}}, \dots, M_{x_{j,m_j}})$  over  $\mathbb{Z}_d$ .

Note that in the definition, for  $s \in \{1, \dots, m_j\}$  the observable  $M_{x_j, s}$  may depend on the whole  $m_j$ -tuple  $x_j$ , and not only on  $x_{j, s}$ .

We introduce a notion of “honest strategy” in a stabilizer game.

**Definition 3.3.** (*Honest strategy*) Let  $\mathcal{G} = (X_j, \{w_x, b_x\})$  be a stabilizer game. A *honest strategy* in  $\mathcal{G}$  is a strategy in which the state  $\rho$  is an  $(\sum_j k_j)$ -qudit  $\ell$ -partite pure state  $|\psi\rangle$  such that the  $j$ th player has  $k_j$  qubits, and the player’s observables  $M_{x_j} = (M_{x_{j,1}}, \dots, M_{x_{j,m}})$  associated with question  $x_j = (x_{j,1}, \dots, x_{j,m_j}) \in X_j^{m_j}$  are precisely the  $m_j$  commuting Pauli observables specified by  $x_j$ . We say that the strategy has *depth*  $d$  if the state  $|\psi\rangle$  can be prepared by a quantum circuit of depth at most  $d$  starting from the  $|0\rangle$  state.

*3.1. Pauli observables.* Recall the notation  $\sigma_r$ , where  $r \in (\mathbb{Z}_d^2)^k$ , introduced in Sect. 2.1 to designate an arbitrary  $k$ -qudit Pauli observable.

**Definition 3.4.** (*Correction value*) Let  $q, r \in (\mathbb{Z}_d^2)^k$ . The *correction value*  $\text{cor}_r(q) \in \mathbb{Z}_d$  is defined such that

$$\omega^{\text{cor}_r(q)} I = [\sigma_q, \sigma_r], \quad (3)$$

where the brackets denote the group commutator,  $[P, Q] = P Q P^{-1} Q^{-1}$ .

(The above definition takes advantage of the fact that the group commutator of Pauli matrices is always a scalar multiple of the identity matrix.) By abuse of notation, we will also sometimes write  $\text{cor}$  as a vector valued function, where if  $r = \text{cor}(r_1, \dots, r_i)$  and  $q = (q_1, \dots, q_i)$  then  $\text{cor}_r(q)$  is defined as  $(\text{cor}_{r_1}(q_1), \dots, \text{cor}_{r_i}(q_i))$ .

The following lemma shows that the function  $\text{cor}$  can be computed locally.

**Lemma 3.5.** (*Cor can be computed locally*) For a string  $s$ , let  $s|_i$  denote the string which is equal to  $s_i$  in the  $i^{\text{th}}$  position and 0 everywhere else. Then for any  $q$  and  $r$ ,

$$\sum_i \text{cor}_{r|_i}(q|_i) = \text{cor}_r(q). \quad (4)$$

*Proof.* First, notice that

$$\text{cor}_{r|_i}(q|_i) = \text{cor}_{r|_i}(q). \quad (5)$$

To see this, recall that  $\text{cor}$  is computed as the phase of the group commutator of a  $P_{r|_i}$  and  $\sigma(q)$ . We can evaluate this group commutator one tensor factor at a time. In all tensor factors other than  $i$ , the commutator will be trivial since the  $r$  operator is identity. Therefore, the commutator does not change if we also set the  $q$  operator to identity.

Next, we need that for any fixed  $q$ , the map  $r \mapsto \text{cor}_r(q)$  is an additive homomorphism. In other words,

$$\text{cor}_{r+r'}(q) = \text{cor}_r(q) + \text{cor}_{r'}(q). \quad (6)$$

To see this, we apply Lemma 3.6 with  $A = \sigma_q$ ,  $B = \sigma_r$ ,  $C = \sigma_{r'}$ .

The lemma follows by combining Equations (5) and (6) with the observation that  $r = \sum_i r|_i$ .  $\square$

**Lemma 3.6.** (*Commutators*) Suppose  $B$  commutes with  $[A, C]$ . Then  $[A, B][A, C] = [A, BC]$ .

*Proof.* Write  $[A, BC]$  as  $A(BC)A^{-1}(BC)^{-1}$ . Note that by definition,  $AB = [A, B]BA$ . Then we have

$$\begin{aligned} [A, BC] &= A(BC)A^{-1}(BC)^{-1} \\ &= ABCA^{-1}C^{-1}B^{-1} \\ &= [A, B]BACA^{-1}C^{-1}B^{-1} \\ &= [A, B]B[A, C]B^{-1} \\ &= [A, B][A, C], \end{aligned}$$

where the last line follows from commutation of  $B$  and  $[A, C]$ .  $\square$

**3.2. Rotated and stretched stabilizer games.** In this section we define stretched stabilizer games which formalize the notion of distributing a stabilizer game out over long “paths”. One property of stretched games is that players on far ends of the paths have outputs which require correction according to a function of the outputs along the intermediate points in the paths. We introduce a notion of *rotated stabilizer game* that captures this scenario by allowing the players to report an additional “rotation string”.

**Definition 3.7.** (*Rotated stabilizer game*) Given a stabilizer game  $\mathcal{G} = (X_j, \{w_x, b_x\})$  the rotated stabilizer game associated with  $\mathcal{G}$ ,  $\mathcal{G}^R$ , is defined as follows. For each  $j \in \{1, \dots, \ell\}$  and question  $x_j \in X_j^{m_j}$ , the  $j$ th player reports an answer  $a_j \in \mathbb{Z}_d^{m_j}$  together with a rotation string  $r_j \in (\mathbb{Z}_d^2)^{k_j}$ . The win condition (2) is replaced by the condition

$$w_x \cdot (a - \text{cor}_r(x)) = b_x, \quad (7)$$

where  $r = (r_1, \dots, r_\ell)$ .

Observe that if  $r$  is the 0 vector then for any  $q$ ,  $\text{cor}_r(q) = 0$ , so the win condition for the rotated game  $\mathcal{G}^R$  reduces to the win condition for  $\mathcal{G}$ . Therefore any strategy for  $\mathcal{G}$  implies a strategy for  $\mathcal{G}^R$  with the same success probability. More generally, it is possible to define a strategy in  $\mathcal{G}^R$  by having the players conjugate their observables in  $\mathcal{G}$  by an arbitrary Pauli observable (the same for all observables), and report as rotation string the string that specifies the observable used for conjugation.

Using Lemma 3.5 it follows that there is a reduction in the other direction as well. Given a strategy for  $\mathcal{G}^R$ , one obtains a strategy for  $\mathcal{G}$  by replacing the answer  $(a_i, r_i)$  from the  $i$ th player in  $\mathcal{G}^R$  by the answer

$$(a_i - \text{cor}_{r_i}(q_i)) \quad (8)$$

in  $\mathcal{G}$ . The following lemma summarizes this observation in terms of rigidity of the rotated game. Recall the definition of a robustly rigid game in Definition 2.7.

**Lemma 3.8.** (*Rotation preserves rigidity*) *Suppose that a stabilizer game  $\mathcal{G}$  is robustly rigid (see Definition 2.7). Let  $\tau = (|\psi\rangle\langle\psi|, \{M_{x_j}\})$  be a rigid strategy and  $f$  the robustness. Then the rotated stabilizer game  $\mathcal{G}^R$  is rigid in the following sense. For any strategy  $\tau = (\rho', \{M'_{x_j}\})$  that has value  $w' = \omega_\tau^*(\mathcal{G})$  in  $\mathcal{G}^R$ , there is a strategy in  $\mathcal{G}$  that is a coarse-graining of  $(\rho', \{M'_{x_j}\})$  according to (8),<sup>3</sup> and that has value  $w'$  in  $\mathcal{G}$ . In particular, up to local isometries the state  $\rho'$  is within distance  $f(1 - w')$  of  $|\psi\rangle\langle\psi|$ .*

<sup>3</sup> Here by “coarse-graining” we mean the strategy that is implied by requiring each player to compute the update (8) locally; Lemma 3.5 shows that this can always be done.

We introduce a notion of “stretched” rotated game, that will be useful when we relate circuit games to stabilizer games.

**Definition 3.9.** (*Stretched stabilizer game*) Let  $\mathcal{G} = (X_j, \{w_x, b_x\})$  be a stabilizer game, and  $\Gamma = (\Gamma_1, \dots, \Gamma_\ell)$  an  $\ell$ -tuple of finite sets, such that for  $j \in \{1, \dots, \ell\}$ ,  $\Gamma_j$  has  $k_j$  designated elements  $(u_{j,1}, \dots, u_{j,k_j})$ . Each element of  $\Gamma_j$  is used to index one out of  $|\Gamma_j|$  qudits that are supposed to be held by the  $j$ th player. To  $\mathcal{G}$  and  $\Gamma$  we associate a “stretched” game  $\mathcal{G}_\Gamma^S$  as follows. In  $\mathcal{G}_\Gamma^S$  the parameter  $k_j$  is replaced by  $k'_j = |\Gamma_j|$ . For any  $k_j$ -qudit Pauli observable asked to player  $j$  in  $\mathcal{G}$ , there is a  $k'_j$ -qubit Pauli observable in  $\mathcal{G}_\Gamma^S$  such that the observable acts as the identity on the additional  $(k'_j - k_j)$  qubits. The win condition in  $\mathcal{G}_\Gamma^S$  is the same as in  $\mathcal{G}$ .

Given a stabilizer game  $\mathcal{G}$  and sets  $\Gamma = (\Gamma_1, \dots, \Gamma_\ell)$ , we write  $\mathcal{G}_\Gamma^{S,R} = (\mathcal{G}_\Gamma^S)^R$  for the rotated stretched stabilizer game associated with  $\mathcal{G}$  and  $\Gamma$ .

**3.3. Repeated games.** For an integer  $r \geq 1$  we consider the game that is obtained by repeating a stabilizer game  $r$  times in parallel, with  $r$  independent sets of  $\ell$  players (that may share a joint entangled state).

**Definition 3.10.** Let  $\mathcal{G}$  be an  $(\ell, k, m)$  stabilizer game, and  $r \geq 1$  an integer. The  $r$ -fold repetition of  $\mathcal{G}$  is the  $(r\ell, k)$  stabilizer game  $\mathcal{G}_r$  that is obtained by executing  $\mathcal{G}$  independently in parallel with  $r$  groups of  $\ell$  players. More formally, the input distribution  $\pi_r$  in  $\mathcal{G}_r$  is the direct product of  $r$  copies of the input distribution  $\pi$  in  $\mathcal{G}$ , and the win condition in  $\mathcal{G}_r$  is the AND of the win conditions in each copy of  $\mathcal{G}$ .

For purposes of randomness expansion, in Sect. 6 we consider repeated games for which the input distribution  $\pi_r$  is not exactly the direct product of  $r$  copies of  $\pi$ , but a derandomized version of it. Similarly, to achieve better robustness, instead of the AND of the winning conditions we may consider a win condition that is satisfied as soon as sufficiently many of the win conditions for the subgames are satisfied. These modifications are explained in Sect. 6.1.

**3.4. The Magic Square game.** For concreteness we give two examples of stabilizer games, the Memin-Peres Magic Square game [Mer90b] and the Mermin GHZ game [Mer90a]. The former is given for illustration; the latter will be used towards randomness expansion in Sect. 6.

**Definition 3.11.** (*Magic Square game*) Consider the following  $3 \times 3$  matrix, where each entry is labeled by a two-qubit Pauli observable:

$$\begin{bmatrix} X \otimes I & I \otimes X & X \otimes X \\ I \otimes Z & Z \otimes I & Z \otimes Z \\ X \otimes Z & Z \otimes X & (XZ) \otimes (ZX) \end{bmatrix}. \quad (9)$$

The Magic Square game is a  $(2, 2, 2)$  stabilizer game over 2-dimensional qubits defined as follows. The sets

$$X_1 = X_2 = \left\{ (X \otimes I, I \otimes X), (I \otimes Z, Z \otimes I), (X \otimes Z, Z \otimes X) \right\}$$

each contain 6 pairs of two qubit-Pauli observables, each pair corresponding to either a row or column of (9). The distribution  $\pi$  is uniform. For any query  $x = (x_1, x_2)$  each player reports two bits associated with the two observables it was asked about. We can associate a third bit to the third observable in the corresponding row or column by taking the parity of the first two bits, except for the case of the third column, where we take the parity plus 1. The constraint  $w_x \cdot a = b$  expresses the constraint that, whenever the questions  $x_1, x_2$  are associated with a row and column that intersect in an entry of the square, the outcomes associated with the intersection should match.

**Definition 3.12.** (*Honest strategy in the Magic Square game*) In the honest strategy, the two players share two EPR pairs. Upon reception of a question that indicates two commuting two-qubit Pauli observables, the player measures both observables on her qubits and reports the two outcomes.

The following robustness result is shown in [WBMS16].

**Theorem 3.13.** *The Magic Square game is robustly rigid, with respect to the honest strategy and with robustness  $f(\delta) = O(\sqrt{\delta})$ .*

Next we recall the Mermin GHZ game.

**Definition 3.14.** (*GHZ game*) The game GHZ is a  $(3, 1, 1)$  stabilizer game over 2-dimensional qubits defined as follows. The sets  $X_1 = X_2 = X_3 = \{0, 1\}$ . The distribution  $\pi$  is uniform over the set  $\{(0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0)\}$ . For all queries  $x$  the vector  $w_x = (1, 1, 1)$ . For  $x = (0, 0, 0)$ ,  $b_x = 0$ , and for all other  $x$ ,  $b_x = 1$ .

It is well-known that there is a honest strategy based on making Pauli measurements on a GHZ state  $|\psi_{\text{GHZ}}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$  (which can be prepared in depth 3) and that succeeds with probability 1 in the game GHZ.

## 4. Lightcone Arguments for Low-Depth Circuits

Let  $N \geq 1$  be an integer. We write  $\text{GRID}_N$  for the set  $\{1, \dots, N\}^2$ , that we often identify with the “grid graph” of degree 4, which is the graph on this vertex set with an edge between  $(i, j)$  and  $(i \pm 1, j \pm 1)$ , with addition taken modulo  $N$ . (As a matter of notation we often identify a graph with its vertex set.)

For an integer  $0 \leq L \leq N$  and  $u \in \text{GRID}_N$  we write  $\text{Box}_L(u)$ , or  $\text{Box}(u)$  when  $L$  is implicit, for the set  $\text{Box}_L(u) = \{u\} + \{-L, \dots, L\}^2 \subseteq \text{GRID}_N$  (with addition again taken mod  $N$ ). In other words,  $\text{Box}_L(u)$  is the closed ball of radius  $L$  around  $u$  in the  $L_\infty$  metric.

*4.1. Lightcones.* Recall that the circuits that we consider are defined over an arbitrary gate set with bounded fan-in  $K$ . Given a circuit  $\mathcal{C}$ , we introduce the natural notion of a *circuit graph*, with vertices at the gates and edges along the wires.

