Fault-tolerant compiling of classically hard IQP circuits on hypercubes

Dominik Hangleiter,[1,*] Marcin Kalinowski,[2,*] Dolev Bluvstein,[2,*] Madelyn Cain,[2] Nishad Maskara,[2] Xun Gao,[2,3] Aleksander Kubica,[4,5] Mikhail D. Lukin,[2] and Michael J. Gullans[1]

[1] *Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland 20742, USA*
[2] *Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[3] *JILA and Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*
[4] *AWS Center for Quantum Computing, Pasadena, California 91125, USA*
[5] *California Institute of Technology, Pasadena, California 91125, USA*

(Dated: May 1, 2024)

Realizing computationally complex quantum circuits in the presence of noise and imperfections is a challenging task. While fault-tolerant quantum computing provides a route to reducing noise, it requires a large overhead for generic algorithms. Here, we develop and analyze a hardware-efficient, fault-tolerant approach to realizing complex sampling circuits. We co-design the circuits with the appropriate quantum error correcting codes for efficient implementation in a reconfigurable neutral atom array architecture, constituting what we call a *fault-tolerant compilation* of the sampling algorithm. Specifically, we consider a family of $[\![2^D, D, 2]\!]$ quantum error detecting codes whose transversal and permutation gate set can realize arbitrary degree-D instantaneous quantum polynomial (IQP) circuits. Using native operations of the code and the atom array hardware, we compile a fault-tolerant and fast-scrambling family of such IQP circuits in a hypercube geometry, realized recently in the experiments by Bluvstein *et al.* [Nature 626, 7997 (2024)]. We develop a theory of second-moment properties of degree-D IQP circuits for analyzing hardness and verification of random sampling by mapping to a statistical mechanics model. We provide strong evidence that sampling from these hypercube IQP circuits is classically hard to simulate even at relatively low depths. We analyze the linear cross-entropy benchmark (XEB) in comparison to the average fidelity and, depending on the local noise rate, find two different asymptotic regimes. To realize a fully scalable approach, we first show that Bell sampling from degree-4 IQP circuits is classically intractable and can be efficiently validated. We further devise new families of $[\![O(d^D), D, d]\!]$ color codes of increasing distance d, permitting exponential error suppression for transversal IQP sampling. Our results highlight fault-tolerant compiling as a powerful tool in co-designing algorithms with specific error-correcting codes and realistic hardware.

I. INTRODUCTION

Quantum computers hold a promise to significantly outperform classical computers at various tasks. However, for many envisioned applications, very low error rates below $\sim 10^{-10}$ are required [1–5], in stark contrast to the state of the art experimental physical error rates of $\sim 10^{-3}$. Quantum error correction (QEC) provides a potential solution to this challenge by encoding error-corrected "logical" qubits across many redundant physical qubits [6–8]. In principle, QEC can exponentially suppress the logical error rate by increasing the code distance d, thereby promising a realistic route to low error rates required for large-scale algorithms. However, implementing QEC in practice is a challenging task. In addition to the large physical qubit overheads, QEC codes typically realize a discrete gate set using native operations [9]. Although universal computation can be realized through various techniques such as magic state distillation [10–12] and code switching [13, 14], these are generally very resource intensive. Thus devising hardware-efficient and fault-tolerant implementations of quantum algorithms is a non-trivial task, requiring co-design of error correcting code and physical implementation with the algorithm. We call this task *fault-tolerant compiling*.

Realizing computationally hard sampling algorithms is an interesting goal for logical qubit processors for a number of reasons. First, using QEC the logical noise rate can be exponentially suppressed and high circuit fidelities can be maintained while system size is increased. In contrast, the non-corrected signal decays exponentially with increasing circuit depth and system size [15, 16]. Second, computationally complex sampling circuits can be implemented using significantly fewer resources compared to universal quantum computation. In particular, a quantum computation does not have to be universal in order to be classically hard to simulate [17, 18]. This opens up intriguing opportunities for co-designing an algorithmic implementation with a QEC code. Specifically, complex quantum sampling circuits can be based on a variety of non-universal gate sets [17–19], which allows us to restrict the circuits to native gate sets of QEC codes. Moreover, they profit from even a limited amount of error detection and correction to improve the sample quality. These features dramatically reduce the overhead in fault-tolerant compiling. Finally, such circuits which are also fast scrambling can be used to benchmark the performance of a quantum processor [15, 16, 20–22].

* DH, MK and DB contributed equally.

In this work, we propose a viable path to systematically improve experimental implementations of complex sampling circuits with encoded qubits on the near-term quantum processors. Our approach is based on D-dimensional color codes with parameters $[\![2^D, D, 2]\!]$ (distance-2 codes with D logical qubits encoded in 2^D physical qubits). For $D \geq 3$, these codes support transversally implemented D-qubit non-Clifford gates, as well as CNOT and SWAP gates realized by qubit permutations. We show that along with transversal CNOT gates these native operations allow us realize arbitrary degree-D instantaneous quantum polynomial (IQP) circuits [17, 23]; sampling from such circuits is believed to be a hard task for classical computers based on complexity-theoretic arguments [24, 25]. The distance-2 codes allow us to detect errors directly from the classical samples, yielding an improvement over bare circuits even without intermediate measurements. Nonetheless, an approach based on error detection is not scalable. To achieve scalability, we devise new families of $[\![O(d^D), D, d]\!]$ color codes based on the $[\![2^D, D, 2]\!]$ family, allowing for repeated rounds of error correction throughout the circuit execution. For these code families and transversal IQP sampling, there is a noise threshold [26] below which the noisy output distribution converges exponentially fast towards the ideal output distribution as the code size is increased.

In order to maximize hardware efficiency, we focus on the capabilities of the recently realized logical quantum processor [27] that is based on reconfigurable arrays of neutral atoms in optical tweezers [28–30]. In this setting, many physical quantum operations can be naturally parallelized, including single-qubit gates on blocks of physical qubits and transversal entangling gates between large blocks. We design a hardware-efficient family of degree-D IQP circuits with connectivity graph given by a $\mathfrak{D}$-dimensional hypercube which we call *hypercube IQP (hIQP)* circuits. We show that this family rapidly converges to uniform IQP circuits and can therefore be thought of as a fault-tolerant compilation of the uniform IQP family. In an hIQP circuit, transversal degree-D and permutation CNOT circuits are performed in each block and the blocks are coupled by transversal CNOT gates. We analyze the conditions under which we expect random degree-D hIQP circuits to be sufficiently scrambling for quantum advantage as well as benchmarking applications. Finally, we address the issue of *efficiently* verifying quantum advantage in this model, by showing that degree-D IQP sampling can be efficiently validated by measuring two copies of a logical degree-$(D+1)$ circuit in the Bell basis.

A. Summary of Results

Our approach to fault-tolerant implementations of computationally hard algorithms is based on fault-tolerant compiling of IQP circuit sampling using high-dimensional color codes. Our goal is to devise a scalable path towards the quantum advantage regime starting from experiments that are possible today, as demonstrated in Ref. [27]. Thus, we address the finite-size and asymptotic regimes of three distinct properties of our proposed algorithm—fault tolerance, complexity, and verification. Concretely, we start from co-designing the quantum code and logical circuits with the reconfigurable atom array hardware as in Refs. [27, 32], giving rise to the hIQP circuits implemented natively on hypercube color codes, see Fig. 1(a-e). In what follows we summarize our key results.

First, in Section II, we analyze efficiently simulatable instances of hIQP circuits in which error detection is performed at the end of the circuit (Fig. 1(b)) in terms of the achievable error reduction. To understand the behaviour of the logical errors, we introduce a notion of the average gate fidelity for logical quantum gates and study the performance of the transversal and permutation gates of the $[\![8, 3, 2]\!]$ code using classical simulation. We analyze the behaviour of the average logical fidelity as a function of the amount of error detection and explain its power-law behaviour using a simple noise model.

Second, in Section III, we study the properties of hIQP circuits that are relevant to quantum advantage demonstrations: complexity and verification. We give a complexity-theoretic argument that sampling from hIQP circuits is classically intractable (Section III A). Our argument is based on an analytical and numerical study of the scrambling properties of the hIQP circuits. We find that already after two rounds of gates on all hypercube edges the output states are close to maximally scrambled. We also show that the runtime of existing classical simulation methods, in particular the recently developed near-Clifford simulator for degree-D circuits [33], roughly scales as $\Omega(2^n)$. The increase in complexity compared to the $O(2^{n/2})$ scaling from Ref. [27] stems from additional gate layers; while the difference to the $O(2^{n/3})$ scaling from Ref. [33] is due to additional random in-block permutation CNOT gates.