**Definition 4.1.** Let  $\mathcal{C}$  be a circuit. The *circuit graph* associated with  $\mathcal{C}$  is a directed graph on vertex set  $V = \mathcal{I} \cup \mathcal{U} \cup \mathcal{O}$ . Here  $\mathcal{I}$  contains one vertex for each input wire,  $\mathcal{O}$  contains one vertex for each output wire, and  $\mathcal{U}$  contains one vertex for each gate. There is an edge from  $u$  to  $v$  if the output of  $u$  is an input of  $v$ . In particular, all vertices of  $\mathcal{I}$  are sources (have indegree 0) and all vertices of  $\mathcal{O}$  are sinks (have outdegree 0). We call vertices in  $\mathcal{I}$  *input vertices* and vertices in  $\mathcal{O}$  *output vertices*.

We typically consider circuits that are spatially local on a 2D grid, in which case we identify the input and output sets of the graph with a grid, i.e.  $\mathcal{I} = \mathcal{O} = \text{GRID}_N$  for some integer  $N \geq 1$ . Note that the circuit graph of a circuit with fan-in  $K$  has in-degree bounded by  $K$ , but has no a priori bound on the out-degree.

**Definition 4.2.** Let  $\mathcal{C}$  be a circuit. For a vertex  $v$  in the circuit graph define its *backward lightcone*  $L_b(v)$  as the set of input vertices  $u$  for which there exists a path in the circuit graph from  $u$  to  $v$ . For an input vertex  $u$  define the *forward lightcone of  $u$* ,  $L_f(u)$ , as the set of output vertices  $v$  such that  $u \in L_b(v)$ .

The following lemma is established in Section 4.2 of [BGK18] during the proof of their Theorem 2. We include the short proof for completeness.

**Lemma 4.3.** [BGK18] *Let  $\mathcal{C}$  be a circuit that has depth  $D$  and maximum fan-in  $K$ . Then the following hold:*

- All backward lightcones are small. That is, for every vertex  $v$  of the circuit graph,  $|L_b(v)| \leq K^D$ .
- Most forward lightcones are small. That is, for any  $\mu \in (0, 1)$ ,

$$\Pr_v[L_f(v) \geq \mu^{-1} K^D] \leq \mu, \quad (10)$$

where the probability is taken over the choice of a uniformly random input vertex  $v \in \mathcal{I}$ .

*Proof.* Every path in the circuit graph has length at most  $D$ . Each vertex has indegree at most  $K$ . Then for any fixed vertex  $v$ , there are at most  $K^D$  distinct paths through the circuit graph ending at  $v$ . Therefore,  $|L_b(v)| \leq K^D$ .

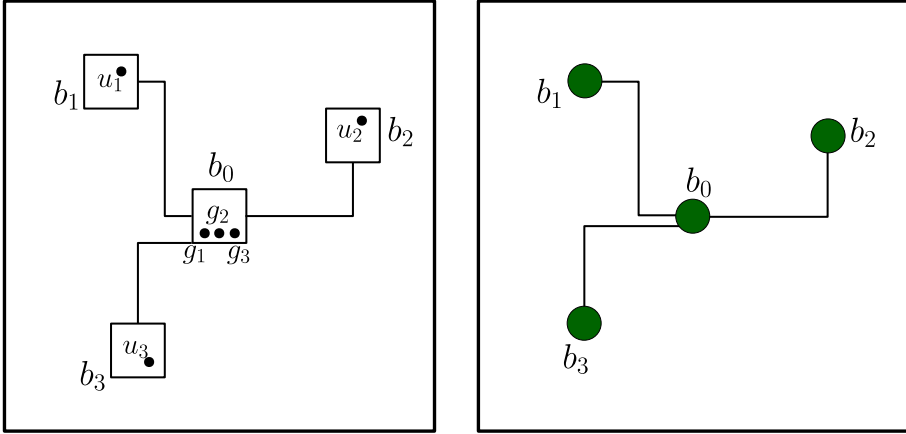
Now consider the directed graph with an edge from  $u$  to  $v$  if  $u \in L_b(v)$ . The in-degree of vertex  $v$  is equal to  $|L_b(v)|$  while its out-degree is  $|L_f(v)|$ . Each vertex has an in-degree of at most  $K^D$ , so there are at most  $nK^D$  edges in the graph, where  $n$  is the number of output wires for the circuit. Fix  $\mu \in (0, 1)$ . By Markov's inequality, at most  $\mu n$  vertices may have out-degree at least  $\frac{1}{\mu} K^D$ .  $\square$

**4.2. Input patterns.** We introduce a method to “plant” queries to the players in a stabilizer game into the input to a circuit. The main definition we need is of an *input pattern*, that specifies locations for each players’ question, as well as paths between these locations. These paths, or “stars”, will be useful in the design of a quantum circuit that implements the players’ strategy as a low-depth quantum circuit; this is explained in Sect. 5.1.

**Definition 4.4.** (*Star*) See Fig. 1. We say that a subset of  $\text{GRID}_N$  is a *box* if it is equal to  $\text{Box}_L(u)$  for some integer  $L$  and vertex  $u$ . A *star*  $\Gamma$  is a collection of disjoint boxes together with a collection of disjoint paths such that

- each path has its endpoints on the boundaries of boxes, and
- contracting each box to a single vertex, and each path to a single edge, results in a star graph, i.e. a graph that has  $\ell$  vertices of degree one, one vertex of degree  $\ell$ , and no other vertices.

We use the term *central box* to refer to the unique box which contains one endpoint of every path. If  $b_0$  is the central box, we may say that the star  $\Gamma$  is *centered at  $b_0$* . By abuse of notation, we often write  $\Gamma$  to refer to the set of vertices contained in the paths and boxes of  $\Gamma$ .



**Fig. 1.** A star centered at box  $b_0$ , connecting  $\ell = 3$  boxes  $b_1, b_2, b_3$  in  $\text{GRID}_N$  (grid edges are not shown on the picture). The paths may be extended inside each box to connect the vertices  $u_1, u_2, u_3$  to  $g_1, g_2, g_3$  in an arbitrary way. Such connections will be used in Sect. 5.1 to define low-depth measurements which distribute a three-qubit state at sites  $g_1, g_2, g_3$  among the qubits  $u_1, u_2, u_3$ . On the right, we show the contraction of the star to a star graph. The paths are contracted to single edges (shown by thick lines) and the boxes are contracted to single vertices (shown by filled-in circles)

The following definition captures exactly the amount of information that we need to remember about a given circuit  $\mathcal{C}$  in order to talk about the spread of correlations within  $\mathcal{C}$ —we will forget everything about the circuit except some information about its lightcones.

**Definition 4.5.** (*Input pattern*) Let  $\mathcal{G}$  be an  $(\ell, k, m)$  stabilizer game. Let  $1 \leq r \leq N$  be integer. An *input pattern* associated with  $(\mathcal{G}, N, r)$  is a tuple  $\mathcal{P} = \{(u^{(i)}, \Gamma^{(i)})\}_{1 \leq i \leq r}$  such that

- each  $u^{(i)} = (u_1^{(i)}, \dots, u_\ell^{(i)})$  is an  $\ell$ -tuple of vertices of  $\text{GRID}_N$ , which we refer to as *input locations*,
- each  $\Gamma^{(i)}$  is a star,
- the vertices of  $u^{(i)}$  are contained in distinct noncentral boxes of  $\Gamma^{(i)}$ . For a vertex  $u$ , we write  $\text{Box}(u)$  for the box that contains  $u$ .

*Remark 4.6.* We often write patterns as  $\mathcal{P} = \{(u^{(i)}, \Gamma^{(i)})\}$  without specifying the range of the index  $i$ , that we generally leave implicit. When we write “a pair  $(u^{(i)}, \Gamma^{(i)})$ ”, without the use of the curly brackets, we mean a pair, for some arbitrary but fixed index  $i$ .

**Definition 4.7.** (*Circuit specification*) A *circuit specification*  $\mathcal{S}$  on  $\text{GRID}_N$  is a triple  $\mathcal{S} = (L_f, \text{BAD}_{in}, \text{BAD}_{out})$  such that for all  $u \in \text{GRID}_N$ ,  $L_f(u) \subseteq \text{GRID}_N$  is a set called the *forward lightcone* associated with  $u$ , and  $\text{BAD}_{in}, \text{BAD}_{out} \subseteq \text{GRID}_N$  are sets called the *bad input set* and *bad output set* respectively.

**Definition 4.8.** For integer  $B, R_{in}, R_{out}$  we say that a circuit specification  $\mathcal{S} = (L_f, \text{BAD}_{in}, \text{BAD}_{out})$  on  $\text{GRID}_N$  is  $(B, R_{in}, R_{out})$ -*bounded* if the following hold:  $|\text{BAD}_{in}| \leq R_{in}$ ,  $|\text{BAD}_{out}| \leq R_{out}$ , and for every  $u \in \text{GRID}_N \setminus \text{BAD}_{in}$  it holds that  $|L_f(u)| \leq B$ .

Given a fixed circuit specification, the following definition captures the conditions that are required for an input pattern so that the circuit game associated with that input pattern can be reduced to a nonlocal game (the reduction is explained in Sect. 5).



The intuition to keep in mind for the definition is as follows: each player in the nonlocal game receives her input from one of the input locations and puts her output along the paths of the star. Each player also puts some outputs inside their box of the star. In order for it to be possible to implement the strategy locally, we must have the outputs of each player be causally independent of the inputs of the other players. We ensure this by checking that the forward lightcone of one player's input misses the locations of the other players' outputs.

**Definition 4.9.** (*Causality-respecting patterns*) Let  $\mathcal{S} = (L_f, \text{BAD}_{in}, \text{BAD}_{out})$  be a circuit specification. Let  $\mathcal{P} = \{(u^{(i)}, \Gamma^{(i)})_i\}$  be an input pattern. We say that a pair  $(u^{(i)}, \Gamma^{(i)})$  is *individually- $\mathcal{S}$ -causal* with respect to  $\mathcal{P}$  if the following hold:<sup>4</sup>

- (a) For each  $k$ , the forward lightcone of  $u_k^{(i)}$  misses  $\Gamma^{(i)}$ , except possibly near  $u_k^{(i)}$ . More precisely,  $L_f(u_k^{(i)}) \cap \Gamma^{(i)} \subseteq \text{Box}(u_k^{(i)})$ .
- (b) For all  $(u^{(j)}, \Gamma^{(j)}) \in \mathcal{P}$  (with  $j \neq i$ ) and for all  $k$ , the forward lightcone of  $u_k^{(j)}$  misses  $\Gamma^{(i)}$  entirely, i.e.  $L_f(u_k^{(j)}) \cap \Gamma^{(i)} = \emptyset$ .
- (c)  $\Gamma^{(i)}$  misses  $\text{BAD}_{out}$ , i.e.  $\Gamma^{(i)} \cap \text{BAD}_{out} = \emptyset$ .

Furthermore, we say that a pair  $(u, \Gamma)$  is  *$\mathcal{S}$ -valid* if the following conditions hold.

- (d) Every input location lies outside of  $\text{BAD}_{in}$ , i.e.  $u_k^{(i)} \cap \text{BAD}_{in} = \emptyset$  for all  $k, i$ .

We say that an input pattern  $\mathcal{P}$  is  *$\mathcal{S}$ -causal* if every  $(u^{(i)}, \Gamma^{(i)}) \in \mathcal{P}$  is individually- $\mathcal{S}$ -causal and  $\mathcal{S}$ -valid with respect to  $\mathcal{P}$ .

Finally, we introduce a distribution on input patterns so that for any circuit specification  $\mathcal{S}$  that is  $(B, R_{in}, R_{out})$ -bounded for sufficiently small parameters  $B, R_{in}$ , and  $R_{out}$ , a sample from the distribution gives an  $\mathcal{S}$ -causal pattern with high probability (see Sects. 4.3 and 4.4).

**Definition 4.10.** (*Random input patterns*) Let  $L, N \geq 1$ ,  $\ell \geq 1$ , and  $1 \leq r \leq N$  be integer such that  $3L\sqrt{\ell+1} \leq M = \lfloor N/\sqrt{r} \rfloor$ . Divide  $\text{GRID}_N$  in  $r$  disjoint squares  $S^{(1)}, \dots, S^{(r)}$  of side length  $M$  each.<sup>5</sup> Partition each square into  $T = \lfloor \frac{M}{2L+1} \rfloor^2$  boxes of side length  $(2L+1)$ , in an arbitrary way. For each possible choice of  $(\ell+1)$  distinct boxes  $b_0, b_1, \dots, b_\ell$  within a square, fix a collection  $\text{STARS}(b_0, \dots, b_\ell)$  of  $L/\ell$  stars such that

- each star has  $b_0$  as its central box and  $b_1, \dots, b_\ell$  as its other boxes,
- the total length of the paths in any star is at most  $2\ell M$ , and
- the paths of the distinct stars are vertex-disjoint.

Consider the following distribution  $\mathcal{D}^{(r)}(N, L)$  on input patterns on  $\text{GRID}_N$ . For each  $i \in \{1, \dots, r\}$  select  $b_0^{(i)}, \dots, b_\ell^{(i)}$  uniformly at random among the  $T$  boxes that partition the  $i$ th square  $S^{(i)}$ , for  $j \in \{1, \dots, \ell\}$ , a vertex  $u_j^{(i)}$  uniformly at random within the  $j$ th selected box. Finally, select a star  $\Gamma^{(i)} \in \text{STARS}(b_0^{(i)}, b_1^{(i)}, \dots, b_\ell^{(i)})$  uniformly at random. Return the input pattern  $\mathcal{P} = \{(u^{(i)}, \Gamma^{(i)})\}_{1 \leq i \leq r}$ .

<sup>4</sup> Recall that we identify a star  $\Gamma$  with the union of the vertex sets of its paths and boxes.

<sup>5</sup> It does not matter where these squares are located, as long as they do not overlap.

**4.3. Single-input patterns.** We would like to show that patterns in the support of  $\mathcal{D}^{(r)}$  are “very nearly”  $\mathcal{S}$ -causal for most  $\mathcal{S}$  in the sense that removing only an exponentially small fraction of inputs yields an  $\mathcal{S}$ -causal pattern. To warm up, we argue that for any  $(B, R_{in}, R_{out})$ -bounded circuit specification  $\mathcal{S}$ , an input pattern sampled from the distribution  $\mathcal{D}^{(1)}$  introduced in Definition 4.10 is  $\mathcal{S}$ -causal with high probability. We use this single-input analysis later to show that in a many-input pattern, most of the inputs are individually- $\mathcal{S}$ -causal.

In this subsection only, we use  $M$  instead of  $N$  to denote the grid size. We do this because the distribution  $\mathcal{D}^{(r)}(N, L)$  can be (informally) thought of as the direct product of  $r$  copies of  $\mathcal{D}^{(1)}(M, L)$ , and the former is of greater interest to us.