We also analytically and numerically study the behaviour of the linear cross-entropy benchmark (XEB) which has been used to benchmark global circuit behaviour [15, 16, 34, 35] (Section III B). We show that—similar to random quantum circuits [16, 22, 36]—the XEB in sparse degree-D IQP circuits (a toy model of hIQP circuits) undergoes a transition as a function of the local noise rate between being a good proxy for the global state fidelity and being much larger than the fidelity. At the same time, we argue that the relation between fidelity and XEB is much tighter in the case of IQP circuits compared to Haar-random circuits, making it a good measure for quantum advantage. In particular, we show that the transition can be arbitrarily shifted by adding fixed gates to an otherwise random circuit. For small instances, we numerically confirm a tight relation between the average fidelity and average linear XEB at low noise rates for both physical and logical noise. In this context, we discuss intriguing aspects of how to think about logical

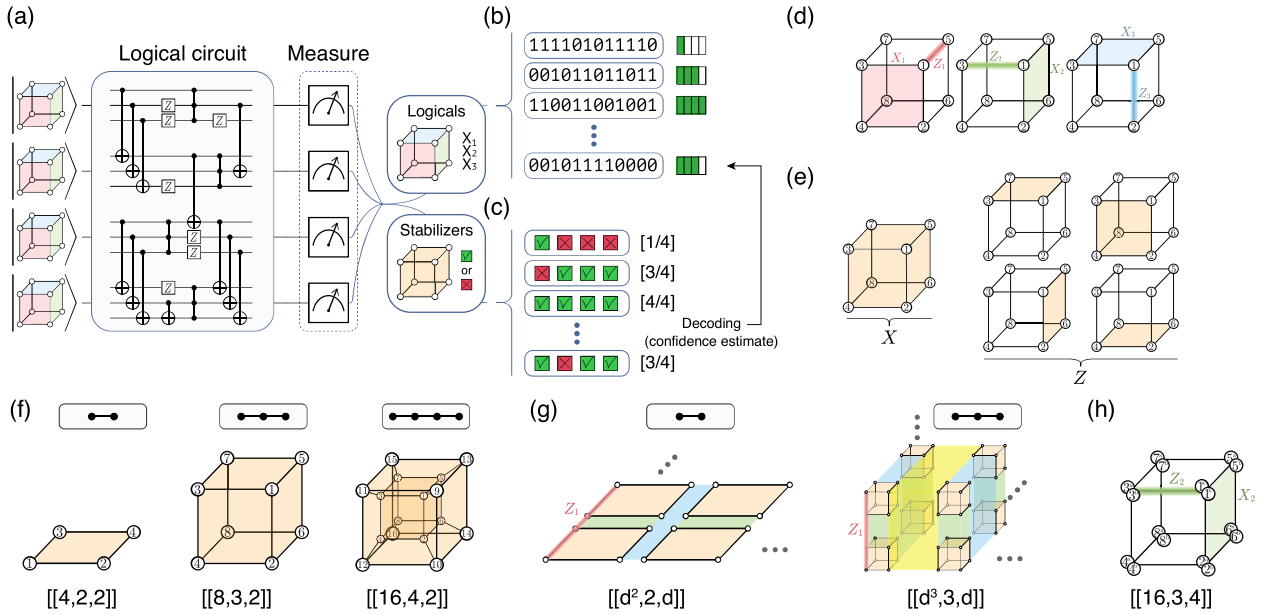

FIG. 1. **Sampling from IQP circuits encoded in hypercube codes.** (a) The $[\![2^D, D, 2]\!]$ code family features a transversal gate set that enables classically hard computation. The logical in-block operations are performed with high-fidelity single-qubit rotations and permutations on physical qubits, and inter-block operations are realized via transversal CNOT gates. (b,c) The measured physical qubit bitstring provides a logical bitstring as well as stabilizer check outcomes. Classical post-processing (decoding) can be used to correct errors or post-select (error detection) based on the confidence for each measured logical bitstring; both procedures improve sampling performance. (d) The $[\![8, 3, 2]\!]$ code is an instance of the family ($D = 3$, $d = 2$). It can be visualized as a cube with logical X and Z operators corresponding to the three orientations of faces and edges, respectively. (e) The single X-basis stabilizer is a product of Pauli-X on all eight qubits, and the Z-basis stabilizers consists of the Pauli-Z products on four independent faces of the cube. (f) A D-dimensional code in the $[\![O(d^D), D, d]\!]$ family supports a transversal $\mathrm{C}^k\mathrm{Z}$ operation within a block [31], where $k \in \{1, \ldots, D-1\}$. For distance $d = 2$, a D-dimensional code is represented by a hypercube of dimension D. The X stabilizer is the product of all qubits, and the Z stabilizers are formed from all independent faces. (g) A distance-d code can be constructed by placing distance-2 codes on a D-dimensional lattice (for even d). This construction preserves the single-qubit realization of transversal in-block gates, and we argue that only codes with even distance can have this property. (h) An alternative approach is to decorate the four vertices with small repetition codes. This results in smaller codes with a higher rate of encoding, but it comes at the expense of introducing multi-qubit gates for transversal in-block operations.

fidelities in encoded sampling circuits (Section III C).

To analyze the scrambling properties as well as the linear XEB of the hIQP circuits, we develop a broadly applicable theory for second-moment quantities of degree-D IQP circuits (Section III D). This theory is based on a mapping of second-moment quantities to a classical statistical-mechanics model, analogous to a similar model for Haar-random circuits [22, 37–40]. It can handle CNOT entangling gates as well as noise on one or two copies of the circuit. As a first application of the statistical-mechanics model we prove that sparse degree-2 IQP circuits anticoncentrate in logarithmic depth. The understanding of the dynamics of the model gleaned from the proof of this result then helps us analyze the asymptotic properties of the hIQP circuits as well as the dynamics of noisy IQP circuits.

To go beyond the inefficient verification of the hIQP circuits via XEB, we propose a natural two-copy hIQP protocol which involves a measurement in the transversal Bell basis at the end of the circuit [41] (Section IV). We show that sampling from the corresponding distribution is classically intractable for degree-4 and higher circuits. Interestingly, the simulation cost of these circuits halves to roughly $\Omega(2^{n/2})$ while the number of logical qubits needs to be doubled. At the same time, the samples can be efficiently classically validated using properties of the two-copy measurement [41]. hIQP Bell sampling thus constitutes a near-term achievable means of performing efficiently validated quantum advantage in a fault-tolerant setting.

Finally, in Section V, we consider the question of scalable fault tolerance of hIQP circuits beyond implementations in the $[\![8, 3, 2]\!]$ color codes. To this end, we devise two code families with the same transversal gate set. The first one is a family of 3D color codes based on the $[\![8, 3, 2]\!]$ code and has parameters $[\![O(d^3), 3, d]\!]$, see Fig. 1(g). The other family is a 3D toric/color code and has parameters $[\![O(d^3), 3, d]\!]$. Both constructions result in families of topological quantum codes and have a fault-tolerance threshold under a local stochastic noise model [26], prov-

ing the scalability of this approach. We also quantitatively compare the performance of error detection in the hIQP circuits using the $[\![8,3,2]\!]$ code with the same circuits implemented in slightly larger codes that support error correction—the $[\![15,1,3]\!]$ code and a $[\![16,3,4]\!]$ code (depicted in Fig. 1(h)). Interestingly, we find that the $[\![8,3,2]\!]$ code with error detection outperforms the other small codes.

Overall, our results highlight that the design of fault-tolerant quantum algorithms requires careful analysis of the algorithm as well as the fault-tolerance properties of the employed error correcting codes for the particular circuits. Most pointedly, this is highlighted by our comparisons of the performance of the 3D hypercube code with error detection with similar small codes where error correction is possible.

B. Implications

Our analysis indicates that sampling from the output distribution of low-depth hIQP circuits is a viable path towards a fault-tolerant quantum advantage over classical computation. hIQP sampling is hardware-efficient on the recently realized logical processor using reconfigurable atom arrays [27]. For intermediate-scale implementations with error detection and/or correction at the end of the computation, the $[\![8,3,2]\!]$ 3D color code has not only the highest rate but also the best fault-tolerance properties when compared to similar small codes. In the experiment of Bluvstein *et al.* [27], hIQP sampling without in-block permutation CNOT gates was performed on 48 logical qubits encoded in 2^4 blocks of the $[\![8,3,2]\!]$ code. These specific circuits exhibit asymptotic quantum advantage but can be simulated in time $\sim 2^{n/3}$ [33]. This classical scaling is considerably more favorable than what we expect for hIQP circuits studied in this work that include in-block CNOT gates. For these circuits, which can be directly implemented using the techniques demonstrated in Ref. [27], we expect the best classical simulation algorithms to run in time at least roughly $\sim 2^n$.

Our results showing the existence of a transition in the relation between XEB and fidelity imply that experiments should require a careful analysis of the noise regime in order to use XEB as a reliable benchmark. Our analysis suggest that the relation between XEB and fidelity may be tighter for IQP circuits compared to random circuits [16, 22, 36], but exactly how so remains an open question. The existence of a transition highlights the need for fault tolerance in experiments benchmarked via XEB since the local error rate needs to be suppressed to $\sim 1/n$ for the XEB to be a good estimator of fidelity. For the error rates and circuit parameters of Ref. [27], we find that the experiments are in the "healthy" noise regime in which the average logical XEB score is a good measure of an appropriately defined logical fidelity.

Our results also suggest a path to scale up hIQP sampling experiments to achieve fault-tolerant quantum advantage. Two natural improvements can be achieved using the $[\![2^D, D, 2]\!]$ hypercube color codes: First, one can perform quantum advantage experiments that can be efficiently validated from the classical samples by using transversal Bell measurements between two copies of similar hIQP circuits based on the $[\![16,4,2]\!]$ code. Note that using the $[\![8,3,2]\!]$ code, a classically simulatable Bell sampling experiment has already been performed for 24 logical qubits by Bluvstein *et al.* [27, Fig. 6]. Second, one can improve the fault-tolerance properties by performing intermediate measurements to detect more errors and improve the final state fidelity. Going beyond error detection will require scaling up the code sizes. We construct some candidate families of 3D color codes with an error correction threshold that support the classically hard hIQP circuits. If the experimental error rate is below the threshold, the quantum output distribution converges exponentially to the target distribution as the code size is increased. However, the rate of these codes decreases with the code distance. Finding high-performing quantum codes similar to [42, 43], that also support native non-Clifford gates remains an important open question to further scale up quantum advantage demonstrations of the type proposed here. For example, one interesting possibility we leave for future work is to use a $[\![64,15,4]\!]$ CSS code where a transversal T-gate implements a logical circuit with 15 CCZ gates [44].

C. Overview

Section II outlines the fault-tolerant logical sampling architecture based on $[\![2^D, D, 2]\!]$ codes and degree-D hIQP circuits with error detection. Section III A describes the hardness properties of our circuits, Section III B the behaviour of the XEB under noise, Section III C the relation between the logical XEB and fidelity in encoded circuits, and Section III D our statistical mechanics mapping. Section IV addresses efficient verification by introducing and showing hardness of degree-D Bell sampling. Section V compares the performance of different small codes that allow limited error correction to the $[\![8,3,2]\!]$ code and describes the new family of $[\![O(d^D), D, d]\!]$ color codes that allow scalable transversal IQP sampling in the presence of noise. Section VI concludes with an outlook. Appendices A to H contain various technical details and proofs.

II. LOGICAL SAMPLING ARCHITECTURE

In this section, we detail the concrete architecture for IQP circuit sampling using logical circuits that employ only transversal and permutation gates. The architecture is co-designed with the capabilities of a neutral-atom logical quantum processor [27–29, 32] in order to generate highly scrambled quantum states that are classically intractable to simulate and at the same time minimize

errors incurred during the computation. To this end, we exploit parallel transversal control of logical qubits by moving atoms in arrays of tweezers, which can naturally realize variable grid-like entangling patterns [27]. These patterns will be reflected in a nested-hypercube structure of the physical qubit quantum circuits we study.

As described in the introduction, our proposed architecture uses a family of distance-2 error detecting codes, which have a family of non-Clifford transversal gates. This circumvents the need for magic state distillation, since the transversal gates efficiently realize classically intractable operations. The computational task we consider is sampling from the distribution of *logical qubit* states. The redundancy afforded by the error-detecting code is used to improve the quality of the sampled distribution, therby yielding an advantage compared to unencoded computations. Moreover, several complex many-qubit operations can be executed much more cheaply than in a comparable bare physical circuit. The associated cost of the relatively complex encoding step is mitigated by the fact that this state preparation process can be made fault-tolerant across the qubit array in a scalable manner through parallelized postselection and re-preparation.

A. A family of distance-2 codes and their native operations

Our architecture is based on the D-hypercubic code with code parameters $[\![2^D, D, 2]\!]$ for $D \geq 2$ [46, Example 3]. The hypercubic code is a color code defined on a D-dimensional hypercube, that is, its stabilizers are given by products of X and Z operators on the faces of the hypercube. The colorability of the D-dimensional hypercube determines the number of redundant stabilizers and hence the number of encoded logical qubits, which for the hypercube is just given by D.

The $[\![2^D, D, 2]\!]$ code supports a logical gate-set comprising all Pauli gates, and logical

$$\mathrm{C}^k\mathrm{Z} \;=\; \left(\mathbb{1}_k - \left|1^k\right\rangle\left\langle 1^k\right|\right) \otimes \mathbb{1} \;+\; \left|1^k\right\rangle\left\langle 1^k\right| \otimes Z \quad (1)$$

gates for $k = 1, \ldots, D-1$ on any $k+1$ of the D encoded qubits. Importantly, for $k \geq 2$, the $\mathrm{C}^k\mathrm{Z}$ gate is a non-Clifford gate. All of these operations are physically implemented with single-qubit rotations on a subset of the physical qubits. These single-qubit operations can be implemented in parallel with very high fidelities of $\sim$0.9999 [47, 48].

The three-dimensional instance of the code with parameters $[\![8, 3, 2]\!]$ is considered "the smallest interesting color code" [49]. It has a high encoding rate and allows us to perform error-detected non-Clifford computations. In the following, we will use the $[\![8, 3, 2]\!]$ code as the main focus of our study to illustrate concepts that apply more generally to the $[\![2^D, D, 2]\!]$ code. The logical operators and stabilizers of the $[\![8, 3, 2]\!]$ code are illustrated in Fig. 1(d,e), the transversal in-block operations in Fig. 2(b).

The Hadamard gate completes the universal gate set for the hypercubic codes, as for other high-dimensional codes, so it cannot have a transversal implementation [50]. Nonetheless, since the $[\![2^D, D, 2]\!]$ code is a CSS code, the logical $\left|+^D\right\rangle$-state can be prepared fault-tolerantly [7, 51]. In Fig. 2(d), we show a circuit that prepares the $|{+}{+}{+}\rangle$ state of a single $[\![8, 3, 2]\!]$ code block in a non-fault-tolerant way. This circuit can be understood as preparing a GHZ state of two $[\![4, 2, 2]\!]$ blocks in opposite bases and then applying a transversal CNOT between these blocks, and directly generalizes to higher D. A fault-tolerant preparation can be achieved with three additional physical flag qubits [45]. Moreover, measurements in both the logical X and the logical Z-basis are transversal.

Since the $[\![2^D, D, 2]\!]$ is a CSS code [7, 51] the transversal CNOT gate implements a transversal CNOT gate on the D encoded qubit pairs, see Fig. 2(e). We entangle code blocks by applying the transversal CNOT gate in parallel on all physical qubit pairs. We can further implement in-block CNOT and SWAP gates between arbitrary pairs of logical qubits within each block and, using this capability, also between any inter-block pair. Both of those gates can be realized just by permutations/relabelling of the physical qubits, see Fig. 2(c), which comes at almost no extra cost. While permutations on their own do not grow the weight of errors, when combined with transversal entangling gates they are no longer fault-tolerant. In order to maintain fault tolerance, a round of error detection should be applied after each permutation gate. We note that, for the circuits studied in this work, fault tolerance is actually maintained for Z errors and a specific sequence of instructions needs to occur to break it for X and Y errors. Moreover, one can efficiently check if fault tolerance is preserved for any random circuit instance. The CNOT gate between two specific qubits in two distinct blocks can be realized using two transversal CNOTs interlaced with in-block CNOT and SWAP gates, see Fig. 2(f) and the inter-block SWAP gate by compiling it from inter-block CNOT gates. Combined with in-block permutation gates, arbitrary configurations of controls and targets can be realized in this way.

Despite the blocked structure of the gates, with these gadgets we can realize any-to-any logical qubit connectivity by implementing SWAP gates between two specific logical qubits in different blocks. This capability enables us to extend the in-block operations to any three qubits in the system—we can simply swap targeted qubits into a block, perform the desired diagonal operation, and then swap them back to their initial locations. Together with the transversal CNOT gates, the capability to measure in the logical X and Z basis also makes the Bell measurement transversal.

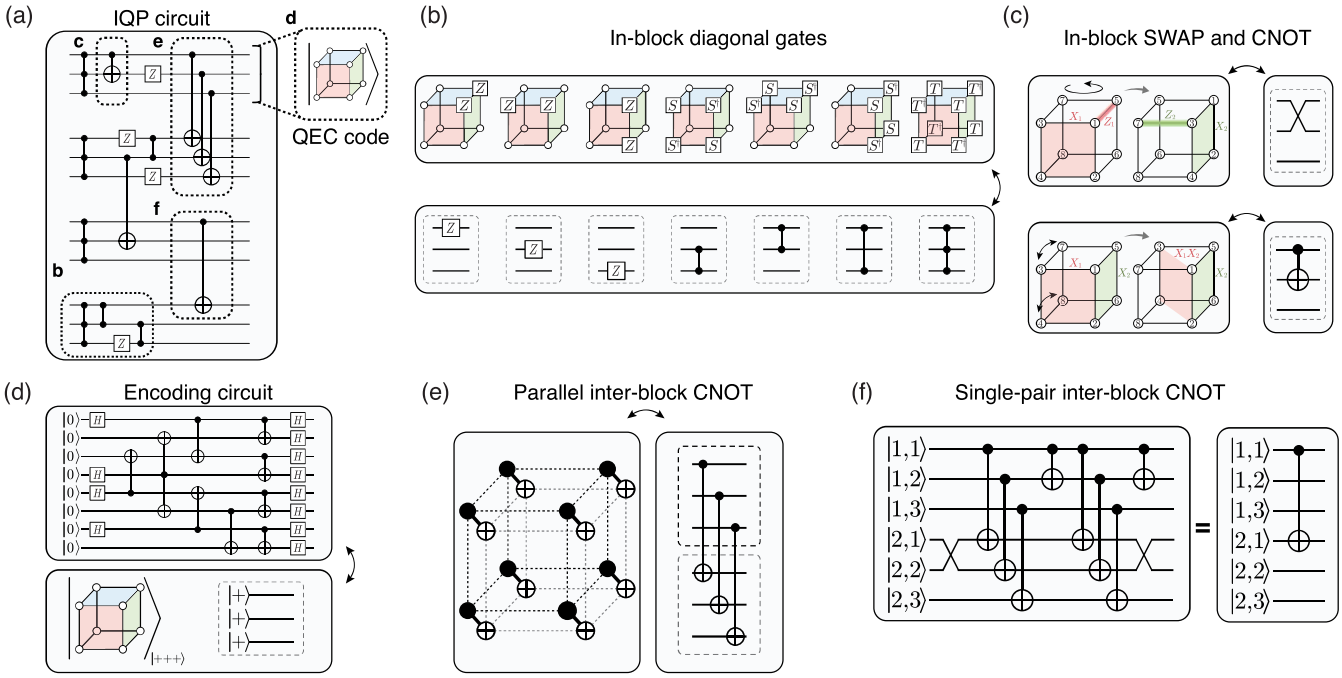

FIG. 2. **IQP circuits as fault-tolerant operations on the $[\![8,3,2]\!]$ code.** (a) Arbitrary degree-3 IQP circuits can be decomposed into the transversal gate-set of the $[\![8,3,2]\!]$ code and its extensions. The logical qubits are grouped into sets of three and encoded into code blocks consisting of eight physical qubits. (b) The diagonal gates between logical qubits within a single block consist of all Z, CZ, and CCZ gates. These operations and are implemented with parallel single-qubit rotations on the physical qubits. (c) Additional transversal in-block operations include the CNOT and SWAP gates, which are realized by simple re-labeling of the qubits. In hardware, re-labeling can be realized by exchanging the qubits' positions. (d) State preparation circuit for preparing a symmetric logical state $|{+}{+}{+}\rangle$. In this preparation circuit and its generalizations, the method can be made fully fault-tolerant by utilizing flag protocols or Steane error detection [45] at each stage of the protocol in a recursive manner. (e) Two encoded blocks can be entangled via a transversal CNOT, which is guaranteed by the CSS nature of the code. (f) Two layers of transversal out-block CNOTs interspersed with in-block CNOTs can realize a CNOT between individual logical qubits from different blocks, which enables transversal realization of arbitrary IQP circuits.