**Lemma 4.11.** *Let  $M \geq 1$ ,  $1 \leq B, L \leq M/4$  and  $0 \leq R_{in}, R_{out} \leq M^2$  be integer. Let  $\mathcal{S} = (L_f, \text{BAD}_{in}, \text{BAD}_{out})$  be a circuit specification for  $\text{GRID}_M$  that is  $(B, R_{in}, R_{out})$ -bounded. Let  $\mathcal{P} = \{(u, \Gamma)\}$  be drawn from the distribution  $\mathcal{D}^{(1)}$  introduced in Definition 4.10. Then the probability that the unique pair  $(u, \Gamma)$  in  $\mathcal{P}$  is not individually- $\mathcal{S}$ -causal with respect to  $\mathcal{P}$  is  $O(L^2(R_{out} + B)/M^2 + (R_{out} + B)/L)$ . Moreover, the probability that  $\mathcal{P}$  is not  $\mathcal{S}$ -valid is  $O(R_{in}/M^2)$ . Overall, the probability that  $\mathcal{P}$  is not  $\mathcal{S}$ -causal is at most*

$$O\left(L^2(R_{out} + B)/M^2 + (R_{out} + B)/L + R_{in}/M^2\right), \quad (11)$$

where the  $O$  notation hides factors polynomial in  $\ell$ .

*Proof.* We check all conditions in Definition 4.9. Since  $\mathcal{P}$  contains only one (input, star) pair, condition (b) (which restricts the interactions between pairs) is satisfied automatically.

Now we check conditions (a) and (c). Call a box bad if it intersects  $\text{BAD}_{out}$ . Under  $\mathcal{D}^{(1)}$  there are  $\lfloor M/(2L+1) \rfloor^2 \geq 1/4(M/2L)^2$  possible box locations. By a union bound, the probability that any box is bad is at most  $16L^2R_{out}/M^2 = O(L^2R_{out}/M^2)$ . There are at least  $L/\ell$  possible choices for the paths of  $\Gamma$ . Since all such choices are disjoint, again by a union bound the probability that  $\Gamma \cap X \neq \emptyset$  for some subset  $X$  is at most  $\ell|X|/L$ . Letting  $X = \text{BAD}_{out} \cup \bigcup_i L_f(u_i)$ , so that  $|X| \leq R_{out} + \ell B$ , we see that the probability of violating condition (a) or condition (c) is at most  $O\left(\frac{\ell R_{out} + \ell^2 B}{L}\right)$ .

Similarly, since the  $u_j$  are chosen independently, for any  $u \neq u' \in \{u_0, u_1, \dots, u_\ell\}$  the probability that  $L_f(u) \cap \text{Box}(u') \neq \emptyset$  is  $16L^2B/M^2 = O(L^2B/M^2)$ .

Finally we check condition (d). Any  $u_j$  is chosen independently among  $(2L+1)^2 \geq M^2/8 = \Omega(M^2)$  possibilities, so the probability that  $u_j \in \text{BAD}_{in}$  is at most  $8R_{in}/M^2 = O(R_{in}/M^2)$ ; we conclude by the union bound, and absorb the parameter  $\ell$  in the  $O(\cdot)$ .  $\square$

**4.4. Arbitrary-input patterns.** We extend the argument from the previous section to the case where the input pattern contains more than one input.

**Lemma 4.12.** (Random input patterns are usually causal) *Let  $N \geq 1$ ,  $1 \leq r \leq N$  and  $1 \leq B, R_{in}, L \leq N/4$  be integer. Let  $\mathcal{S} = (L_f, \text{BAD}_{in}, \emptyset)$  be a circuit specification for  $\text{GRID}_N$  that is  $(B, R_{in}, 0)$ -bounded. Then the probability that an input pattern  $\mathcal{P}$  chosen according to  $\mathcal{D}^{(r)}(N, L)$  (as defined in Definition 4.10) is not  $\mathcal{S}$ -causal is at most  $O(r^2B(r(L^2 + R_{in})/N^2 + 1/L))$ .*

*Proof.* Let  $\mathcal{P} = \{(u^{(i)}, \Gamma^{(i)})\}$  be an input pattern chosen according to  $\mathcal{D}^{(r)}(M, L)$ . For  $i \in \{1, \dots, r\}$  we let  $\mathcal{P}^{(i)}$  be the single-pair pattern  $\{(u^{(i)}, \Gamma^{(i)})\}$ . Let  $X_i$  be the indicator variable that the pair  $(u^{(i)}, \Gamma^{(i)})$  is not individually- $\mathcal{S}$ -causal with respect to  $\mathcal{P}$ . Let  $Y_i$  be the indicator that  $(u^{(i)}, \Gamma^{(i)})$  is not  $\mathcal{S}$ -valid. To see that  $\mathcal{P}$  is  $\mathcal{S}$ -causal, it suffices to check that each input is  $\mathcal{S}$ -valid and individually- $\mathcal{S}$ -causal with respect to  $\mathcal{P}$ . This is true if and only if  $\sum_i X_i = 0$  and  $\sum_i Y_i = 0$ . We first bound the latter event.

**Claim 4.13.** *It holds that*

$$\Pr\left(\sum_j Y_j \neq 0\right) = O(r^2 R_{in}/N^2). \quad (12)$$

*Proof.* Applying the second bound in Lemma 4.11 and using that  $M = \lfloor N/\sqrt{r} \rfloor$  it follows that for any  $i \in \{1, \dots, r\}$ ,

$$\Pr(Y_i \neq 0) = \Pr(\mathcal{P}^{(i)} \text{ is not } \mathcal{S}\text{-valid}) \leq 8R_{in}/M^2 = O(rR_{in}/N^2).$$

The claim follows by a union bound over the  $r$  patterns  $\mathcal{P}^{(i)}$ .  $\square$

Next we turn to the  $X_i$ .

**Claim 4.14.**

$$\Pr\left(\sum_i X_i \neq 0\right) = O(r^2 B(rL^2/N^2 + 1/L)) + \sum_i \Pr\left(\sum_{j \neq i} Y_j \neq 0\right). \quad (13)$$

*Proof.* For any  $i \in \{1, \dots, r\}$  let

$$\text{BAD}_{out}^{(i)} = \bigcup_{k \neq i} \left( \cup_j L_f(u_j^{(k)}) \right),$$

and define a specification  $\mathcal{S}^{(i)} = (L_f, \text{BAD}_{in}, \text{BAD}_{out}^{(i)})$ . With these definitions, it follows that

$$\begin{aligned} \Pr(X_i = 0) &= \Pr(\mathcal{P}^{(i)} \text{ is individually-}\mathcal{S}\text{-causal}) \\ &\geq \Pr(\mathcal{P}^{(i)} \text{ is individually-}\mathcal{S}^{(i)}\text{-causal}). \end{aligned} \quad (14)$$

Indeed condition (c) of being individually- $\mathcal{S}^{(i)}$ -causal (see Definition 4.9) implies all conditions of being individually- $\mathcal{S}$ -causal for  $\mathcal{S} = (L_f, \text{BAD}_{in}, \emptyset)$ .

In the event that  $\mathcal{P}^{(j)}$  is  $\mathcal{S}$ -valid for all  $j \neq i$  (that is, when  $\sum_{j \neq i} Y_j = 0$ ) we know that  $L_f(u_j^{(k)}) \leq B$  for each  $(j, k) \in \{1, \dots, \ell\} \times \{1, \dots, r\}$ , and thus that  $|\text{BAD}_{out}^{(i)}| \leq \ell r B = O(rB)$ . Using that the marginal distribution of a single pair  $(u^{(i)}, \Gamma^{(i)})$  from  $\mathcal{P}$  is equal to  $\mathcal{D}^{(1)}(M, L)$  (when seen as a distribution on the square  $\mathcal{S}^{(i)}$  associated with  $(u^{(i)}, \Gamma^{(i)})$ ), it follows from the bound in Lemma 4.11 that for any  $i \in \{1, \dots, r\}$ ,

$$\Pr\left(X_i \neq 0 \mid \sum_{j \neq i} Y_j = 0\right) = O(rB(rL^2/N^2 + 1/L)). \quad (15)$$

Applying the union bound,

$$\begin{aligned}
\Pr\left(\sum_i X_i \neq 0\right) &\leq \sum_i \Pr(X_i \neq 0) \\
&\leq \sum_i \Pr\left(X_i \neq 0 \mid \sum_{j \neq i} Y_j = 0\right) + \Pr\left(\sum_{j \neq i} Y_j \neq 0\right) \\
&\leq O(r^2 B(rL^2/N^2 + 1/L)) + \sum_i \Pr\left(\sum_{j \neq i} Y_j \neq 0\right),
\end{aligned}$$

where the last line follows from (15).  $\square$

To conclude the proof of the lemma we write

$$\begin{aligned}
\Pr(\mathcal{P} \text{ is not } \mathcal{S}\text{-causal}) &= \Pr\left(\sum_i X_i + Y_i \neq 0\right) \\
&\leq \Pr\left(\sum_i X_i \neq 0\right) + \Pr\left(\sum_j Y_j \neq 0\right) \\
&\leq O(r^2 B(rL^2/N^2 + 1/L)) + \sum_i \Pr\left(\sum_{j \neq i} Y_j \neq 0\right) + O(r^2 R_{in}/N^2) \\
&\leq O(r^2 B(rL^2/N^2 + 1/L)) + O(r^3 R_{in}/N^2) + O(r^2 R_{in}/N^2) \\
&= O\left(r^2 B(r(L^2 + R_{in})/N^2 + 1/L)\right),
\end{aligned}$$

where the third line uses (13) and (12) and the fourth uses (12).  $\square$

The previous lemma shows that a random input pattern  $\mathcal{P}$  is  $\mathcal{S}$ -causal with high probability. In this case we can define a game from  $\mathcal{P}$  so that in the game, a shallow circuit with specification  $\mathcal{S}$  can be simulated by a set of spacelike-separated players. This simulation is perfect when  $\mathcal{P}$  is exactly  $\mathcal{S}$ -causal. More generally, a weaker simulation argument still applies if a small constant fraction of inputs in  $\mathcal{P}$  are not  $\mathcal{S}$ -causal. The next lemma shows that this condition can be guaranteed to hold with much higher probability, exponentially close to 1 rather than inverse-polynomially close. This bound will be used in the proof of Theorem 1.2.

**Lemma 4.15.** (Random input patterns are mostly causal with high probability) *Let  $N \geq 1$ ,  $1 \leq r \leq N$  and  $1 \leq B, R_{in}, L \leq N/4$  be integer. Let  $\mathcal{S} = (L_f, \text{BAD}_{in}, \emptyset)$  be a circuit specification for  $\text{GRID}_N$  that is  $(B, R_{in}, 0)$ -bounded. Consider an input pattern  $\mathcal{P}$  chosen according to  $\mathcal{D}^{(r)}(N, L)$ . Let*

$$\begin{aligned}
\mathcal{P}_{VAL} &= \{(u, \Gamma) \in \mathcal{P} \mid (u, \Gamma) \text{ is } \mathcal{S}\text{-valid}\}, \\
\mathcal{P}_{CAUS} &= \{(u, \Gamma) \in \mathcal{P} \mid (u, \Gamma) \text{ is individually-}\mathcal{S}\text{-causal with respect to } \mathcal{P}_{VAL}\}.
\end{aligned}$$

*Then there exist universal constants  $C, C' > 0$  such that if  $p = C'rB(rL^2/N^2 + \ell/L)$  then for any  $t > 0$ ,*

$$\Pr\left(|\mathcal{P}_{CAUS}| \geq r(1-p) - 2t\right) \geq 1 - 2\exp\left(-t^2/8r\right), \quad (16)$$

$$\Pr\left(|\mathcal{P}_{VAL}| \geq r(1 - CrR_{in}/N^2) - t\right) \geq 1 - 2\exp\left(-2t^2/r\right). \quad (17)$$

For later convenience we note that (16) and (17) can be combined by a union bound to obtain

$$\Pr\left(|\mathcal{P}_{VAL} \cap \mathcal{P}_{CAUS}| \geq r(1 - CrR_{in}/N^2 - 2p) - 3t\right) \geq 1 - 4 \exp\left(-\frac{t^2}{8r}\right). \quad (18)$$

*Proof.* The proof relies on concentration arguments to bound the probabilities in (16) and (17). The second bound, (17), is easier to show, because it can be expressed as a bound on a sum of independent random variables. The following claim establishes the bound.

**Claim 4.16.** *There is a universal constant  $C > 0$  such that for any  $t > 0$ ,*

$$\Pr\left(|\mathcal{P}_{VAL}| \geq r(1 - CrR_{in}/N^2) - t\right) \geq 1 - 2 \exp\left(-\frac{2t^2}{r}\right).$$

*Proof.* For  $i \in \{1, \dots, r\}$  let  $V_i$  be the indicator variable for the event that  $(u^{(i)}, \Gamma^{(i)})$  is  $\mathcal{S}$ -valid. Since in  $\mathcal{D}^{(r)}(N, L)$  the  $(u^{(i)}, \Gamma^{(i)})$  are chosen independently within the disjoint squares  $S^{(i)}$ , it follows from the definition of  $\mathcal{S}$ -valid that the  $V_i$  are independent. Using the bound shown in Lemma 4.11 it follows that for any  $i \in \{1, \dots, r\}$ ,

$$\mathbb{E}[V_i] = 1 - CR_{in}/M^2, \quad (19)$$

for some constant  $C > 0$  and where  $M = \lfloor N/\sqrt{r} \rfloor$ . We conclude by applying Hoeffding's inequality:

$$\begin{aligned} \Pr\left(|\mathcal{P}_{VAL}| \leq r(1 - CR_{in}/M^2) - t\right) &= \Pr\left(\sum_{i=1}^r V_i \leq r(1 - CR_{in}/M^2) - t\right) \\ &\leq \Pr\left(\frac{1}{r} \left| \sum_{i=1}^r V_i - \sum_{i=1}^r \mathbb{E}[V_i] \right| \geq t/r\right) \\ &\leq 2 \exp\left(-2t^2/r\right). \end{aligned}$$

□

The proof of the remaining bound (16) is made a little delicate by the fact that the condition that a pair  $(u, \Gamma) \in \mathcal{P}$  is individually- $\mathcal{S}$ -causal is a global condition, so that  $\mathcal{P}_{CAUS}$  is not directly expressible as a sum of independent random variables. To get around this, we first make a few definitions.

Let  $\mathcal{P} = \{(u^{(i)}, \Gamma^{(i)})\}$  be an input pattern chosen at random according to the distribution  $\mathcal{D}^{(r)}(M, L)$ . For  $i \in \{1, \dots, r\}$  define

$$\text{BAD}_{out}^{<i} = \bigcup_{\substack{k < i \text{ s.t.} \\ (u^{(k)}, \Gamma^{(k)}) \text{ is } \mathcal{S}\text{-valid}}} \left( \bigcup_j L_f(u_j^{(k)}) \right), \quad \text{BAD}_{out}^{>i} = \bigcup_{\substack{k > i \text{ s.t.} \\ (u^{(k)}, \Gamma^{(k)}) \text{ is } \mathcal{S}\text{-valid}}} \left( \bigcup_j L_f(u_j^{(k)}) \right),$$

and  $\text{BAD}_{out}^{(i)} = \text{BAD}_{out}^{<i} \cup \text{BAD}_{out}^{>i}$ . From the assumption that the circuit specification  $\mathcal{S}$  is  $(B, R_{in}, 0)$ -bounded, and since  $\text{BAD}_{out}^{(i)}$  is defined as a union of the lightcones of only the *valid* pairs  $(u^{(k)}, \Gamma^{(k)})$ , it follows that for all  $i$ ,  $|\text{BAD}_{out}^{(i)}|, |\text{BAD}_{out}^{<i}|, |\text{BAD}_{out}^{>i}|$  are each at most  $rB$ .