B. Random logical IQP sampling

Our proposal is based on sampling from encoded degree-D IQP circuits. These circuits comprise Z, and $\mathrm{C}^k\mathrm{Z}$ gates for $k = 1, \ldots, D-1$ between arbitrary subsets of qubits with state preparation and measurement in the X basis. We can implement those circuits as follows. Consider a system comprising b blocks of the $[\![2^D, D, 2]\!]$ code. We can then prepare the entire system in the $\left|+^{bD}\right\rangle$ state, and implement arbitrary in-block degree-D IQP circuits using the transversal gate set. In order to apply an IQP gate to an arbitrary subset of the qubits, we can use the SWAP gadget discussed in the previous section (Section II A) to swap the involved logical qubits into a single block, implement the transversally realized IQP gate in that block, and then swap the logical qubits back (or to a different block). Finally, we can measure all blocks in the logical X basis. Thus, we can implement sampling from arbitrary encoded degree-D IQP circuits.

We will now consider random degree-D IQP circuits. In a *uniformly* random degree-D IQP circuit C every gate from the gate set $\{Z,\ \mathrm{CZ}, \ldots,\ \mathrm{C}^{D-1}\mathrm{Z}\}$ is applied with probability one half to every subset of qubits with the corresponding size, i.e., for every k-subset of qubits, a $\mathrm{C}^{k-1}\mathrm{Z}$ gate is applied with uniform probability. Bremner *et al.* [24] have shown that approximately sampling from the output distribution p_C of an n-qubit uniformly random degree-D IQP circuit C with probabilities

$$p_C(x) = |\langle x| H^{\otimes n} C |{+}^n\rangle|^2, \tag{2}$$

is classically intractable under reasonable complexity-theoretic assumptions for any $D \geq 3$.

In fact, the same is believed to hold true for a *sparse* ensemble of IQP circuits comprising only $O(n \log n)$ gates from a slightly different gate set comprising the control-phase gate CS and the T gate [25]. This ensemble is significantly more hardware-efficient to implement than uniformly random IQP circuits since it can be implemented using only $O(\log n)$ parallel gate layers. Recently, Paletta *et al.* [52] have demonstrated that this ensemble of IQP circuits can be implemented fault-tolerantly in unit depth using a certain family of quantum codes they call *tetrahelix codes*. Tetrahelix codes are constructed from another three-dimensional color code, namely, a tetrahedral code whose smallest instance is the $[\![15, 1, 3]\!]$ Reed-Muller code

with a transversal T gate. Similar constructions make use of fault-tolerant measurement-based quantum computing [53–55].

A plausible analogous and resource-efficient approach that we could take would be to define a sparse ensemble of uniform degree-D IQP circuits, argue that it is hard to simulate, and then implement this ensemble using the gadgets from the previous section. This would yield a possible path towards demonstrating quantum advantage using only native operations of the $[\![2^D, D, 2]\!]$ code. However, this approach has the disadvantage that the SWAP gadgets required to implement $\mathrm{C}^k\mathrm{Z}$ gates on arbitrary subsets of qubits require an overhead compared to a single gate: three CNOT gates are required to implement a single SWAP gate between an arbitrary pair of qubits, each of which requires two transversal inter-block CNOT gates, in-block qubit permutations, and additional error correction gadgets. As detailed in the next subsection, we develop for a more hardware-efficient approach to demonstrating quantum advantage that we show converges asymptotically to the behavior of random IQP circuits. In this sense, it constitutes a (fault-tolerant) compilation of random IQP circuits.

We close this subsection with some general remarks on the relation between logical circuits realizable via transversal gates and IQP circuits. In the case of qubit stabilizer codes, a rigorous proof exists that all transversal gates must be within a finite level of the Clifford hierarchy [56, 57]. The set of unitaries in the hierarchy has also been partially classified and it is conjectured (as well as proven for the 3rd level of the hierarchy) that they all take the form of so-called generalized semi-Clifford operations [58, 59]. Moreover, as noted above, the set of transversal gates forms a discrete group on an error detecting code [60]. The finite groups formed out of generalized semi-Clifford unitaries in the kth level of the hierarchy have generators of the form (up to some technical caveats) [61]

$$U = C^\dagger P D C, \tag{3}$$

where C is a Clifford unitary, D is a diagonal circuit in the kth level (classified in [62]), and P is a permutation circuit in the kth level. Up to the basis change by the Clifford circuit C, these unitaries are exactly of the IQP type. As a result, there seems to be fundamental connection between transversal computation on stabilizer codes and IQP-like circuit dynamics. Interestingly, if one considers non-stabilizer codes, the family of transversal operations that is achievable on qubit error detecting codes includes gates that lie completely outside the Clifford hierarchy [63, 64].

C. Hypercube IQP circuits

In order to optimize the hardware efficiency, we focus on a gate set which can be fully parallelized, namely, inter-block transversal CNOT gates as well as in-block degree-D IQP and CNOT circuits. Any in-block degree-D circuit can be compiled in a single layer of physical single-qubit gates which are powers of the T gate, and an in-block CNOT circuit just requires physically permuting the atoms constituting the code block. In order to achieve fast scrambling of the quantum circuits while exploiting the long-range parallelization possible in the reconfigurable atom arrays, we will consider an interaction graph between code blocks given by a $\mathfrak{D}$-dimensional hypercube. In general, quantum dynamics in hypercube geometry exhibit fast (Page) scrambling [65, 66] due to the expansion properties of hypercube graphs.

When compiling the hypercube IQP circuits, our first observation is that CNOT gates leave the family of degree-D IQP circuits invariant under conjugation. In other words, a degree-D IQP circuit with CNOT gates is equivalent to some other degree-D IQP circuit. To see this, we exploit the fact that a circuit comprised of CNOT gates just acts as a linear map on bit strings, see Appendix A 2. Given this observation, we define the ensemble of *hypercube IQP (hIQP)* circuits (with degree D, dimension $\mathfrak{D}$, and hyperdepth $\mathfrak{L}$) as follows, see Fig. 3(a):

1. Prepare the logical $\left|+^D\right\rangle^{\otimes 2^{\mathfrak{D}}}$ state.
2. Perform $\mathfrak{L}$ *hypercube layers* each comprising
   i. uniformly random logical degree-D circuits and uniformly random logical CNOT gates in every block alternated with
   ii. transversal CNOT gates with random orientation between blocks

   on all $\mathfrak{D}$ sets of parallel edges of the hypercube.
3. Perform a last layer of uniformly random logical degree-D in-block circuits.
4. Measure all logical qubits transversally in the logical X-basis.

After encoding of the code blocks, a hyperdepth-$\mathfrak{L}$ hIQP circuit (i.e., a circuit with $\mathfrak{L}$ hypercube layers) thus requires $\mathfrak{D} \cdot \mathfrak{L} + 1 \in O(\log n)$ physical gate layers each comprising a layer of in-block degree-D circuits, a layer of in-block CNOT gates, and (except the last) a layer of transversal CNOT gates. Finally, the transversal CNOT gates along parallel edges of the hypercube can be physically realized in the reconfigurable processor by interlacing a pair of two-dimensional grids of qubits, see ED Fig. 6 of Ref. [27]

We can also consider variants of random hyperdepth-$\mathfrak{L}$ hIQP or combinations thereof that are motivated by experimental feasibility, or fault-tolerant properties. In one variant, random IQP gates are applied to a block if and only if it was the target of a CNOT gate in the previous layer. This is because diagonal gates commute through the control of a CNOT gate so that the two random in-block IQP circuits before and after the gate give rise to a new random IQP circuit. In this variant, $\mathrm{C}^{D-1}\mathrm{Z}$

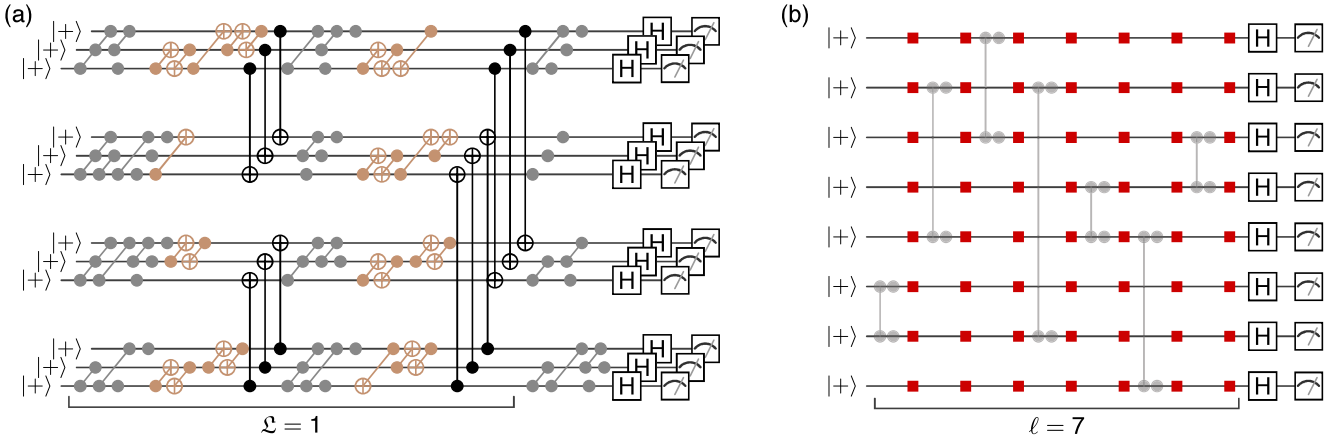

FIG. 3. **IQP circuit families.** (a) Random $\mathfrak{D} = 2$ hypercube IQP (hIQP) circuit on 4 blocks of the $[\![8,3,2]\!]$ code with hyperdepth $\mathfrak{L} = 1$. In-block transversal single-qubit gates realize a degree-3 in-block circuit (gray). In-block CNOT gates are realized through permutations of the physical qubits (ochre). Transversal CNOT gates couple the code blocks (black). The measurement is performed transversally in the X basis. (b) In a sparse degree-2 IQP (sIQP) circuit as we consider it here, in every one of the ℓ circuit layers, a random degree-2 circuit consisting of a random CZ and Z gates (vertices) is applied to a random pair of qubits, which in the noisy setting is followed by a parallel layer of Pauli noise (red squares).

gates are also applied deterministically, and single-qubit Z gates are only applied in the last circuit layer, since they commute through the circuit. This variant of hIQP circuitswhich furthermore did not involve the in-block CNOT gates as realized by physical atom permutations, was implemented in Ref. [27]. This has the advantage that of requiring less error detection to maintain fault tolerance compared to the circuits considered here. As was shown recently, however, the resulting symmetry in the circuit can be exploited to reduce the exponent in the time required to simulate these circuits [33]. This motivates including the in-block random CNOT layers.

The logical hIQP circuit has several interesting interpretations in terms of nested hypercubes. Consider for concreteness the $[\![8,3,2]\!]$ code. First, observe that the interaction graph of the entire physical circuit is a $(\mathfrak{D}+3)$-dimensional hypercube of individual physical qubits, since the $[\![8,3,2]\!]$ code blocks themselves are defined on a three-dimensional cube. Then, consider the encoding circuit of the $|{+}{+}{+}\rangle$ state. This state is created by transversally entangling two logical $|{+}{+}\rangle$ states of the $[\![4,2,2]\!]$ code, see Fig. 1(d). Thus, the $\mathfrak{D}$-dimensional hypercube of $[\![8,3,2]\!]$ codes can be reinterpreted as a $(\mathfrak{D}+1)$-dimensional hypercube of $[\![4,2,2]\!]$ codes. This idea generalizes to encoding the logical $|{+}^D\rangle$ state of the $[\![2^{D-1}, D-1, 2]\!]$ code using two sets of logical $|{+}^{D-1}\rangle$ states of the $[\![2^{D-1}, D-1, 2]\!]$ code. If the single-qubit rotations are altered, we can thus understand the physical hIQP circuit as state preparation of a more general $[\![2^{\mathfrak{D}+D}, \mathfrak{D}+D, 2]\!]$ code.

Finally, we also note that the output states of degree-D IQP circuits before the last layer of Hadamard gates are also known as hypergraph states with hyperedges comprising up to D qubits [67]. Hypergraph states have been studied in the quantum information literature [68, 69] and can serve as a resource for measurement-based quantum computing [70–72].

D. Fault-tolerant sampling with error detection

The main advantage of sampling using logical encodings is the resilience to errors. Typically, quantum error correction consists of many rounds of syndrome extraction repeated throughout the circuit, in order to prevent errors from spreading and, thus, suppress the probability of logical faults. While this approach will be eventually necessary to realize large-scale quantum algorithms, our fault-tolerant compilation of IQP circuits allows us to benefit from using encoded qubits with limited rounds of mid-circuit measurement. In the extreme case, we can perform the syndrome measurement only at the end which significantly simplifies experimental implementation. We focus primarily on the $[\![8,3,2]\!]$ color code, which is the smallest code with the desired properties, due to its high rate and small size.

In Fig. 4(a), we show the linear XEB score for hIQP circuits with $\mathfrak{D}=4$ and $\mathfrak{L}=2$ using $[\![8,3,2]\!]$ codes, with various amounts of postselection based on evaluating the X-type stabilizer from the measurement data. For each measurement, we set a threshold on the number of violated stabilizers of which there are one per block and discard the samples that fail this test. The fraction of the remaining samples is the *acceptance ratio*. We observe a significant increase in the XEB score as we increase the amount of postselection in the system. The XEB curve approximately obeys a power law $\propto x^{7/8}$, which is consistent with a simple independent and identically distributed (i.i.d.) error model description; see Appendix C. Roughly speaking, the scaling arises from the fact that

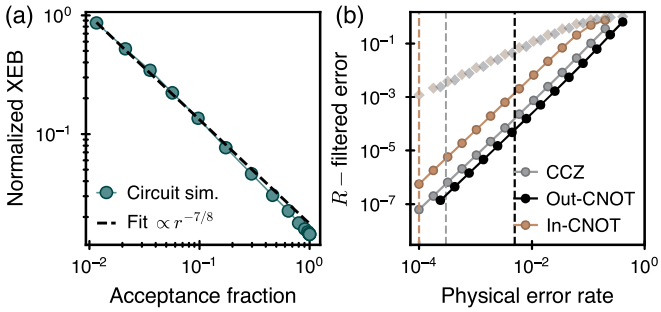

FIG. 4. **Performance of encoded sampling in hIQP circuits.** (a) The logical XEB score can be improved with increased postselection, at the cost of fewer samples. The relationship between the XEB and acceptance fraction follows an approximate power law, which for the $[\![8,3,2]\!]$ code has an exponent close to $-7/8$. The simulation was performed on a single random instance of ($\mathfrak{D}=4$, $\mathfrak{L}=2$) hIQP circuit (128 physical, 48 logical qubits) with single-qubit depolarizing noise corresponding to 99.9% and 99.5% fidelities for one- and two-qubit gates, respectively. The circuit includes only the Clifford subset of available operations, and the data points are averaged over $\sim 10^6$ samples. The normalized XEB is evaluated relative to the ideal value for this circuit $\chi_{\text{ideal}}=3$. (b) Relative performance of logical operations within a circuit can be compared using the notion of R-filtered fidelities with $R=0$ (opaque) and $R=2$ (semi-transparent). The three main operations in our circuits are the in-block diagonal gates, out-block CNOT gates, and in-block CNOT gates which are realized on the physical level by single-qubit gates, atom transport, and two-qubit gates, respectively. The vertical lines correspond to characteristic error rates for the relevant operations on the physical level. See Appendix B for more details.

one out of the eight physical qubits does not participate in any logical operator. Using this approach, Ref. [27] realized IQP circuits on 48 logical qubits with logical fidelities above those achievable with physical qubit implementations.

Importantly, the in-block CZ and CCZ gates (single-qubit rotations), as well as the in-block CNOT (atom transport) many-body logical operations in the $[\![8,3,2]\!]$ code are implemented with single-qubit operations on the physical level. Since these single-qubit operations can be realized with high fidelity, a logical implementation of degree-D IQP circuits can be advantageous compared to a physical one even in the absence of error detection. In fact, we expect that errors in the circuits studied here are dominated by the inter-block CNOT gates, rather than the in-block CZ and CCZ gates.

To quantify the relative performance of our circuit elements we introduce and use the concept of R-filtered fidelity, which measures how a given (many-body) operation affects an input state with up to weight-R physical errors, see Appendix B for details. Refining the fidelity in this way is helpful in analyzing encoded computations. There, in contrast to bare computations, entanglement builds up between syndrome and logical degrees of freedom in a context-dependent way, giving rise to an intrinsically non-Markovian error model. For example, for a computation encoded in a distance-2 fault-tolerant circuit, the 0-filtered fidelity of any single-qubit operation is always 1. This is because no single-qubit error can propagate to a logical one. In contrast, the 1-filtered fidelity is far from unity, since a "background" physical error can combine with the single-qubit operation to cause a logical fault. In Fig. 4(b), we plot the R-filtered fidelity and see that the inter-block CNOT indeed introduces the most errors out of our basic operations. The capability to realize the non-Clifford CCZ gates with high fidelity is therefore a central element of our proposal, since in most scenarios it is the most costly gate: physical implementations of the CCZ often have much lower fidelity than two-qubit gates, and typical QEC codes require expensive magic state distillation protocols to realize them. Thus, we expect that this fault-tolerant approach to classically hard circuits can result in very high XEB scores even in early logical quantum processors, as demonstrated in Ref. [27].

III. COMPLEXITY AND VERIFICATION OF (H)IQP CIRCUITS

Let us now turn to a detailed investigation of random hIQP circuits as defined in Section II C and their properties relevant to quantum advantage demonstrations. We will address two questions in detail. The first question regards the classical complexity of sampling from the output distribution of the circuits and, second, how to verify the correctness of the samples. We do so both in the finite-size and the asymptotic regimes. As mentioned above, the complexity of various ensembles of IQP circuits has been studied by Bremner *et al.* [24, 25] using a complexity-theoretic argument for the hardness of sampling from random quantum circuits based on Stockmeyer's algorithm; see Ref. [19] for a review of that argument. Here, we will argue that this argument applies also to random hyperdepth-2 hIQP circuits. To do so, we make use of a measure of scrambling called *anticoncentration* that features in this argument and serves as an indicator of classical hardness [18, 19, 24]. We study anticoncentration of random hIQP circuits in Section III A.

The second question regards verification. Verifying classical samples from random quantum circuits is a challenging problem, one that in fact requires exponentially many samples if the goal is unconditional verification [73]. What has become the standard approach to verification in this scenario is to make use of the linear *cross-entropy benchmark (XEB)* [15, 74]. This benchmark can be evaluated using only a small (polynomial) number of samples but involves computing ideal output probabilities, which can require exponential time. It is appealing, however, because it has been argued that it can be used as a proxy for the fidelity of the quantum pre-measurement states averaged over instances of the random circuit [15, 40]. It may also witness the achievement of a computationally

complex task [36, 75, 76]. The XEB has been studied in detail for random universal circuits, revealing the value of the XEB that we would obtain for ideal circuits, as well as the behaviour of the XEB and fidelity under local circuit noise [16, 22, 36, 39]. We study the linear XEB for hIQP circuits in Section III B.

It turns out that both the anticoncentration property and the linear XEB can be written as a second moment of potentially noisy hIQP circuits. The tool of our analytical study of both complexity and verification of hIQP circuits will be a mapping of the behvaiour of such second-moment quantities for general degree-D IQP circuits of varying circuit depth to the dynamics of four classical states. We introduce this mapping and discuss its most important properties in Section III D, deferring details to Appendices D and F.