Recall that in  $\mathcal{D}^{(r)}(M, L)$  each  $u^{(i)}$  is chosen within a square  $S^{(i)}$  of side length  $M = \lfloor N/\sqrt{r} \rfloor$ . We identify  $S^{(i)}$  with  $\text{GRID}_M^{(i)}$ , and introduce a single-pair pattern  $\mathcal{P}^{(i)} = \{(u^{(i)}, \Gamma^{(i)})\}$  that we think of as a pattern on  $\text{GRID}_M^{(i)}$ . We further define a specification

$$\mathcal{S}^{(i)} = (L_f, \text{BAD}_{in}, \text{BAD}_{out}^{(i)})$$

on  $\text{GRID}_M^{(i)}$ .

Let  $X_i$  be the indicator variable for the event that the pair  $(u^{(i)}, \Gamma^{(i)})$  is not individually- $\mathcal{S}^{(i)}$ -causal with respect to  $\mathcal{P}_{VAL}$ . The following claim relates the  $X_i$  to  $\mathcal{P}_{CAUS}$ .

**Claim 4.17.** *It holds that  $|\mathcal{P}_{CAUS}| \geq \sum_{i=1}^r X_i$ .*

*Proof.* Note that whenever  $\mathcal{P}^{(i)}$  is individually- $\mathcal{S}^{(i)}$ -causal with respect to  $\mathcal{P}_{VAL}$ , it is also individually- $\mathcal{S}$ -causal with respect to  $\mathcal{P}_{VAL}$ . This follows by noting that for the given definition of  $\text{BAD}_{out}^{(i)}$ , condition (c) of being individually- $\mathcal{S}^{(i)}$ -causal with respect to  $\mathcal{P}$  implies conditions (b) and (c) of being individually- $\mathcal{S}$ -causal with respect to  $\mathcal{P}_{VAL}$ . Therefore,  $\sum_i X_i$  is an upper bound on the number of  $\mathcal{P}^{(i)}$  which are not individually- $\mathcal{S}$ -causal with respect to  $\mathcal{P}_{VAL}$ .  $\square$

The previous claim reduces our task to showing a high-probability lower bound on  $\sum X_i$ . The random variables  $X_i$  are dependent. To obtain a bound, we apply a Martingale argument to two related sequences of random variables, defined as follows. First introduce specifications

$$\mathcal{S}^{(<i)} = (L_f, \text{BAD}_{in}, \text{BAD}_{out}^{<i}) \quad \text{and} \quad \mathcal{S}^{(>r-i)} = (L_f, \text{BAD}_{in}, \text{BAD}_{out}^{>(r-i)})$$

on grid  $\text{GRID}_M^{(i)}$ , let  $Y_i$  (resp.  $Z_i$ ) as the indicator variable for the event that  $(u^{(i)}, \Gamma^{(i)})$  is not individually- $\mathcal{S}^{(<i)}$ -causal (resp. individually- $\mathcal{S}^{(>r-i)}$ -causal) with respect to  $\mathcal{P}^{(i)}$ . The next claim relates  $Y_i$  and  $Z_{r-i}$  to  $X_i$ .

**Claim 4.18.** *For each  $i \in \{1, \dots, r\}$ , it holds that  $X_i = Y_i \vee Z_{r-i}$ .*

*Proof.* The claim follows by noting that, in Definition 4.9, the conditions (a) for  $X_i$ ,  $Y_i$  and  $Z_{r-i}$  are equivalent. Condition (b) for  $X_i$  is equivalent to the event that condition (c) for both  $Y_i$  and  $Z_{r-i}$  is true. Finally, condition (c) for  $X_i$  is vacuous, and conditions (b) for both  $Y_i$  and  $Z_{r-i}$  are vacuous.  $\square$

The following claim almost finishes the proof.

**Claim 4.19.** *There is a universal constant  $C' > 0$  such that if  $p = C'rB(rL^2/N^2 + \ell/L)$  then for any  $t > 0$ ,*

$$\Pr\left(\sum_{i=1}^r Y_i \geq t + rp\right) \leq \exp\left(\frac{-t^2}{2r}\right),$$

$$\Pr\left(\sum_{i=1}^r Z_i \geq t + rp\right) \leq \exp\left(\frac{-t^2}{2r}\right).$$

*Proof.* The proof is based on a Martingale tail bound. For any  $i \in \{1, \dots, r\}$ , let  $Y_{<i} = \{Y_k | k < i\}$  and  $Z_{>i} = \{Z_k | k > i\}$ . Note that  $Y_i$  (resp.  $Z_i$ ) depends only on the underlying circuit  $\mathcal{C}$  together with the selection of pairs in squares  $S^{(j)}$  for  $j \leq i$  (resp.  $j \geq i$ ). It follows from the bound shown in Lemma 4.11 that

$$\mathbb{E}[Y_i | Y_{<i}] = O(rB(rL^2/N^2 + \ell/L)), \quad (20)$$

and similarly

$$\mathbb{E}[Z_i | Z_{>i}] = O(rB(rL^2/N^2 + \ell/L)). \quad (21)$$

Let  $p$  denote the maximum of the bounds on the right-hand side of (20) and (21), and assume  $p \leq 1$ . For any  $n \in \{1, \dots, r\}$  define  $\bar{Y}_n = \sum_{i=1}^n Y_i - np$  and  $\bar{Z}_{r-n} = \sum_{i=r-n}^r Z_i - np$ . Then

$$\mathbb{E}[\bar{Y}_n | Y_{<n}] = \mathbb{E}[Y_n - p + \bar{Y}_{n-1} | Y_{<n}] \leq \bar{Y}_{n-1},$$

and

$$\mathbb{E}[\bar{Z}_{r-n} | Z_{>r-n}] = \mathbb{E}[Z_{r-n} - p + \bar{Z}_{r-n+1} | Z_{>r-n}] \leq \bar{Z}_{r-(n-1)}.$$

Additionally, it always holds that

$$|\bar{Y}_n - \bar{Y}_{n-1}| \leq |Y_n - p| \leq \max(1 - p, p) \leq 1,$$

where the last inequality follows from the assumption that  $p \leq 1$ . Similarly,

$$|\bar{Z}_{r-n} - \bar{Z}_{r-(n-1)}| \leq |Z_{r-n} - p| \leq \max(1 - p, p) \leq 1.$$

Thus, both  $\bar{Y}_n$ , and  $\bar{Z}_{r-n}$  form super-martingale sequences for increasing  $n$ . Defining  $\bar{Y}_0 = \bar{Z}_r = 0$  and applying Azuma's inequality gives

$$\Pr(\bar{Y}_n - \bar{Y}_0 \geq t) = \Pr(\bar{Y}_n \geq t) \leq \exp\left(\frac{-t^2}{2n \max(1 - p, p)^2}\right) \leq \exp\left(\frac{-t^2}{2n}\right),$$

and

$$\Pr(\bar{Z}_{r-n} - \bar{Z}_r \geq t) = \Pr(\bar{Z}_{r-n} \geq t) \leq \exp\left(\frac{-t^2}{2n \max(1 - p, p)^2}\right) \leq \exp\left(\frac{-t^2}{2n}\right).$$

Setting  $n = r$  proves the claim.  $\square$

Using Claim 4.18 it follows that  $\sum_{i=1}^r X_i \leq \sum_{i=1}^r Y_i + \sum_{i=1}^r Z_i$ . Therefore, for any  $t > 0$

$$\begin{aligned} \Pr\left(\sum_{i=1}^r X_i \geq 2t + 2rp\right) &\leq \Pr\left(\sum_{i=1}^r Y_i + Z_i \geq 2t + 2rp\right) \\ &\leq \Pr\left(\sum_{i=1}^r Y_i \geq t + rp\right) + \Pr\left(\sum_{i=1}^r Z_i \geq t + rp\right) \\ &\leq 2 \exp\left(\frac{-t^2}{2r}\right), \end{aligned}$$

where the last inequality follows from Claim 4.19. Replacing  $t$  with  $t/2$  and using Claim 4.17 proves (16).  $\square$

**4.5. Derandomization.** Lemma 4.12 states that if a pattern is chosen according to the distribution  $\mathcal{D}^{(r)}(M, L)$ , then it is  $\mathcal{S}$ -causal with probability that is close to 1, regardless of the choice of  $\mathcal{S}$ . The following lemma shows that it is possible to partially derandomize the distribution, at little loss in the success probability.

**Lemma 4.20.** *Let  $N \geq 1$  be an integer and  $\eta > 0$ . Let  $L, B, R_{in}, r$  be integer such that*

$$B = O(N^\eta), \quad L = O(N^{4/7+\eta}), \quad \text{and} \quad r = O(N^{2/7-2\eta}), \quad (22)$$

*Then using  $O(\log^2 N)$  uniformly random bits it is possible to sample from a distribution  $\tilde{\mathcal{D}}^{(r)}$  on input patterns  $\mathcal{P}$  for  $\text{GRID}_N$  such that for any circuit specification  $\mathcal{S}$  that is  $(B, R_{in}, 0)$ -bounded,  $\mathcal{P}$  is  $\mathcal{S}$ -causal with probability  $1 - O(N^{-\eta})$ , and a random  $(u, \Gamma) \in \mathcal{P}$  is valid with probability  $O(R_{in}/N^2 + N^{-\eta})$ .*

*Proof.* The choice of parameters made in the lemma is such that  $r^3 B L^2 / N^2 = O(N^{-\eta})$  and  $r^2 B / L = O(N^{-\eta})$ , so Lemma 4.12 gives that  $\mathcal{P}$  sampled according to  $\mathcal{D}^{(r)}$  is  $\mathcal{S}$ -causal with probability  $1 - O(N^{-\eta})$ , and a random  $(u, \Gamma) \in \mathcal{P}$  is valid with probability  $O(R_{in}/N^2)$ .

A pattern in the support of  $\mathcal{D}^{(r)}(M, L)$  can be specified using  $O(r \log(N))$  uniformly random bits: for each  $i \in \{1, \dots, r\}$ , there are  $O(\log N)$  random bits to specify the locations of the  $u_j^{(i)}$ , and  $O(\log N)$  additional bits to specify the star that connects the  $u_j^{(i)}$ . (Recall we fixed a small set of possible stars when we defined the distribution.) Given any choice of such random bits, and a fixed circuit specification, by Savitch's theorem it is possible to decide whether the pattern is  $\mathcal{S}$ -causal in  $O(\log^2 N)$  space, given read-only access to the circuit graph determining  $\mathcal{S}$ . This allows us to apply the INW pseudorandom generator for small-space circuits [INW94] with  $O(\log^2 N)$  seed to obtain the claimed result, with the additional error  $O(N^{-\eta})$  being due to the pseudorandom generator.  $\square$

## 5. Circuit Games

Let  $N \geq 1$  be an integer grid size. Let  $r \geq 1$  be an integer number of repetitions. Let  $\mathcal{G}$  be an  $(\ell, k, m)$  stabilizer game. In this section we design a circuit game  $G = G_{\mathcal{G}, N, r}$  associated with  $(\mathcal{G}, N, r)$  in a way that the circuit game has similar completeness and soundness properties as  $\mathcal{G}$  (more precisely, as a rotated, stretched game obtained from  $\mathcal{G}$ , using the stars in an input pattern  $\mathcal{P}$  associated with  $(\mathcal{G}, N, r)$  that is provided as input to the circuit to define the length of the stretches; see Sect. 3.2 for the definition of rotated and stretched games).

We first give a general definition that specifies what we mean by a ‘‘circuit game’’.

**Definition 5.1.** (*Circuit game*) Given input and output sets  $\mathcal{I}, \mathcal{O}$  respectively, a *circuit game* is a relation  $\mathcal{R} \subseteq \mathcal{I} \times \mathcal{O}$ , together with a probability distribution  $\pi$  on  $\mathcal{I}$ . We say that a circuit  $\mathcal{C}$  wins the circuit game  $(\mathcal{R}, \pi)$  with probability  $p$  if, on average over an input  $x \in \mathcal{I}$  sampled according to  $\pi$ , the circuit returns an output  $y \in \mathcal{O}$  such that  $(x, y) \in \mathcal{R}$  with probability  $p$ .

To specify the relation associated with the circuit game we will construct from  $\mathcal{G}$  it is convenient to first introduce a quantum circuit that succeeds in the game with certainty. This is done in Sect. 5.1. In Sect. 5.2 we give the definition of the circuit game.



*5.1. Definition and completeness.* Informally, the game  $G_{\mathcal{G},N,r}$  is obtained by “planting  $r$  copies of  $\mathcal{G}$  in the grid  $\text{GRID}_N$ ”. Let  $\mathcal{P}$  be an input pattern associated with  $(\mathcal{G}, N, r)$  (see Definition 4.5). Let  $k = \max_j k_j$  be the maximum number of qudits used by a player in the honest strategy for  $\mathcal{G}$ ,  $|\psi\rangle$  the  $(k\ell)$ -qubit state used in the strategy (padded if needed), and  $D$  the depth of a circuit that prepares  $|\psi\rangle$  from  $|0\rangle$ . Assume that  $D \geq 2$ . We describe a depth  $(D + 1)$  quantum circuit  $\mathcal{C}_{ideal}$  that takes an input from

$$\mathcal{I} = \{\mathcal{P} : \text{input pattern for } \mathcal{G}\} \times \{\{(x_1^{(i)}, \dots, x_\ell^{(i)})\}_{1 \leq i \leq r} : r\text{-tuple of queries in } \mathcal{G}\}, \quad (23)$$

and returns a string in the output set

$$\mathcal{O} = \prod_{i=1}^r \prod_{j=1}^{\ell} \left( (\mathbb{Z}_d^{2k})^{\Gamma_j^{(i)}} \times (\mathbb{Z}_d^{m_j})^{u_j^{(i)}} \right). \quad (24)$$

In (24) we have used the vertices in  $\Gamma_j^{(i)}$  and the  $u_j^{(i)}$  to label indices of the elements of  $\mathcal{O}$ . Note that these are always distinct. In Sect. 5.1.2 we show how to modify the format for the input and the output in a way that the circuit can be made geometrically local on a 2D grid.

The computation performed by the circuit  $\mathcal{C}_{ideal}$  proceeds in three stages:

- In the first stage, the circuit initializes a lattice of qudits as follows. Each vertex in  $\text{GRID}_N$  is associated with  $4k$  qudits, organized in 4 groups of  $k$  that we call the “left”, “right”, “top” and “bottom” groups associated with that vertex. Each of these groups is initialized in a maximally entangled state with the group from the neighboring grid vertex that is closest to it, i.e. the “top” group at vertex  $(i, j)$  is associated with the “bottom” group at vertex  $(i, j + 1)$ , etc. In addition, for each center location  $g^{(i)}$  of a star  $\Gamma^{(i)}$  in  $\mathcal{P}$ , the circuit creates the state  $|\psi\rangle$  on the  $(k\ell)$  qudits associated with the  $\ell$  vertices  $(g_1^{(i)}, \dots, g_\ell^{(i)})$ ; for each vertex  $g_j^{(i)}$ , a group of qudits is used that is not connected to the next vertex in the path  $\Gamma_j^{(i)}$ . (This replaces the creation of the maximally entangled state, for that group of qudits.) This step can be implemented in depth  $\max(2, D)$ .
- In the second stage, the circuit implements an entanglement transfer protocol as described in Sect. 5.1.1, using each of the  $\ell$  simple paths that form a star  $\Gamma^{(i)}$  from  $\mathcal{P}$  to route the qudits of  $|\psi\rangle$ . The measurement outcomes in  $(\mathbb{Z}_d)^{2k}$  from the teleportation measurements obtained at each vertex in  $\Gamma^{(i)}$  are recorded at the location at which they are obtained, and will eventually form part of the output of the circuit. This step can be completed in depth 1.
- In the last stage, the circuit implements the honest quantum strategy for the game  $\mathcal{G}$ , using locations  $u_j^{(i)}$  indicated in  $\mathcal{P}$  to specify the  $k$  qudits to be used by the  $j$ th player (the group used is the one closest to the endpoint of the path  $\Gamma_j^{(i)}$ ), and  $x_j^{(i)}$  as the player’s question. The outcomes obtained are returned as part of the output. This step can be completed in depth 1, and can be executed in parallel with the previous step.