A. Complexity of hIQP circuits

To set the stage for our results on the scrambling and complexity properties of random hIQP circuits, let us recap the results on uniform IQP circuits by Bremner *et al.* [24]. They show that output distributions of circuits comprising $\{Z,\mathrm{CZ},\mathrm{CCZ}\}$ in uniformly random locations, i.e. uniform degree-3 circuits, are classically intractable to simulate under certain complexity-theoretic assumptions. Specifically, there is no efficient classical sampling algorithm for uniformly random IQP circuits, assuming that the output probabilities (2) are #P-hard to approximate on average over random IQP circuits, unless the widely believed conjecture that the polynomial hierarchy does not collapse is false. The key assumption on which the result hinges is therefore the approximate average-case hardness of computing the outcome probabilities. To give evidence for this conjecture, Bremner *et al.* [24] show approximation hardness for degree-3 IQP circuits *in the worst case* and a strong version of the so-called *anticoncentration* property.

At a high level, the anticoncentration property ensures that the output distribution of random IQP circuits has barely any structure that might be exploited by an approximate classical sampling algorithm to speed up the computation compared to the worst case. Technically, the (strong) anticoncentration property is the statement that the average second moment of the output distribution of a random circuit family $\mathcal{C}$ as measured by the ideal linear XEB score is constant as

$$\overline{\chi}_{\mathcal{C}} := 2^{2n} \cdot \mathbb{E}_{C\leftarrow\mathcal{C}}\left[p_C(x)^2\right] - 1 \in O(1). \tag{4}$$

The anticoncentration property ensures that in order to sample from the correct output distribution it is not merely sufficient to identify the dominant outcomes in the distribution. Rather, almost all probabilities need to be computed to exponential precision in order to perform the sampling correctly. In the following, we will use anticoncentration as a key—but not the sole indicator—that sampling from non-uniform IQP circuits remains intractable. For uniform degree-D (uIQP) circuits with any $D \geq 2$, Bremner *et al.* [24] show that

$$\overline{\chi}_{\mathrm{uIQP}} = 2 - 2^{-n+1}. \tag{5}$$

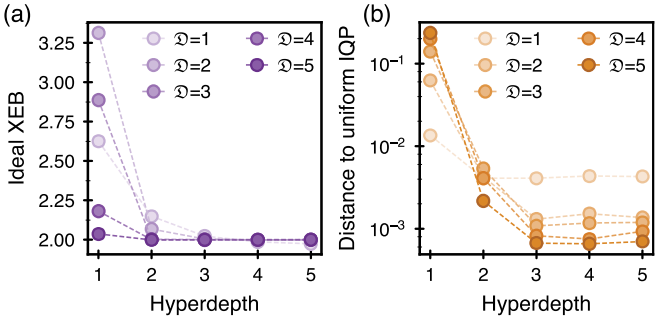

FIG. 5. **Convergence of hypercube IQP to uniform IQP.** The effective IQP circuits implemented by random hIQP circuits converge to uniform IQP circuits as a function of hyperdepth. For each data point, we average $\sim 10^6$ random degree-2 hIQP circuits. (a) Convergence of the ideal XEB score of hIQP circuits to the uniform value $\overline{\chi}_{\mathrm{uIQP}} = 2$ as a function of hyperdepth. (b) Total-variation distance between the distribution over IQP circuits defined by hIQP circuits and uniform IQP.

How do hIQP circuits fare in terms of their complexity? We expect degree-3 and higher hIQP circuits to be classically hard to simulate already at constant hyperdepth because of their large number of non-Clifford gates and the expander properties of the hypercube. Therefore, we expect that the quantum dynamics scramble very quickly in this geometry. Fast scrambling on hypercubes has in fact been observed in Ref. [66].

We provide analytical evidence for this in two steps. We first show that degree-2 circuits anticoncentrate if they have $\Omega(n \log n)$ gates, but not for any constant number of gates. To this end, we consider a model of sparse random degree-2 IQP (sIQP) circuits in which a CZ gate and Z gates are applied with probability $1/2$ to ℓ random qubit pairs, see Fig. 3(b). For us, sparse IQP serves as a toy model of low-depth hIQP circuits, which have significantly more structure and are therefore more difficult to analyze.

Theorem 1 (Anticoncentration of sparse IQP)**.** *The ideal linear XEB score of sparse degree-2 IQP circuits acting on n qubits with uniformly random Z gates and random CZ gates acting on ℓ random pairs is given by*

$$\overline{\chi}_{\mathrm{sIQP}}(\ell) = \overline{\chi}_{\mathrm{uIQP}} + 2^{-\Omega(\log n)} \qquad \text{if } \ell \geq \frac{11}{4} n \log n \tag{6}$$

$$\overline{\chi}_{\mathrm{sIQP}}(\ell) \in \Omega(2^n) \qquad \text{if } \ell \in O(n). \tag{7}$$

This result is analogue to the sparse anticoncentration result of Bremner *et al.* [77, Lemma 6] for degree-2 circuits. Next, we show that the ideal XEB score of hIQP circuits approaches the uniform score as $\mathfrak{L} \to \infty$. This complements the sparse IQP result by showing that the

structure of hIQP circuits is asymptotically irrelevant to the XEB.

Theorem 2 (Anticoncentration of deep hIQP)**.** *The ideal linear XEB score of degree-2, hyperdepth-$\mathfrak{L}$ hIQP circuits acting on $2^{\mathfrak{D}}$ blocks of D qubits satisfies*

$$\overline{\chi}_{\mathrm{hIQP}}(\mathfrak{L}) \xrightarrow{\mathfrak{L}\to\infty} \overline{\chi}_{\mathrm{uIQP}} \tag{8}$$

We defer the proofs of Theorems 1 and 2 to Appendix E. We note that Theorems 1 and 2, while formulated for degree-2 circuits, provide upper bounds for the XEB value of IQP circuits with additional higher-degree gates. This is because all gates commute and therefore additional gates cannot increase the XEB score.

The two anticoncentration results complement each other. While Theorem 1 implies that uniformly sparse degree-D circuits can anticoncentrate in logarithmic circuit depth, Theorem 2 implies anticoncentration of hIQP circuits as $\mathfrak{L} \to \infty$. Since IQP and hIQP share the same gate set, this suggests that the ensemble of degree-D hIQP circuits approaches degree-D uniform IQP in this limit and, thus, that very deep hIQP circuits are also classically hard to simulate. Together, they lend credence to our claim that hIQP circuits anticoncentrate in constant hyperdepth, meaning that the total circuit comprises $\Theta(n \log n)$ (nonuniform) IQP gates.

To better understand the behaviour of the hypercube circuits at low depths, we numerically compute the ideal XEB value for random hIQP circuits as a function of hyperdepth $\mathfrak{L}$ and hypercube dimension $\mathfrak{D}$. We do so by classically sampling random degree-2 circuits, which are efficiently simulatable. The results, shown in Fig. 5(a), demonstrate that the XEB of hIQP circuits quickly converges to the uniform value as a function of hyperdepth, with hyperdepth-2 circuits being basically converged. We also compare the full ensemble of degree-2 hIQP circuits to uniform IQP and find a very similar behaviour. Indeed, for any $\mathfrak{L} \geq 2$ the hIQP distribution quickly converges to uniform IQP, see Fig. 5(b).

Together, these observations provide evidence that after hyperdepth 2 any structure which might be exploited by a classical simulation algorithm has essentially vanished, rendering classical simulation inefficient, and we have the following conjecture.

Conjecture 3 (hIQP approximate average-case hardness)**.** *Approximating the output probabilities of hyperdepth-2, degree-3 hIQP circuits up to error $O(2^{-n})$ is #P hard for a constant fraction of the instances.*

Note that the output probabilities of degree-D circuits are related to the *normalized gap* of degree-D Boolean polynomials [24] and hence we can equivalently phrase Conjecture 3 in terms of approximating the gap of those polynomials, see Ref. [78] for an in-depth discussion. Conjecture 3 implies hardness of sampling under a widely believed complexity-theoretic conjecture by a standard argument, see Ref. [19] for details.

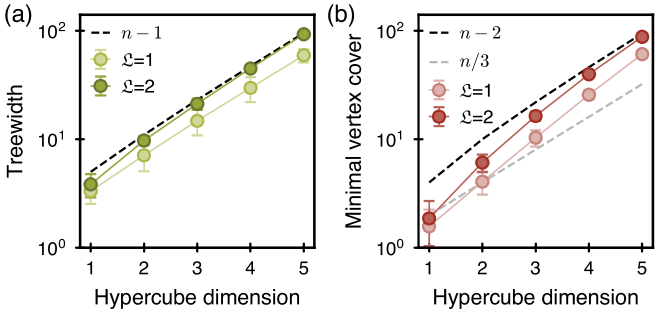

FIG. 6. **Evidence for classical hardness.** To give evidence for the hardness of simulating random hIQP circuits, we analyze the properties of the effective hypergraphs defined by degree-3 hIQP circuits for 100 random instances at hyperdepth $\mathfrak{L} = 1$ and $\mathfrak{L} = 2$. Error bars are one standard deviation. (a) The treewidth determines the complexity of tensor-network contractions [79]. (b) The minimum vertex cover of the degree-3 hyperedges determines the run-time of the near-Clifford simulator of Ref. [33].

Theorem 4 (Hardness of hIQP sampling)**.** *There is no efficient classical algorithm that samples from the output distribution of hyperdepth-2, degree-3 hIQP circuits up to constant total-variation distance error, unless Conjecture 3 is false or the polynomial hierarchy collapses to its third level.*

How are these complexity-theoretic results reflected in the concrete runtime of classical simulation algorithms? To answer this question, we now discuss the runtime scaling of the available classical algorithms.

The most important classes of general-purpose simulation algorithms are tensor-network algorithms. The runtime of these algorithms is governed by the treewidth of the circuits' interaction graph [80]. While the hyperdepth-1 circuits can be simulated in time $O(2^{n/2})$ [27], as discussed in more detail in Ref. [27], we expect the treewidth of hyperdepth-2 circuits to (nearly) saturate n. Tensor-network techniques can also be applied to the degree-D IQP circuit equivalent to the logically implemented hIQP circuit which comprises degree-D and CNOT gates [79, 81, 82]. We show the treewidth of the hypergraphs defined by the effective IQP circuits in Fig. 6(a), and find that it is significantly smaller than n by a constant factor for hyperdepth $\mathfrak{L} = 1$ but nearly saturating the maximal value of $n - 1$ at $\mathfrak{L} \geq 2$.