The following lemma states that outputs generated by this circuit satisfy the win condition for an associated rotated, stretched game.

**Lemma 5.2.** *Let  $\mathcal{G}$  be an  $(\ell, k, m)$  stabilizer game,  $1 \leq r \leq N$ , and  $\mathcal{P}$  an input pattern associated with  $(\mathcal{G}, N, r)$ . Let  $x^{(1)}, \dots, x^{(r)}$  be an arbitrary tuple of  $r$  queries for  $\mathcal{G}$ .*

For all  $i \in \{1, \dots, r\}$  and  $j \in \{1, \dots, \ell\}$  let  $(r_j^{(i)}, a_j^{(i)}) \in (\mathbb{Z}_d^{2k})^{\Gamma_j^{(i)}} \times \mathbb{Z}_d^{m_j}$  be the outputs generated by an execution of  $\mathcal{C}_{ideal}$  on input  $\mathcal{P}$  and  $(x^{(1)}, \dots, x^{(r)})$ . Then for any  $i \in \{1, \dots, r\}$ ,  $\{(r_j^{(i)}, a_j^{(i)})\}_{j \in \{1, \dots, \ell\}}$  is a valid  $\ell$ -tuple of answers for the players in the rotated stretched game  $\mathcal{G}_{\Gamma^{(i)}}^{S,R}$ , on query  $x^{(i)}$ .

*Proof.* The lemma follows from the definition of  $\mathcal{C}_{ideal}$ , the properties of the entanglement transfer protocol stated in Lemma 5.6, and the definition of the rotated, stretched game.  $\square$

*5.1.1. Entanglement transfer.* We introduce a simple procedure for routing entanglement along a path, such that nearest neighbors on the path have been initialized in a maximally entangled state. This is a standard calculation; for completeness we include the details.

**Lemma 5.3.** (Entanglement transfer I) *Let  $d \geq 2$  and  $|\text{EPR}_d\rangle = \frac{1}{\sqrt{d}} \sum_i |ii\rangle$  a maximally entangled state on  $d$ -dimensional qudits. Let  $a, b \in \mathbb{Z}_d$  and*

$$|\psi\rangle_{ABCD} = (X_A^a \otimes Z_B^b |\text{EPR}_d\rangle_{AB}) \otimes |\text{EPR}_d\rangle_{CD}$$

*a maximally entangled state on four qudits. Then upon measuring the qudits in registers B and C in the Bell basis  $\{X^x \otimes Z^y | \text{EPR}_d\rangle : x, y \leq d\}$ , the post-measurement state is equal (up to global phase) to*

$$(X_A^{a+x} \otimes Z_D^{b-y} |\text{EPR}_d\rangle)_{AD} \otimes (X_B^x \otimes Z_C^y |\text{EPR}_d\rangle)_{BC}. \quad (25)$$

*Proof.* We evaluate the post-measurement state by computing the result of applying the measurement projector onto the state  $X^x \otimes Z^y |\text{EPR}_d\rangle$ . Let  $\omega = e^{2\pi i/d}$ .

$$\begin{aligned} & (X^x \otimes Z^y |\text{EPR}_d\rangle \langle \text{EPR}_d|_{BC} X^{-x} \otimes Z^{-y}) X_A^a \otimes Z_B^b |\text{EPR}_d\rangle_{AB} \otimes |\text{EPR}_d\rangle_{CD} \\ &= X_B^x \otimes Z_C^y \left( \frac{1}{d} \sum_{k,l} |kk\rangle \langle ll| \right) X_B^{-x} \otimes Z_C^{-y} X_A^a \otimes Z_B^b \frac{1}{d} \sum_{i,j} |i\rangle |i\rangle |j\rangle |j\rangle \\ &= X_B^x \otimes Z_C^y \left( \frac{1}{d} \sum_{k,l} |kk\rangle \langle ll| \right) X_B^{-x} \otimes Z_C^{-y} \frac{1}{d} \sum_{i,j} \omega^{bi} |i+a\rangle |i\rangle |j\rangle |j\rangle \\ &= X_B^x \otimes Z_C^y \left( \frac{1}{d} \sum_{k,l} |kk\rangle \langle ll| \right) \frac{1}{d} \sum_{i,j} \omega^{bi-yj} |i+a\rangle |i-x\rangle |j\rangle |j\rangle \\ &= X_B^x \otimes Z_C^y \frac{1}{d^2} \sum_{i,j,k} \omega^{bi-yj} \delta_{i-x,j} |i+a\rangle |k\rangle |k\rangle |j\rangle \\ &= X_B^x \otimes Z_C^y \frac{1}{d^2} \sum_{j,k} \omega^{b(j+x)-yj} |j+x+a\rangle |k\rangle |k\rangle |j\rangle \\ &= \omega^{bx} X_A^{a+x} \otimes Z_D^{b-y} |\text{EPR}_d\rangle_{AD} \otimes X_B^x \otimes Z_C^y |\text{EPR}_d\rangle_{BC}. \end{aligned}$$

$\square$

**Lemma 5.4.** (Entanglement transfer II) *Let  $n \geq 1$ , and  $L_1, \dots, L_n, R_1, \dots, R_n$  qudit registers such that each  $L_i$  is maximally entangled with  $R_i$ . Suppose one performs  $(n - 1)$  Bell basis measurements on qudit pairs  $(R_1, L_2), \dots, (R_{n-1}, L_n)$ , so that the post measurement state of the  $i$ th pair is  $X^{x_i} \otimes Z^{y_i} |\text{EPR}_d\rangle$ . Let  $x = \sum_i x_i$  and  $z = \sum_i z_i$ . Then the post measurement state of the remaining pair  $(L_1, R_n)$  is  $X^x \otimes Z^{-z} |\text{EPR}_d\rangle$ .*

*Proof.* The  $(n - 1)$  Bell basis measurements commute, so we can think of them as being performed in sequence, performing the measurement on  $(R_k, L_{k+1})$  at the  $k$ th step. Using Lemma 5.3 and induction, one can check that after the  $k$ th measurement, the qudits  $(L_1, R_{k+1})$  are in post-measurement state  $X^{\sum_{i=1}^k x_i} \otimes Z^{-\sum_{i=1}^k z_i} |\text{EPR}_d\rangle$ .  $\square$

Suppose that we have prepared a state  $|\phi\rangle$  of  $k$  qudits in one part of the grid and we would like to teleport it to another part of the grid. We show how to design a depth-2 circuit that accomplishes the teleportation.

**Definition 5.5.** (Low-depth state teleportation protocol) *Let  $N \geq 1$  and  $A$  and  $B$  be ordered lists of vertices in  $\text{GRID}_N$  with  $|A| = |B| = k$ . Pick  $\{\Gamma_j\}_{1 \leq j \leq k}$  a set of  $k$  vertex-disjoint, even-length paths on the grid, each with one endpoint in  $A$  and one endpoint in  $B$ . For any  $j \in \{1, \dots, k\}$  denote the vertices of  $\Gamma_j$  as  $v_0, \dots, v_\ell$ , with endpoints  $v_0 \in A, v_\ell \in B$ .*

From this set of paths define a depth-2 measurement circuit as follows: In the first layer, to each of the “odd-even” qudit pairs  $(v_1, v_2), (v_3, v_4), \dots, (v_{\ell-1}, v_\ell)$  apply a gate taking the two-qudit state  $|00\rangle$  to  $|\text{EPR}_d\rangle$ . In the second layer, measure each of the “even-odd” qudit pairs  $(v_0, v_1), (v_2, v_3), \dots, (v_{\ell-2}, v_{\ell-1})$  in the Bell basis and return the  $2\ell$ -bit string of outcomes.

Now we prove that, up to local corrections, and with a technical condition on  $|\phi\rangle$ , our “teleportation circuit” in fact brings  $|\phi\rangle$  from region  $A$  on the grid to region  $B$ .

**Lemma 5.6.** *Suppose that a grid  $\text{GRID}_N$  of qudits is initialized in a state that is in tensor product between the qudits of  $A$  and  $\text{GRID}_N \setminus A$  and equals  $|\psi\rangle = |\phi\rangle_A \otimes |0\rangle_{\text{GRID}_N \setminus A}$ . Furthermore suppose that  $|\phi\rangle$  is locally maximally mixed in the sense that the marginal density matrix on any individual qudit in  $A$  is the maximally mixed state.*

*Suppose we execute the low-depth teleportation described in Definition 5.5. For  $j \in \{1, \dots, k\}$  and  $i \in \{1, \dots, \ell\}$  let  $X^{a_i^j} \otimes Z^{b_i^j} |\text{EPR}_d\rangle$  be the post-measurement state of the  $i$ th pair along the  $j$ th path of the teleportation circuit.*

*Then there is a unitary operator  $P$  such that the post-measurement state  $|\psi'\rangle$  of the grid is  $|\psi'\rangle = (P |\phi\rangle)_B \otimes |\text{aux}\rangle_{\text{GRID}_N \setminus B}$ . Furthermore,  $P$  is a tensor product of  $k$  Pauli operators such that  $P$  acts on the  $j$ th qudit of  $B$  as  $X^{a^j} Z^{-b^j}$ , where  $a^j = \sum_i a_i^j$  and  $b^j$  is defined similarly.*

*Proof.* We analyze the circuit one path at a time. Fix  $i \in \{1, \dots, k\}$ . Let  $x_i \in A$  be one endpoint of  $P_i$  and  $y_i$  the other. Suppose that  $|\phi\rangle$  is any locally maximally mixed state on a set  $C$  of qudits with  $x_i \in C$ . Let  $C' = C \setminus \{x_i\}$ . Then there is a unitary  $V : C' \rightarrow C'_1 \otimes C'_2$  such that

$$I \otimes V |\phi\rangle_{x_i C} = |\text{EPR}_d\rangle_{x_i C'_1} \otimes |\phi'\rangle_{C'_2}. \quad (26)$$

Now suppose we apply  $V$  and then apply the entangling gates and measurements along  $\Gamma_i$ . By Lemma 5.4, the state of qudits  $y_i$  and  $C'_1$  is

$$X^{a^i} \otimes Z^{-b^i} |\text{EPR}_d\rangle_{y_i C'_1} = X^{a^i} Z^{-b^i} \otimes I |\text{EPR}_d\rangle_{y_i C'_1}. \quad (27)$$

The equality in (27) can be verified by noticing that  $Z \otimes Z^\dagger$  stabilizes  $|\text{EPR}_d\rangle$ . Applying  $V^\dagger$  gives

$$(I \otimes V^\dagger)(X^{a^i} Z^{-b^i} \otimes I) |\text{EPR}_d\rangle_{y_i C'_1} \otimes |\phi'\rangle_{C'_2} = (X^{a^i} Z^{-b^i} \otimes I) |\phi\rangle_{y_i C'}. \quad (28)$$

Notice that the operation along the path commutes with  $V$ . Therefore, applying  $V$ , applying that operation, and then applying  $V^\dagger$  is equivalent to just applying that operation.

Notice that the resulting state continues to be locally maximally mixed. Therefore, we can apply the above repeatedly until all of the path circuits have been applied. Then the final state is as desired.  $\square$

*5.1.2. Geometric locality.* We explain how to modify the input and outputs sets  $\mathcal{I}$  and  $\mathcal{O}$  specified in (23) and (24) respectively in a way that both input and output are Boolean strings of the same length that can be organized in a 2-dimensional pattern and such that the circuit described in Sect. 5.1 can be implemented in the same depth  $(D + 1)$  using only geometrically local gates.

Recall that the input to the circuit consists of an input pattern  $\mathcal{P} = \{(\Gamma^{(i)}, u^{(i)})\}_{1 \leq i \leq r}$  together with an  $r$ -tuple of queries  $\{(x_1^{(i)}, \dots, x_\ell^{(i)})\}_{1 \leq i \leq r}$  in  $\mathcal{G}$ . Recall also that we think of the circuit as being organized on an  $N \times N$  grid of vertices, such that each vertex contains 4 groups of  $k$  qudits, each group facing one of the vertex' nearest neighbors on the grid. We index the input and output sets by grid vertices, with each vertex associated with an element taken from a constant-size alphabet  $\Sigma$  that is defined in (29) below.

Each star  $\Gamma^{(i)}$  specifies  $\ell$  paths  $\Gamma_j^{(i)}$  from the central box to the noncentral boxes. Assign to one point in each noncentral box the label  $u_j^{(i)}$ . Also assign to  $\ell$  points inside the central box the labels  $g_j^{(i)}$ . Extend the paths  $\Gamma_j^{(i)}$  so that their endpoints are  $u_j^{(i)}$  and  $g_j^{(i)}$ . We naturally distribute each question  $x_j^{(i)}$  at the grid vertex indicated by  $u_j^{(i)}$ .

For each edge  $(v, w)$  in the path, at vertex  $v$  (resp.  $w$ ) we include a symbol that indicates that a teleportation measurement is to be performed between the group of  $k$  qudits nearest to vertex  $w$  (resp.  $v$ ). For any grid vertex  $v$ , the output of the circuit at vertex  $v$  is either an answer in  $\mathcal{G}$ ,  $a_j^{(i)} \in (\mathbb{Z}_d)^{m_j}$ , or a teleportation measurement outcome, which is an element of  $(\mathbb{Z}_d)^{2k}$ . A question  $x_j^{(i)}$  is an element of  $(\mathbb{Z}_d^k)^{m_j}$ . This leads us to a circuit specification that considers the input and output sets

$$\mathcal{I} = \mathcal{O} = \Sigma^{\text{GRID}_N}, \quad \text{where } \Sigma = \{0, 1\}^{2mk \lceil \log d \rceil}, \quad (29)$$

where  $m = \max_j m_j$  and we fixed an arbitrary embedding of the natural input and output alphabets in  $\Sigma$ . Note that not all input strings are used; since we consider the parameters  $d, m, k$  to be constants (depending only on the type of stabilizer game chosen), the cardinality of the alphabet  $\Sigma$  is constant.

*5.2. Circuit game definition.* Having specified the ideal (or, ‘‘honest’’) quantum circuit that we have in mind, we are ready to give a formal definition of the circuit game associated with  $r$  copies of a stabilizer game  $\mathcal{G}$ . Recall the definition of the distribution  $\mathcal{D}^{(r)}$  on input patterns given in Definition 4.10. (For clarity, we omit the arguments  $N, L$ , for which we will eventually make an appropriate choice.)

**Definition 5.7.** Let  $\mathcal{G}$  be an  $(\ell, k, m)$  stabilizer game and  $1 \leq r \leq N$  integer. The circuit game  $G_{\mathcal{G}, N, r}$  is a game on the input and output sets defined in (29). The input distribution  $\pi$  is obtained by independently sampling an input pattern  $\mathcal{P}$  according to  $\mathcal{D}^{(r)}$  and a tuple of  $r$  independent queries  $(x^{(1)}, \dots, x^{(r)})$  for  $\mathcal{G}$ , and encoding them as an element of  $\mathcal{I}$  as described in Sect. 5.1.2. The relation  $\mathcal{R} \subseteq \mathcal{I} \times \mathcal{O}$  is defined as the support of the output distribution of the circuit described in Sect. 5.1, when it is provided an input in the support of  $\pi$ .

Skipping ahead, we note that in Sect. 6.2 we consider a slight variation of the circuit game  $G_{\mathcal{G}, N, r}$  from Definition 5.7, where the  $r$  query tuples to  $\mathcal{G}$  are no longer independent and the win condition is relaxed to allow failure in some of the game instances. These modifications allow us to obtain a circuit game whose inputs can be sampled using few random bits (polylogarithmic in  $N$ ), and that can be won with high probability by a circuit whose gates are subject to a limited amount of noise.

**5.3. Soundness.** We describe a reduction from circuit strategies (i.e. circuits with constant fan-in and bounded depth) in the circuit game introduced in Definition 5.7 to strategies for the players in the game  $\mathcal{G}$ . The next lemma refers to the notions of lightcone, input pattern, and circuit specification introduced in Sects. 4.1 and 4.2, and of rotated, stretched and repeated game introduced in Sects. 3.2 and 3.3.

**Lemma 5.8.** (Circuit locality implies local simulation) *Let  $G = G_{\mathcal{G}, N, r}$  be a circuit game as in Definition 5.7. Let  $\mathcal{P} = \{(u^{(i)}, \Gamma^{(i)})\}_{1 \leq i \leq r}$  be an input pattern in the support of the input distribution for  $G$ . Let  $\mathcal{C}$  be a circuit with fan-in  $K$  and depth  $D$  that wins with probability  $p$  in  $G_{\mathcal{G}, N, r}$ , for some  $0 \leq p \leq 1$ , conditioned on the input pattern being  $\mathcal{P}$ .*

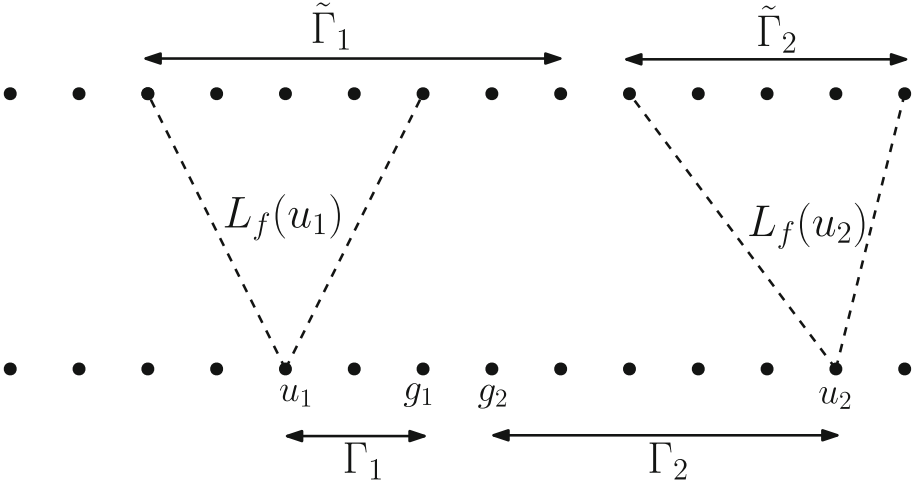
*Let  $\eta > 0$ . Let  $L_f$  be the lightcone function obtained from the circuit graph, and  $\mathcal{S} = (L_f, \text{BAD}_{in}, \emptyset)$  an associated circuit specification. Assume that  $\mathcal{P}$  is  $\mathcal{S}$ -causal and that a fraction at least  $1 - \delta$  of all input pairs  $(u, \Gamma) \in \mathcal{P}$  are  $\mathcal{S}$ -valid, for some  $\delta \in [0, 1]$ .*

*Then there exists an  $(r\ell)$ -player strategy in the  $r$ -repeated rotated stretched game  $\mathcal{G}' = ((\mathcal{G}_r)_{\Gamma}^{\mathcal{S}})^R$ , for some  $\Gamma$  depending on  $\mathcal{C}$  that is defined in the proof, such that with probability at least  $p$  the strategy succeeds in a fraction at least  $1 - \delta$  of the game instances.*

*Proof.* An input to the circuit  $\mathcal{C}$  consist of two parts: the pattern  $\mathcal{P} = \{(u^{(i)}, \Gamma^{(i)})\}_{1 \leq i \leq r}$ , and the queries  $\{(x_1^{(i)}, \dots, x_\ell^{(i)})\}_{1 \leq i \leq r}$ , that are embedded in the input to the circuit as described in Sect. 5.1.2. By assumption there is a fraction at most  $\delta$  of  $(u^{(i)}, \Gamma^{(i)})$  that are  $\mathcal{S}$ -valid. For the remainder of the argument, ignore those vertices (equivalently, relabel  $r$  to  $(1 - \delta)r$ ). When designing a strategy for the players in the game, the players associated to ignored vertices ignore their question and return a random answer.

The assumption that  $\mathcal{P}$  is  $\mathcal{S}$ -causal implies that for each  $i \in \{1, \dots, r\}$  each of the vertices  $v \in \Gamma^{(i)}$  has a backwards lightcone that includes at most one of the input locations  $u_j^{(i)}$ . Moreover, all vertices in  $\Gamma^{(i)} \cap \text{Box}(u_j^{(i)})$  have a backwards lightcone that includes no other input location than  $u_j^{(i)}$ .

We define an  $(r\ell)$ -player strategy in the rotated, stretched game  $\mathcal{G}'$ . For each  $i \in \{1, \dots, r\}$ , partition  $\Gamma^{(i)}$  into sets  $\tilde{\Gamma}_j^{(i)}$ ,  $j \in \{1, \dots, \ell\}$ , such that the only input location in the backwards lightcone of any vertex in  $\tilde{\Gamma}_j^{(i)}$  is  $u_j^{(i)}$ . Note that the lightcones may intersect at other grid vertices.



**Fig. 2.** Illustration of the construction in the proof of Lemma 5.8. A circuit with two input locations  $u_1, u_2$  (case  $r = 1$  and  $\ell = 2$ ). The star  $\Gamma$  has two simple paths linking the center vertices  $g_i$  with  $u_i$ , for  $i = 1, 2$ . The forward lightcones of  $u_1$  and  $u_2$  do not intersect at any vertex in  $\Gamma$ . The output vertices along  $\Gamma$  are partitioned into  $\tilde{\Gamma}_1$  and  $\tilde{\Gamma}_2$  in a way that all vertices within a lightcone of an input location  $u_i$  are associated with the index  $i$

Recall that by definition each wire of  $\mathcal{C}$  is associated with the space  $\mathbb{C}^{d'}$ , where  $d' = |\Sigma|$  with  $\Sigma$  the input alphabet for the circuit game. We now describe an unambiguous way to generate a density matrix  $\sigma$  on  $(\mathbb{C}^{d'})^{\otimes r\ell}$ , together with an assignment of each qudit of  $\sigma$  to a player  $(i, j)$ , using the circuit  $\mathcal{C}$  and the fixed input pattern  $\mathcal{P}$ .

We define  $\sigma$  as the output of the circuit  $\mathcal{C}$ , when certain inputs have been fixed, and certain wires have been traced out. For all input grid vertices that are not an input location  $u_j^{(i)}$ , hard-wire the input to  $|0\rangle$ . Execute the circuit until a vertex  $v$  of the circuit graph that is in the forward lightcone of an input vertex associated with location  $u_j^{(i)}$  has to be considered. Since no input has been hard-wired for that vertex, the circuit cannot proceed. There are two cases:

- If the forward lightcone of  $v$  intersects the forward lightcone of two different input locations  $u_j^{(i)}$ , then no vertex in the forward lightcone of that location can be on any of the stars  $\Gamma^{(i')}$  for  $i' \neq i$  (as otherwise the vertex would be a vertex of  $\Gamma$  whose backwards lightcone contains two distinct input locations). In that case, trace out the vertex.
- In all other cases, vertex  $v$  is in the forward lightcone of a single input location  $u_j^{(i)}$ . In this case, give the circuit wire associated with that vertex (in the state that it currently is) to player  $(i, j)$ .

Finally, split the unassigned vertices on any path  $\Gamma_j^{(i)}$  in an arbitrary way among the players; the set  $\tilde{\Gamma}_j^{(i)}$  is defined as the set of vertices from the star  $\Gamma^{(i)}$  assigned to the  $j$ th player. All remaining unassigned vertices are traced out.

This procedure specifies the state  $\sigma$  shared by the players (see Fig. 2 for an illustration). It remains to define their observables. Once the game starts, each player uses its input  $x_j^{(i)}$  in location  $u_j^{(i)}$  (encoded as an input to the circuit game, as specified in

Sect. 5.1.2), and proceeds to complete the execution of the circuit on the qudits that it holds. If a gate has an output wire that points to a qudit that is not in the player's possession, the player measures the qudit and ignores the outcome. Finally, the player measures all qudits in the locations  $\Gamma_j^{(i)}$  in the computational basis, and returns them as its answer (decoded as an answer in  $\mathcal{G}$ , as specified in Sect. 5.1.2).

The fact that this strategy for the players has the same success probability in  $\mathcal{G}'$  as the circuit  $\mathcal{C}$  in the circuit game  $G_{\mathcal{G},N,r}$  follows from the fact that the success criterion in  $G_{\mathcal{G},N,r}$  only involves output vertices that are along the stars  $\Gamma^{(i)}$ , and it can be verified that the joint operations performed by the players in the above-defined strategy correctly compute the reduced density matrix computed by the circuit on all those output vertices. Finally, using Lemma 5.6 it can be verified that the success criterion in the circuit game matches the win condition for the rotated stretched game.  $\square$

*Remark 5.9.* The proof of Lemma 5.8 establishes a stronger statement than claimed in the lemma, that will be useful later. Specifically, the reduction from a circuit to a strategy for the players in  $\mathcal{G}'$  constructed in the proof applies whenever the input pattern  $\mathcal{P}$  chosen in the circuit game is  $\mathcal{S}$ -causal for  $\mathcal{S}$  the circuit specification derived from  $\mathcal{C}$ . Moreover, whenever this is the case the reduction yields a strategy for the players that *exactly* reproduces the (suitably decoded) output distribution of the circuit, on *any* choice of queries  $x^{(i)}$ .

We end this section with the proof of Theorem 1.2, that specifies a circuit game for which there is a very large separation between the optimal winning probabilities of classical low-depth and quantum circuits.

*Theorem 1.2.* (Exponential Soundness, restated). There exists universal constants  $c, c' > 0$ , a family of distributions  $\{\mathcal{D}_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $\mathcal{D}_N$  is a distribution on  $\{0, 1\}^{N^2}$ , and a family of efficiently verifiable relations  $\{\mathcal{R}_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $\mathcal{R}_N \subseteq \{0, 1\}^{N^2} \times \{0, 1\}^{N^2}$ , such that the following holds:

- (Completeness) There exists a family of depth-3 geometrically local (in 2D) quantum circuits  $\{\mathcal{C}_N\}_{N \in \mathbb{N}}$  such that for any  $N \geq 1$  and any input  $x$  in the support of  $\mathcal{D}_N$  it holds that  $(x, \mathcal{C}_N(x)) \in \mathcal{R}_N$  with probability 1.
- (Soundness) For any family of classical circuits  $\{\mathcal{C}_N\}_{N \in \mathbb{N}}$  such that for every  $N \geq 1$ ,  $\mathcal{D}_N$  has depth at most  $c \log N$ , the probability that  $(x, \mathcal{C}_N(x)) \in \mathcal{R}$  for  $x \sim \mathcal{D}_N$  is  $O(\exp(-N^{c'}))$ .

*Proof.* Let  $G_N = G_{\mathcal{G},N,r}$  be the circuit game introduced in Definition 5.7, where the stabilizer game  $\mathcal{G}$  is instantiated as the GHZ game from Definition 3.14, and let  $\pi$  and  $\mathcal{P}$  be the input distribution and input pattern introduced in Definition 5.7 respectively.

*Completeness:* By Lemma 5.2 the circuit  $\mathcal{C}_{ideal}$ , as defined from the game  $G_N$  at the beginning of Sect. 5.1, wins the game  $G_N$  with probability at least  $(\omega^*(\mathcal{G}))^r$ , where  $\omega^*(\mathcal{G})$  is the optimal entangled winning probability for the stabilizer game  $\mathcal{G}$ . Since  $\omega^*(\mathcal{G}) = 1$  it follows that  $\mathcal{C}_{ideal}$  succeeds at  $G_N$  with probability 1. As noted after Definition 3.14, the shared state  $|\psi\rangle_{GHZ}$  in the optimal strategy for the GHZ game can be prepared starting from  $|000\rangle$  by a circuit of depth 3. It follows by definition that  $\mathcal{C}_{ideal}$  has depth at most 4. This establishes the completeness part of the corollary.

*Soundness:* Fix an  $\eta \in (0, \frac{1}{7})$ . Let  $\mathcal{C}$  be a classical circuit with fan-in  $K$  and depth  $D \leq c \log N$  (for  $c$  a sufficiently small constant depending on  $\eta$ ) that wins with probability  $p_{win}$  in  $G_{\mathcal{G},N,r}$ . We show that if  $p_{win}$  is too large, then there is a good classical strategy for the players in the game  $\text{GHZ}^{\lfloor r/2 \rfloor}$ , the  $\lfloor r/2 \rfloor$ -repeated GHZ game defined in Definition 3.10.