Another family of simulation algorithms are near-Clifford simulators [83, 84]. Since the number of non-Clifford gates of a random degree-3 hIQP circuit scales as $n \cdot \mathfrak{D} \cdot \mathfrak{L}/3$ we would expect those algorithms to have a large runtime. Nonetheless, IQP circuits appear to be prone to relatively low-rank stabilizer decompositions. Notably, very recently Ref. [33] has demonstrated that in the absence of the in-block permutation gates, there is a decomposition of the output states of hIQP circuits of any depth into $2^{n/3}$ stabilizer states. To this end they exploited the fact that the effective IQP hypergraph has

a vertex cover of size $n/3$. For the hIQP circuits considered here, the minimal vertex cover approaches its maximal size given by $n - D + 1$, however, see Fig. 6(b). We note that one could attempt to classically optimize the minimum vertex cover over CNOT circuits just before the X-basis measurement, since the addition of such a circuit can be simulated in classical post-processing, see Appendix A 2. A closely related approach to that of Ref. [33] has also been explored in Ref. [85] where the authors find low-rank stabilizer decompositions of 2-local classically hard IQP circuits using large independent sets of the IQP graph.

Further specific simulation algorithms for IQP circuits can exploit the fact that one can additively approximate their outcome probabilities in conjunction with sparsity of the output distribution [86–88]. But such algorithms fail due to anticoncentration. This is because anticoncentration implies that the output distribution has exponentially large support and hence the individual probabilities need to be estimated with exponential precision. The best known algorithm for exactly computing output probabilities of IQP circuits scales as $\tilde{\Omega}(2^{0.9965n})$ [89] and exploits a clever way to find roots of degree-3 polynomials. As discussed in Ref. [78] it seems unlikely that this scaling can be improved beyond $2^{n/2}$ for uniform degree-3 circuits. Importantly, the best known algorithm for general IQP circuits thus has a runtime of $\Omega(2^{0.9965n})$ per amplitude.

The results above all apply to the computation of outcome probabilities of IQP circuits. It is not immediately clear that they also imply simulability in terms of sampling. The best available approaches to sampling are, first, the gate-by-gate simulation algorithm of Bravyi *et al.* [90]. This algorithm efficiently turns any method to compute amplitudes of degree-D phase states where some qubits are measured in the X basis and others in the Z basis into a sampling algorithm. For instance, this has been done in Ref. [33] to simulate the experiments in Ref. [27]. Second, one can use rejection sampling if the output distribution is known to be very flat [91]. It is not clear, however, that the output distributions of IQP circuits are sufficiently flat for this approach to work.

Finally, let us note that when comparing to a noisy experiment, the question of classical simulability becomes more subtle. Now, we can also allow the classical algorithm to exploit "the noise", but what exactly this means is up for debate. One could either require a classical algorithm to output samples close to the physically noisy samples. In this scenario, under a simple noise model, any constant amount of local noise makes IQP circuits classically simulatable after constant depth [77, 92]. However, these algorithms are not competitive in practice. Alternatively, one could merely require the samples to pass a test such as the cross-entropy benchmark with a similar score, or "spoof the benchmark" [36, 39]. This approach relates closely to the question of verifying the classical samples, a topic we discuss in detail in the next section. For hIQP circuits Ref. [27, ED Fig. 8] finds that an approach similar to that of Gao *et al.* [36] is not successful. Ultimately, the question is whether quantum error correction is needed for scalable quantum advantage—progress towards this goal has been made in restricted IQP settings [53, 77]. While we do not study noisy simulability of IQP circuits in this work, the question is therefore highly interesting and deserves further study, in particular in the context of encoded circuits.

B. Verification of noisy IQP circuits

We now turn to the question how to verify samples from potentially noisy implementations of random hIQP circuits. We will show that IQP circuits have appealing properties in this respect. Generally speaking, unconditionally verying samples drawn from random quantum circuits which anticoncentrate cannot be done from less than exponentially many samples [73]. This is why sample-efficient benchmarks are typically used to validate the correctness of an experiment [15, 74, 93]. To the extent that a benchmark does not reflect all intricacies of the targeted task, 'spoofing' the experiment—that is, achieving a similar score on the benchmark—can become easier compared to simulating the full distribution. Nevertheless, these benchmarks remain an important tool in the verification of sampling experiments. Most sampling experiments have used the average linear XEB [15, 16, 34]

$$\overline{\chi} \coloneqq \mathbb{E}_C \, 2^n \sum_x q_C(x) p_C(x) - 1, \tag{9}$$

between the ideal and noisy output distributions of a random circuit C denoted by p_C and q_C, respectively, as a benchmark of quantum advantage. The linear XEB has the favourable property that it can be estimated efficiently from few samples by averging the ideal probabilities $p_C(x_i)$ corresponding to samples x_i from a randomly chosen quantum circuits. It has the unfavourable property that computing those ideal probabilities can require exponential time.

The second feature which makes the XEB appealing is that it can be used to measure the many-body average fidelity

$$\overline{F} \coloneqq \mathbb{E}_C \langle C | \rho_C | C \rangle \tag{10}$$

of the pre-measurement state ρ_C compared to the target state $|C\rangle$ for sufficiently scrambling circuits or dynamics [15, 35, 74, 94]. For circuits with Haar-random two-qubit gates and local noise models, this relation holds so long as the single-qubit error rate per gate layer $\epsilon < \gamma/n$ for some constant γ [36, 40]. For error rates $\epsilon > \gamma/n$ on the other hand, the XEB decays much slower than the fidelity, namely with the circuit depth [16, 22, 36, 95]. The slower decay for large noise rates constitutes a vulnerability of XEB that can be exploited when designing classical algorithms using bespokely planted noise at specific locations of the targeted ideal circuit. These algorithms

achieve similar XEB scores [36, 96] as the first experimental demonstration of quantum advantage [15]. In fact, it has been argued by Gao *et al.* [36] that polynomial-time simulations should be able to achieve a comparably high XEB score asymptotically that decays only with the circuit depth.

Intuitively, the deviation between XEB and fidelity for locally Haar-random circuits can be understood in terms of the comparison of error events close to the beginning and the end of the circuit, respectively [36]. Consider a noise model in which single-qubit Pauli errors occur uniformly across the entire circuit volume. Errors at the beginning of the circuit tend to affect XEB and fidelity in the same way. To see this, consider their effect on the initial state which can be gleaned from propagating them backwards into a string with weight determined by the backwards lightcone. Any Pauli error which propagates into a string with a single Pauli-X term will flip a bit of the all-zero input state giving rise to an orthogonal state. This will cause the fidelity as well as the linear XEB to be zero.[97] Pauli errors which propagate into a string with only Z errors do not affect XEB and fidelity. In contrast, an error close to the end of the circuit will affect XEB and fidelity differently. For the XEB, a comparably high fraction of such errors propagate into a low-weight Z-type string at the end of the circuit. Such errors do not affect the outcome probabilities. But they do affect the fidelity: to compute it, we compute the overlap with an ideal state, and therefore an error towards the end of the circuit has a large lightcone. Thus, errors which propagate into a low-weight Z-type string at the end of the circuit will not affect XEB but do affect fidelity, giving rise to the deviation.

Consider in contrast degree-D IQP circuits. Here, all Z-type errors commute with the circuit and therefore yield an orthogonal state. When commuting a single-qubit X error through a degree-k gate, it will turn into an $X \otimes Z^{\otimes(k-1)}$ error. Unless there are cancellations we can therefore commute an X error to the beginning or the end of the circuit, where it will have picked up a Z-string, which yields an orthogonal state. The only deviation between XEB and fidelity should therefore stem from X errors in the very last circuit layer.

In the following, we will rigorously assess this intuition by analyzing the relationship between XEB and fidelity for different circuit ensembles and noise regimes. We find that, indeed, at low noise rates $\lesssim 1/n$ there is a correspondence between XEB and fidelity for general local noise. However, that correspondence breaks down for high noise rates and while the fidelity decays exponentially in the circuit volume, the XEB only decays exponentially in the circuit depth, as is the case for Haar-random circuits [16, 22, 36, 74]. We then show the transition between XEB and fidelity can be shifted to higher noise rates by adding a fixed layer of gates, which does not have an analogue for Haar-random circuits. Altogether, these results give evidence that the intuition discussed above gives the correct mechanism,

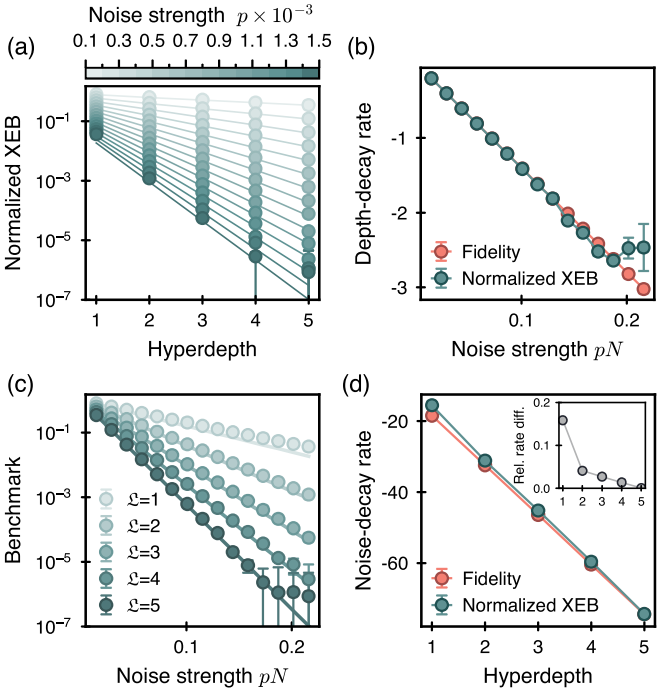

FIG. 7. **XEB versus fidelity in hIQP circuits at low noise rates.** Measures of quality for noisy random degree-2 hIQP circuit ensemble at $\mathfrak{D} = 4$ with circuit-level single-qubit depolarizing noise following a gate application with total error probability p, i.e., each $X/Y/Z$ error occurs with probability $p/3$. Single-qubit and two-qubit fidelities are therefore $1-p$ and $1-2p$, respectively. To obtain the data, we average $\sim 10^4$ circuit realizations with 10^7 noise realizations each and measure the average fidelity $\overline{F}$ and normalized average XEB $\overline{\chi}/\overline{\chi}_{\mathrm{hIQP}}$. (a) Decay of the normalized XEB (circles) and fidelity (lines) with hyperdepth for various noise rate p. (b) Decay rate with hyperdepth α of the normalized XEB (teal) and fidelity (orange) fitted as $\propto \exp(\alpha\mathfrak{L})$, respectively, shown as a function of noise strength pN. (c) Decay of the normalized XEB (circles) and fidelity (lines) with noise strength pN for various hyperdepths. (d) Decay rate with noise strength β of the normalized XEB (teal) and fidelity (orange) fitted as $\propto \exp(\beta \times pN)$, respectively, shown as a function of the hyperdepth $\mathfrak{L}$. The inset shows the relative difference $\beta_{\overline{\chi}}/\beta_{\overline{F}}$ between the decay rates of normalized XEB $\beta_{\overline{\chi}}$ and fidelity $\beta_{\overline{F}}$. The last two values of pN from are ommited from fitting due to their large uncertainty.

while highlighting that error cancellations do in fact play a significant role at high noise rates.