For an input pattern  $\mathcal{P}$  in the support of  $\mathcal{D}^{(r)}$ , say that  $\mathcal{P}$  has a large  $\mathcal{S}$ -causal subpattern if there exists a subpattern  $\mathcal{P}' \subseteq \mathcal{P}$  such that  $\mathcal{P}'$  is  $\mathcal{S}$ -causal and  $|\mathcal{P}'| \geq \lfloor r/2 \rfloor$ . Define  $q$  as

$$q = 1 - \Pr(\mathcal{P} \text{ has a large } \mathcal{S}\text{-causal subpattern}). \quad (30)$$

We first show an upper bound on  $q$ . For any pattern  $\mathcal{P}$  let  $\mathcal{P}_{\text{core}} = \mathcal{P}_{\text{VAL}} \cap \mathcal{P}_{\text{CAUS}}$ , where as in Lemma 4.15,

$$\mathcal{P}_{\text{VAL}} = \{(u, \Gamma) \in \mathcal{P} \mid (u, \Gamma) \text{ is } \mathcal{S}\text{-valid}\},$$

$$\mathcal{P}_{\text{CAUS}} = \{(u, \Gamma) \in \mathcal{P} \mid (u, \Gamma) \text{ is individually-}\mathcal{S}\text{-causal with respect to } \mathcal{P}_{\text{VAL}}\}.$$

By construction  $\mathcal{P}_{\text{core}}$  is  $\mathcal{S}$ -causal. To prove an upper bound on  $q$  it suffices to place a lower bound on the probability that  $|\mathcal{P}_{\text{core}}| \geq \lfloor r/2 \rfloor$ . Recall that the classical circuit  $\mathcal{C}$  has fan-in  $K$ , depth  $D \leq c \log N$ , and circuit specification  $\mathcal{S} = (L_f, \text{BAD}_{in}, \emptyset)$ . By Lemma 4.3, we may choose  $c$  as a function of  $\eta$  such that  $\mathcal{S}$  is  $(B, R_{in}, 0)$ -bounded, where  $B = O(N^\eta)$  and  $R_{in} = O(N^{2-\eta})$ . By (18) we get

$$\Pr\left(|\mathcal{P}_{\text{core}}| \geq r(1 - CrR_{in}/N^2 - 2p - 3t/r)\right) \geq 1 - 4 \exp\left(-t^2/8r\right), \quad (31)$$

where  $p = C'rB(rL^2/N^2 + \ell/L)$ . Here  $\ell = 3$ , and the other parameters depend on  $N$ . To set parameters, first recall that  $B = O(N^\eta)$  and  $R_{in} = O(N^{2-\eta})$ . Set  $t = r/10$ ,  $r = \Theta(N^\eta)$ , and  $L$  such that  $L = O(N^{1-3\eta})$  and  $L = \Omega(N^{4\eta})$ , which is possible as long as  $\eta < 1/7$ . Then  $p = O(1)$  and  $rR_{in}/N^2 = O(1)$ . By choosing the constants appropriately, we can ensure that

$$1 - CrR_{in}/N^2 - 2p - 3t/r > 1/2.$$

With this choice of parameters, (31) implies that  $|\mathcal{P}_{\text{core}}| \geq r/2$  with probability at least  $1 - 4 \exp(-C''r)$ , for some constant  $C'' > 0$ . To conclude, note that whenever  $\mathcal{P}$  contains a  $\mathcal{S}$ -causal subpattern  $\mathcal{P}' \subseteq \mathcal{P}$  such that  $|\mathcal{P}'| \geq r/2$  it follows from Lemma 5.8 and Remark 5.9 that the circuit  $\mathcal{C}$  implies a strategy for  $(r\ell/2)$  classical players in the repeated game  $\text{GHZ}^{\lfloor r/2 \rfloor}$ . Using that the maximum success probability of classical players in GHZ is  $3/4$  and that the classical value multiplies under repetition (since the players are distinct) it follows that the implied strategy has success probability at most  $(3/4)^{\lfloor r/2 \rfloor}$ .  $\square$

## 6. Randomness Generation

In this section we give the construction of a circuit game such that any low-depth circuit that succeeds in the game with non-negligible probability must generate outputs that have large min-entropy, even conditioned on the inputs and side information that may be correlated with the initial state of the circuit.

The main idea for the construction is to embed a large number of copies of a simple stabilizer game  $\mathcal{G}$  in a circuit game, as described in Sect. 5. (We use the Mermin 3-player GHZ game [Mer90a], though a similar reduction could be performed starting from any stabilizer game whose quantum value is larger than its classical value.) Using the reduction from Lemma 5.8, it follows from Remark 5.9 that the output distribution of any circuit that wins with non-negligible probability in the circuit game can be deterministically mapped to a (stretched, rotated) variant of the  $r$ -fold parallel repetition, with  $r$  independent groups of players, of  $\mathcal{G}$ . This reduction implies that, to bound the



output entropy of the circuit, it suffices to place a lower bound on the output entropy of any strategy in the parallel repeated game.

To accomplish this last step we employ the framework based on the Entropy Accumulation Theorem (EAT) [DFR16] introduced in [AFDF+18], including the improvements from [DF18]. This framework allows to place a linear (in the number of repetitions) lower bound on the amount of min-entropy generated in the sequential repetition of a nonlocal game, using a lower bound on the function that measures the von Neumann entropy generated in a single instantiation of the game as a function of the success probability. Our setting of parallel repetition is more constrained (thus in principle easier to analyze) than the sequential setting, but the results from [AFDF+18,DF18] still give the best rates for both settings.

In Sect. 6.1 we start by establishing a bound on the single-round randomness for the three-player GHZ game that takes the form required to apply the framework from [AFDF+18]. In Sect. 6.2 we combine this bound with the reduction from circuit games to nonlocal games shown in Sect. 5 to deduce a family of randomness-generating circuit games.

*6.1. Randomness generation from the GHZ game.* We briefly recall the formalism from [AFDF+18], when tailored to our setting (in particular, we focus on processes specified by quantum strategies in a nonlocal game, instead of arbitrary quantum channels in [AFDF+18]). The main definition that is needed is that of a *min-tradeoff function*, which specifies a lower bound on the amount of randomness generated in a single execution of a nonlocal game, as a function of the players' probability of winning in the game. We give the definition for stabilizer games, as introduced in Definition 3.1.

**Definition 6.1.** (*Min-tradeoff function*) Let  $\mathcal{G}$  be an  $(\ell, k, m)$  stabilizer game. Fix a set of measurements  $\{M_{x_j}\}$  for the  $\ell$  players in the game. For any  $\omega \in [0, 1]$ , let  $\Sigma(\omega)$  denote the set of states  $\rho_{\mathbb{P}_1 \dots \mathbb{P}_k \mathbb{R}}$  such that when the players' state is initialized to  $\rho$  (with player  $i$  being given register  $\mathbb{P}_i$ ), the players' strategy wins the game with probability at least  $\omega$ .<sup>6</sup>

Then a real affine function  $f$  on  $[0, 1]$  is called an (affine) *min-tradeoff function* for  $\mathcal{G}$  and  $\{M_{x_j}\}$  if it satisfies

$$f(\omega) \leq \min_{\rho \in \Sigma(\omega)} H(A|QR)_{\mathcal{M}(\rho)},$$

where the entropy is evaluated on the post-measurement state  $\mathcal{M}(\rho)$  obtained after application of the players' measurements, and  $Q$  and  $A$  are random variables that represent inputs (distributed according to  $\pi$ ) and outputs for the players in  $\mathcal{G}$ . If  $f$  is a min-tradeoff function for a game  $\mathcal{G}$  and every possible set of measurements  $\{M_{x_j}\}$  for the players, then we simply say that  $f$  is a min-tradeoff function for  $\mathcal{G}$ .

To illustrate the definition we apply results from [WBA18] to give a min-tradeoff function for the Mermin GHZ game introduced in Definition 3.14. It is well-known (and easily verified) that the best classical strategy for this game succeeds with probability  $\frac{3}{4}$ . This in particular implies that any strategy that wins with probability strictly larger than  $\frac{3}{4}$  cannot be deterministic, hence generates randomness. The following bound shown

<sup>6</sup> We may without loss of generality assume that the dimension of  $\mathbb{R}$  is no more than the sum of the dimensions of the players' private registers  $\mathbb{P}_j$ , themselves fixed by the measurements  $\{M_{x_j}\}$ . Therefore, the set  $\Sigma(\omega)$  can be taken to be a compact set.

in [WBA18] provides a tight lower bound on the conditional entropy present in the outputs of any strategy that succeeds with sufficiently large probability.<sup>7</sup>

**Lemma 6.2.** [WBA18] *Let  $\tau = (\rho, \{M_{x_j}\})$  be a strategy with success probability  $\omega \geq \frac{7}{8}$  in the game GHZ, where  $\rho$  is a density matrix on the provers' registers  $P_1 \cdots P_\ell$  and an arbitrary auxiliary register  $R$ . Then*

$$H(A_1 A_2 | R Q) \geq f_{\text{GHZ}}(\omega) = -\log\left(\frac{5}{4} - \omega + \sqrt{3}\sqrt{\left(\omega - \frac{1}{2}\right)(1 - \omega)}\right), \quad (32)$$

where  $Q$  is a random variable that denotes the query to the players, and  $A_1, A_2$  are random variables that denote the answers  $a_1, a_2 \in \mathbb{Z}_2$  from the first two players.

Note that for  $\omega = 1$ , the bound (32) gives 2 bits of entropy, which is clearly optimal. If  $\omega = 1 - \varepsilon$  for small  $\varepsilon$ , the bound degrades as  $2 - O(\sqrt{\varepsilon})$ .

Following the framework from [AFDF+18], the bound provided in Lemma 6.2 already implies a linear lower bound on the entropy generated by the sequential repetition of the GHZ game. For our purposes it will be convenient to have a bound on the entropy generated by the stretched, rotated variant of the GHZ game, as introduced in Sect. 3.2.

**Corollary 6.3.** (Min-tradeoff function for GHZ) *Let  $\frac{7}{8} < p_s < 1$ . Let  $\Gamma$  be a tuple of sets as in Definition 3.9. Then the function  $g_{p_s} : [0, 1] \rightarrow \mathbb{R}$  defined by*

$$g_{p_s}(q) = f_{\text{GHZ}}(p_s) + (q - p_s) \frac{df_{\text{GHZ}}}{d\omega}(p_s)$$

is a min-tradeoff function for the rotated stretched game  $\text{GHZ}_\Gamma^{S,R}$ .

*Proof.* First we observe that the bound (32) from Lemma 6.2 applies equally to the the rotated stretched game  $\text{GHZ}_\Gamma^{S,R}$ , for any fixed  $\Gamma$ . Indeed, fix a strategy  $(\rho, \{M_{x_j}\})$  in  $\text{GHZ}_\Gamma^{S,R}$ . Using Lemma 3.8 we obtain a strategy  $\tau' = (\rho, \{M'_{x_j}\})$  for GHZ that has the same success probability. Furthermore, in the coarse-graining of the strategy the answers  $A_1, A_2, A_3 \in \mathbb{Z}_2$  in  $\tau'$  are a deterministic function of the answers  $A_1, A_2, A_3 \in \mathbb{Z}_2 \times (\mathbb{Z}_2^{|\Gamma|})$  in  $\text{GHZ}_\Gamma^{S,R}$ , where the second factor is for the stretched rotation string. Therefore the same bound on randomness generation that applies to GHZ applies to  $\text{GHZ}_\Gamma^{S,R}$  (as long as all outputs in the game are included, which is the case for the definition of a min-tradeoff function).

To conclude, note that the right-hand side of (32) is a convex function of  $\omega$ , hence it is at least its tangent at any point.  $\square$

Recall that our goal is to generate a large amount of randomness by requiring a circuit to play multiple copies of the game GHZ in parallel. Towards this, we introduce a partially derandomized variant of the repeated game, where the inputs are chosen according to a very biased distribution in order to save on the randomness required to generate them. The resulting game is a direct analogue of the protocol for randomness expansion from the CHSH game given in [DF18].

**Definition 6.4.** Let  $r \geq 1$  be an integer. Let  $p, \gamma \in [0, 1]$ . Let  $\Gamma$  be a tuple of sets as in Definition 3.9. Let  $\text{GHZ}_{r,\Gamma,p,\gamma}^{S,R}$  denote the  $r$ -fold repetition, as in Definition 3.10, of the rotated, stretched game GHZ with the following two modifications:

<sup>7</sup> See (18) in [WBA18]. The authors prove a stronger bound, that applies to the min-entropy and extends to all success probabilities in the range  $[3/4, 1]$ . We only state the weaker bound that will be sufficient for us. To see how the bound stated in Lemma 6.2 is obtained from (18) in [WBA18], use the replacement  $\omega = \frac{1}{2} + \frac{M}{8}$ .

- The  $r$  queries, instead of being sampled independently, are chosen according to the following distribution: first, select a subset  $S \subseteq \{1, \dots, r\}$  by including each element independently with probability  $p$ ; second, select queries  $x^{(1)}, \dots, x^{(r)} \in X$  such that  $x^{(i)}$  is sampled as in GHZ when  $i \in S$ , and  $x^{(i)} \leftarrow \bar{x}$  for  $i \notin S$ , where  $\bar{x}$  is an arbitrary, but fixed, query in GHZ;
- It is only required that the win condition is satisfied for a fraction at least  $(1 - \gamma)$  of the tuples of answers  $a^{(i)}$  such that  $i \in S$  (there is no requirement for  $i \notin S$ ).

Corollary 6.3 shows that any sufficiently successful strategy in the (rotated, stretched) game GHZ must generate outputs that contain a constant amount of entropy. It is therefore natural to expect that a strategy for the repeated game from Definition 6.4 should generate a linear (in the number of repetitions) amount of entropy. The difficulty is to obtain a bound on the entropy generated, conditioned on having produced outputs that satisfy the win condition of the game, but without placing an implicit assumption on the intrinsic winning probability of the strategy (which would be difficult to estimate). Moreover, the fact that the strategy involves all players simultaneously measuring parts of the same entangled state introduces correlations that prevent a direct treatment using techniques appropriate for the simpler case of i.i.d. (identically and independently distributed) outputs.

The Entropy Accumulation Theorem, as applied in [AFDF+18], is designed specifically to address these difficulties, and indeed guarantees that the repeated game introduced in Definition 6.4 generates a linear amount of min-entropy. Here we use the improved results from [DF18],<sup>8</sup> that give good bounds even when the “test probability”  $p$  from Definition 6.4 can be very small.

**Lemma 6.5.** *Let  $r \geq 1$ ,  $p, \gamma \in [0, 1]$ , and  $\varepsilon > 0$ . Let  $(\rho, \{M_{x_j}\})$ , where  $\rho$  is a density matrix on the player’s private registers together with an ancilla register  $\mathbf{E}$ , be a strategy for the  $(3r)$  players in  $\text{GHZ}_{r,\Gamma,p,\gamma}^{S,R}$  that succeeds with probability at least  $\varepsilon$ . Let  $Q = (Q^{(1)}, \dots, Q^{(r)})$  (resp.  $A = (A^{(1)}, \dots, A^{(r)})$ ) be random variables associated with the players’ answers in each copy of GHZ; note that each  $Q^{(i)}$  (resp.  $A^{(i)}$ ) is itself a 3-tuple. Let  $\rho_{\text{AQE}}^S$  denote the state of AQE conditioned on the players succeeding in the game (the players’ private registers are traced out). Then*

$$H_{\min}^{\varepsilon}(A|QE)_{\rho^S} \geq f_{\text{GHZ}}(1 - \gamma)r - O\left(\frac{r}{\sqrt{p}} \left(\ln \frac{1}{\varepsilon \text{Tr}(\rho^S)}\right)^{1/2}\right). \quad (33)$$

Note that the lower bound provided in (33) is non-trivial as soon as  $p = \Omega(\log N/r)$  and  $\varepsilon$  is at least inverse-polynomially large (smaller  $\varepsilon$  is also possible, but requires a larger  $p$ ).

*Proof.* The proof is identical to the proof of [DF18, Theorem 6.1], that applies to the CHSH game. The only change needed is to use the min-tradeoff function from Corollary 6.3 instead of the min-tradeoff function  $g^*$  for the CHSH game used in [DF18]. The bound (33) follows from the bound stated in [DF18, Theorem 6.1] by noting that  $\frac{df}{d\omega}(1 - \gamma)$  is bounded by a universal constant.  $\square$

<sup>8</sup> The results in [AFDF+18, DF18] apply to a much more general scenario, and in particular allow one to prove bounds on the entropy generated by an arbitrary sequential process, as long as it satisfies a certain Markov condition. Our setting, which considers parallel repetition, is easier, and trivially satisfies the required Markov condition.

**6.2. Randomness generation from low-depth circuits.** Using the technique to embed a stabilizer game into a circuit game described in Sect. 5 we can leverage the randomness generation results from the previous section to obtain a family of circuit games for certified randomness expansion. First we define the circuit games that we consider by introducing a randomness-efficient, noise-tolerant modification of the game from Definition 5.7. Even though we will eventually instantiate the definition with  $\mathcal{G} = \text{GHZ}$ , we give the definition for a general stabilizer game.

**Definition 6.6.** Let  $\mathcal{G}$  be an  $(\ell, k, m)$  stabilizer game and  $N \geq 1$  an integer. Let  $L, B, R_{in}, r$  be parameters that satisfy the constraints (22) for some constant  $\eta > 0$ . Let  $p, \gamma \in [0, 1]$ . The circuit game  $\tilde{\mathcal{G}}_{\mathcal{G}, N, r, p, \gamma}$  is defined as the game  $G_{\mathcal{G}, N, r}$  with the following modifications:

1. The input pattern  $\mathcal{P}$  is sampled according to the distribution  $\tilde{D}^{(r)}$  from Lemma 4.20;
2. The tuple of queries  $(x^{(1)}, \dots, x^{(r)})$  for  $\mathcal{G}$  is sampled by first, selecting a subset  $S \subseteq \{1, \dots, r\}$  by including each element independently with probability  $p$ ; second, selecting queries  $x^{(1)}, \dots, x^{(r)} \in X$  such that  $x^{(i)}$  is sampled as in  $\mathcal{G}$  when  $i \in S$ , and  $x^{(i)} \leftarrow \bar{x}$  for  $i \notin S$ , where  $\bar{x}$  is an arbitrary, but fixed, query in  $\mathcal{G}$ ;
3. It is only required that the win condition is satisfied for a fraction at least  $(1 - \gamma)$  of the tuples of answers  $a^{(i)}$  such that  $i \in S$  (there is no requirement for  $i \notin S$ ).

As shown in Lemma 4.20, it is possible to sample an input pattern from  $\tilde{D}^{(r)}$  using  $O(\log^2 N)$  random bits. In addition, it is possible to sample inputs as in Definition 6.6 using  $O(H(p)r)$  random bits. So, if  $p = \Theta(\log N/r)$ , then it is possible to sample inputs in  $\tilde{\mathcal{G}}_{\mathcal{G}, N, r, p, \gamma}$  using  $O(\log^2 N)$  random bits. (To this count, one may add  $O(\log^3 N)$  random bits, sufficient to extract near-uniform random bits from the output of the circuit by using a seed-efficient randomness extractor [DPVR12].) The following theorem places a lower bound on the amount of randomness generated by a circuit that succeeds with non-negligible probability in the circuit game from Definition 6.6, when the stabilizer game  $\mathcal{G}$  is instantiated as the GHZ game from Definition 3.14. Together with the preceding comments, the theorem establishes the randomness certification part of Theorem 1.1 (the soundness part follows immediately since classical circuits cannot increase entropy present in their input distribution).

**Theorem 6.7.** Let  $r, p, \gamma, N$  be as in Definition 6.6, for some  $\eta > 0$ . Let  $\eta' > 0$ . Let  $D$  be such that  $D \leq c \log N$ , for some sufficiently small constant  $c > 0$ , depending on  $\eta, \eta'$ . Let  $\mathcal{C}$  be a (classical or quantum) circuit with gates of constant fan-in and depth at most  $D$ . Assume that  $\mathcal{C}$  succeeds in the game  $\tilde{\mathcal{G}}_{\text{GHZ}, N, r, p, \gamma}$  with probability at least  $\delta$ , for some  $\delta = \Omega(N^{-\eta'})$ . Suppose the circuit is executed on its input, described by random variable  $I$ , as well as an auxiliary state  $|\psi\rangle_{\text{CE}}$  such that the circuit acts on register  $\mathcal{C}$ , and the register  $\mathcal{E}$  is available to an adversary. Let  $O$  be a random variable that represents the circuit output,  $O = \mathcal{C}(I)$ . Let  $\rho_{\text{OIE}}^s$  denote the state of the inputs, outputs, and side information, conditioned on the circuit winning in the circuit game. Then there exists an  $\eta''' > 0$  such that for any  $\varepsilon = \Omega(N^{-\eta''})$ , it holds that

$$H_{\min}^{\varepsilon}(O|IE)_{\rho^s} \geq (\kappa - f(\gamma))r - O\left(\frac{r}{\sqrt{p}}\left(\ln \frac{1}{\varepsilon \text{Tr}(\rho^s)}\right)^{1/2}\right), \quad (34)$$

where the implicit constant depends on  $\eta, \eta', \eta'''$ .

*Proof.* Let  $\mathcal{S}$  be the circuit specification obtained from the circuit  $\mathcal{C}$ . It follows from Lemma 4.3 that by choosing the constant  $c$  small enough, we force  $\mathcal{S}$  to be  $(B, R_{in}, 0)$ -bounded where  $R_{in} = O(N^{2-\eta})$  and  $B = O(N^\eta)$ . With this choice of parameters,

it follows from Lemma 4.20 and Lemma 4.12 that a pattern  $\mathcal{P}$  sampled from  $\tilde{D}^{(r)}$  is  $\mathcal{S}$ -causal with probability at least  $1 - O(N^{-c'})$ , for some  $c' > 0$ . Since  $\mathcal{C}$  succeeds in  $\tilde{\mathcal{G}} = \tilde{\mathcal{G}}_{\text{GHZ}, N, r, p, \gamma}$  with probability  $\delta$ , it follows that conditioned on success of  $\mathcal{C}$ , the pattern  $\mathcal{P}$  chosen as part of the input is  $\mathcal{S}$ -causal with probability at least  $1 - O(N^{-c'}\delta^{-1})$ .

Let  $\delta' > 0$  and let  $\mathcal{P}$  be a pattern which is  $\mathcal{S}$ -causal on which  $\mathcal{C}$  succeeds with probability at least  $\delta'$  when the pattern  $\mathcal{P}$  is fixed. For any such pattern, Lemma 5.8 together with Remark 5.9 imply that the circuit's behavior can be simulated by a strategy for the players in the stretched rotated game  $\text{GHZ}_{r, \Gamma, p, \gamma}^{\mathcal{S}, R}$ , for some collection of sets  $\Gamma$  that is determined from  $\mathcal{S}$ . Using Lemma 6.5 it follows that, conditioned on the input  $I$  to the circuit being of the form  $(\mathcal{P}, x)$  for some query  $x \in X$ , the lower bound (33) holds.

If  $\mathcal{P}$  is such that  $\mathcal{C}$  succeeds with probability less than  $\delta'$ , then there is no bound on the min-entropy. However, the probability that this happens, conditioned on winning, is at most  $\delta'/\delta$ . (To see this, apply Bayes' rule directly.) Choosing  $\delta' = \sqrt{\delta}$  we get an  $\eta'''$ , depending on  $\eta'$ , such that (34) holds.  $\square$

*Acknowledgement.* The authors thank Adam Bouland for helpful discussions and members of the Caltech theory reading group (Matthew Weidner, Andrea Coladangelo, Jenish Mehta, Chinmay Nirkhe, Rohit Gurjar, Spencer Gordon) for posing some of the questions answered in this work. We thank Isaac Kim, Jean-Francois Le Gall, and Robin Kothari for useful discussions following the initial announcement of our results. Thomas Vidick is supported by NSF CAREER Grant CCF-1553477, AFOSR YIP award number FA9550-16-1-0495, MURI Grant FA9550-18-1-0161, a CIFAR Azrieli Global Scholar award, and the the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1733907). Jalex Stark is supported by NSF CAREER Grant CCF-1553477, ARO Grant W911NF-12-1-0541, and NSF Grant CCF-1410022. Matthew Coudron is supported by Canada's NSERC and the Canadian Institute for Advanced Research (CIFAR), and through funding provided to IQC by the Government of Canada and the Province of Ontario.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Aar18. Aaronson, S.: Certified randomness from quantum supremacy. Personal communication (2018)
- AC17. Aaronson, S., Chen, L.: Complexity-theoretic foundations of quantum supremacy experiments. In: Proceedings of the 32nd Computational Complexity Conference, p. 22. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2017)
- AFDF+18. Arnon-Friedman, R., Dupuis, F., Fawzi, O., Renner, R., Vidick, T.: Practical device-independent quantum cryptography via entropy accumulation. *Nat. Commun.* **9**(1), 459 (2018)
- BCM+18. Brakerski, Z., Christiano, P., Mahadev, U., Vazirani, U., Vidick, T.: Certifiable randomness from a single quantum device (2018). arXiv preprint [arXiv:1804.00640](https://arxiv.org/abs/1804.00640)
- BCP+14. Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V., Wehner, S.: Bell nonlocality. *Rev. Mod. Phys.* **86**(2), 419 (2014)
- Bel64. Bell, J.S.: On the Einstein–Podolsky–Rosen paradox. *Physics* **1**, 195–200 (1964)
- BGK17. Bravyi, S., Gosset, D., Koenig, R.: Quantum advantage with shallow circuits (2017). arXiv preprint [arXiv:1704.00690](https://arxiv.org/abs/1704.00690)
- BGK18. Bravyi, S., Gosset, D., Koenig, R.: Quantum advantage with shallow circuits. *Science* **362**(6412), 308–311 (2018)

- BWKST19. Bene Watts, A., Kothari, R., Schaeffer, L., Tal, A.: Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits. In: Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (2019)
- CHTW04. Cleve, R., Høyer, P., Toner, B., Watrous, J.: Consequences and limits of nonlocal strategies. In: Proceedings of the 19th IEEE Conference on Computational Complexity (CCC'04), pp. 236–249. IEEE Computer Society (2004)
- CZX+18. Chen, Z.-Y., Zhou, Q., Xue, C., Yang, X., Guo, G.-C., Guo, G.-P.: 64-qubit quantum circuit simulation. *Sci. Bull.* **63**(15), 964–971 (2018)
- DF18. Dupuis, F., Fawzi, O.: Entropy accumulation with improved second-order. Technical report (2018). [arXiv:1805.11652](https://arxiv.org/abs/1805.11652)
- DFR16. Dupuis, F., Fawzi, O., Renner, R.: Entropy accumulation (2016). arXiv preprint [arXiv:1607.01796](https://arxiv.org/abs/1607.01796)
- DHKL18. Dalzell, A.M., Harrow, A.W., Koh, D.E., LaPlaca, R.L.: How many qubits are needed for quantum computational supremacy? (2018). arXiv preprint [arXiv:1805.05224](https://arxiv.org/abs/1805.05224)
- DPVR12. De, A., Portmann, C., Vidick, T., Renner, R.: Trevisan's extractor in the presence of quantum side information. *SIAM J. Comput.* **41**(4), 915–940 (2012)
- Eke91. Ekert, A.K.: Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991)
- Gall18. Gall, F.L.: Average-case quantum advantage with shallow circuits (2018). arXiv preprint [arXiv:1810.12792](https://arxiv.org/abs/1810.12792)
- GVW+15. Giustina, M., Versteegh, M.A., Wengerowsky, S., Handsteiner, J., Hochrainer, A., Pelhal, K., Steinlechner, F., Kofler, J., Larsson, J.-Å., Abellán, C., et al.: Significant-loophole-free test of Bell's theorem with entangled photons. *Phys. Rev. Lett.* **115**(25), 250401 (2015)
- HBD+15. Hensen, B., Bernien, H., Dréau, A., Reiserer, A., Kalb, N., Blok, M., Ruitenberg, J., Vermeulen, R., Schouten, R., Abellán, C., et al.: Experimental loophole-free violation of a Bell inequality using entangled electron spins separated by 1.3 km (2015). arXiv preprint [arXiv:1508.05949](https://arxiv.org/abs/1508.05949)
- HM17. Harrow, A.W., Montanaro, A.: Quantum computational supremacy. *Nature* **549**(7671), 203 (2017)
- HNS18. Huang, C., Newman, M., Szegedy, M.: Explicit lower bounds on strong quantum simulation (2018). arXiv preprint [arXiv:1804.10368](https://arxiv.org/abs/1804.10368)
- INW94. Impagliazzo, R., Nisan, N., Wigderson, A.: Pseudorandomness for network algorithms. In: Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, pp. 356–364. ACM (1994)
- Mer90a. Mermin, N.D.: Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.* **65**(15), 1838 (1990)
- Mer90b. Mermin, N.D.: Simple unified form for the major no-hidden-variables theorems. *Phys. Rev. Lett.* **65**(27), 3373 (1990)
- MFIB18. Markov, I.L., Fatima, A., Isakov, S.V., Boixo, S.: Quantum supremacy is both closer and farther than it appears (2018). arXiv preprint [arXiv:1807.10749](https://arxiv.org/abs/1807.10749)
- NC02. Nielsen, M.A., Chuang, I.: *Quantum Computation and Quantum Information* (2002)
- RBH05. Raussendorf, R., Bravyi, S., Harrington, J.: Long-range quantum entanglement in noisy cluster states. *Phys. Rev. A* **71**(6), 062313 (2005)
- RUV13. Reichardt, B., Unger, F., Vazirani, U.: A classical leash for a quantum system: Command of quantum systems via rigidity of CHSH games. *Nature* **496**(7446), 456–460 (2013)
- SBT+18. Schäfer, V., Ballance, C., Thirumalai, K., Stephenson, L., Ballance, T., Steane, A., Lucas, D.: Fast quantum logic gates with trapped-ion qubits. *Nature* **555**(7694), 75 (2018)
- SMSC+15. Shalm, L.K., Meyer-Scott, E., Christensen, B.G., Bierhorst, P., Wayne, M.A., Stevens, M.J., Gerrits, T., Glancy, S., Hamel, D.R., Allman, M.S., et al.: Strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**(25), 250402 (2015)
- Tom15. Tomamichel, M.: *Quantum Information Processing with Finite Resources: Mathematical Foundations*, vol. 5. Springer, Berlin (2015)
- VV14. Vazirani, U., Vidick, T.: Fully device-independent quantum key distribution. *Phys. Rev. Lett.* **113**(14), 140501 (2014)
- WBA18. Woodhead, E., Bourdoncle, B., Acín, A.: Randomness versus nonlocality in the Mermin-Bell experiment with three parties (2018). arXiv preprint [arXiv:1804.09733](https://arxiv.org/abs/1804.09733)
- WBMS16. Wu, X., Bancal, J.-D., McKague, M., Scarani, V.: Device-independent parallel self-testing of two singlets. *Phys. Rev. A* **93**(6), 062121 (2016)
- WEH18. Wehner, S., Elkouss, D., Hanson, R.: Quantum internet: a vision for the road ahead. *Science* **362**(6412), eaam9288 (2018)