First, we show that in a general setting, at low noise rates, the XEB of IQP circuits can be used as a proxy for the average fidelity. Specifically, we consider general noise occurring in encoded circuits as modelled by quantum instruments and any quantum circuit that is equivalent to an IQP circuit. In this setting, we show that if the noise rate $\epsilon < c/n \log n$ and the circuit depth $\ell \in \Omega(\log n)$, then

$$\overline{\chi} = 2\overline{F} - 2^{-n+1} + O(c). \tag{11}$$

We elaborate the setting and concrete noise model in

Appendix F 1.

We quantitatively observe this behaviour in numerical simulations of degree-2 IQP circuits with circuit-level noise, see Fig. 7. There, we compare the average fidelity with the average noisy XEB normalized by the ideal XEB score $\overline{\chi}/\overline{\chi}_{\mathrm{hIQP}}$. This *normalized XEB* removes some finite-size differences between the average XEB and fidelity at very low noise rates and is therefore a better estimator for the average fidelity (see Refs. [21, 98] for more details). At the same time, the normalization does not change the noisy behaviour qualitatively, i.e., in terms of the decay in terms of the noise and depth shown in Fig. 7(b,d). We find that for sufficiently low noise rates, the (normalized) average XEB and fidelity match extremely well. Together, the analytical and numerical results justify the use of the XEB as a proxy for the average fidelity at sufficiently low (Pauli) noise rates and sufficiently high depth $\Omega(\log n)$.

But how far does the correspondence between XEB and fidelity carry over to the high-noise regime? We find that the relationship between XEB and fidelity breaks down for high local Pauli noise rates and we qualitatively recover its behaviour in Haar-random circuits. Concretely, we consider the sparse degree-2 model with a layer of single-qubit noise following every gate, see Fig. 3(b). We further focus on local X-Y symmetric Pauli noise with X and Y noise rates $p_x = p_y =: q_\perp$ and Z noise rate p_z. Defining $q := q_\perp + 2p_z$ we have the following result.

Theorem 5 (XEB versus fidelity transition)**.** *The average XEB for a local random degree-2 IQP circuits with ℓ gates and ℓ layers of X-Y symmetric Pauli noise is bounded as*

$$\overline{\chi} \geq 2^{1-n}[(1-q_\perp)^\ell + (1-q)^\ell]^n + 2^{-2\ell/n}[(1-q)(1-q_\perp)]^\ell, \quad (12)$$

$$\overline{\chi} \leq 2^{1-n}[1+(1-q)^\ell]^n + p(n)2^{-2\ell/n}, \quad (13)$$

where $p(n) = n^{7/2}/(2\pi)^{3/2}$.

We defer the proof of Theorem 5 to Appendix F 2.

The lower bound on the XEB shows directly that it has two regimes depending on the strength of the noise. When $(p_x + p_y + p_z) < 2\log 2/n^2$, then the XEB is dominated by the exponential decay that is linear in the circuit volume $n\ell$. Whereas for larger noise rates, the XEB decay is dominated by the decay that is linear in the effective depth ℓ/n (that one layer of gates requires $\ell = n/2$ in this model). This sharp change in behavior of the XEB is exactly analogous to what has been found in noisy Haar random circuits [16, 22, 36]. In both cases, it indicates that the XEB ceases to be good proxy for the fidelity, which always decays at a rate linear in the circuit volume until it reaches a value near $1/2^n$. The upper bound of the XEB shows that the two scaling regimes observed in the lower bound are actually tight at low and large enough noise rates.

One way to understand the deviation between the intuition we discuss above and the lower bound for high noise rates is to consider the random gate choices in the sparse IQP ensemble. In the model considered in Theorem 5, for every one of the ℓ randomly chosen qubit pairs a CZ gate is applied with probability 1/2. Think of a circuit comprising ℓ/n parallel gate layers, and consider an X-error occurring in the $(\ell/n - k)$-th layer of the circuit. Then the probability that no gate is applied in the last k layers decays as 2^{-k}, leading to a deviation between XEB and fidelity on that order of magnitude.

This reasoning suggests that the deviation should vanish if we ensure that every X-noise event in the circuit will meet a degree-2 or higher gate when we commute it to the end of the circuit, and that its resulting attached Z-type string is not cancelled by any other gates. As a first step towards this, we can add a fixed parallel layer of CZ gates between all qubit pairs at the very end of an otherwise random degree-2 circuit. This will only leave deviations due to cancellations of Z strings. We find that a fixed gate layer indeed has the effect of lowering the bounds on the average XEB from Theorem 5, an effect that does not have an analogue for Haar-random circuits.

Lemma 6 (Shifting the transition)**.** *The average XEB for the local random degree-2 IQP model on an even number n of qubits with ℓ gates and ℓ layers of X-Y-symmetric Pauli noise with a last layer of parallel CZ gates between all neighbouring qubits is bounded as*

$$\overline{\chi} \geq 2^{1-n}[(1-q_\perp)^\ell + (1-q)^\ell]^n \quad (14)$$

$$+ 2^{-4\ell/n}[(1-q)(1-q_\perp)]^{2\ell}. \quad (15)$$

$$\overline{\chi} \leq 2^{1-n}[1+(1-q)^\ell]^n + p(n)2^{-3\ell/n}. \quad (16)$$

We show Lemma 6 in Appendix F 3. The new upper bound implies that the transition in the XEB has been shifted by a factor of 3/2 to larger circuit depths.

The bounds can be exponentially lowered further by adding additional, non-overlapping fixed gate layers to the circuit. More formally, this expectation can be understood via the proofs of the analytical results in this section. In these proofs, we make use of a mapping of second-moment quantities of degree-2 IQP circuits to a classical statistical model whose states are evolved deterministically by the circuit. In the following section, we will detail this statistical model, and sketch the proofs of our results regarding the ideal and noisy average XEB values.

Before we do so, let us briefly summarize the findings of this section. We have found that—as for Haar-random circuits—the linear XEB is a good estimator of the global state fidelity in the regime of relatively low noise. This is witnessed by our noisy simulations of hIQP circuits shown in Fig. 7 as well as our tight bounds in Theorem 5. That correspondence extends to very general types of noise. As for Haar-random circuits—the correspondence breaks down for large noise rates at which the XEB only decays with the circuit depth while the fidelity

decays with the circuit volume. We have argued, however, that XEB should be a better estimator of fidelity in IQP circuits compared to Haar-random circuits because of the way the noise affects the circuit. This intuition is made rigorous in Lemma 6, which shows that fixed gates in the circuit shift the transition further towards higher noise rates until it eventually vanishes. Altogether, our results shine light on the differences and similarities between random IQP and Haar-random circuits, which we hope will further illuminate the underlying mechanisms.

An open question relevant to our findings is to what extent the noisy XEB score respects typicality or, in other words, how it behaves across different instances of the random circuits. For Haar-random circuits, it is well known that the XEB indeed concentrates rapidly around the mean and hence that typical instances are close to the mean value, since these circuits are higher designs. Second, while we have shown that a transition exists, the precise nature of that transition remains an intriguing field for future studies. This is true in particular when considering the logical XEB and fidelity of encoded circuits subject to physical noise, as we will discuss it in the next section.

Another question left open by our study is to understand the reasons for the transition between hyperdepth $\mathfrak{L} = 1$ and $\mathfrak{L} \geq 2$ circuits. We find this transition in the behaviour of the noisy XEB, as well as in measures of complexity of the underlying IQP graph. This matches the results of Hashizume *et al.* [66] who find the same scrambling transition in a setting of only CZ entangling gates on the hypercube. A more detailed study of a similar scrambling transition in a sparse random model analogous to the random sparse IQP circuits we studied here is given in Ref. [99]. But the exact mechanism and type of transition in the hypercube circuits remains unclear. From a naive interpretation of the statistical-mechanics mapping—elaborated in Section III D—we expect an exponentially decaying deviation with every circuit layer from fully scrambled circuits rather than a sharp phase transition at a given point.

C. Verification of noisy encoded IQP circuits

In the previous section, we showed that the XEB and fidelity in IQP circuits match for local logical noise with low rates. We now assess this correspondence for encoded IQP circuits subject to physical noise. For encoded qubits, there is an ambiguity in defining the logical fidelity in relation to the XEB: The XEB is evaluated directly from measurement outcomes in the X-basis and is not affected by (physical and logical) X errors just before the measurement, but the fidelity can be drastically reduced. This reduction is determined by the physical error density and happens because we are not applying any correction in the Z basis and therefore the state after X measurements and postselection remains outside of the code space. Therefore, a genuine comparison, akin to

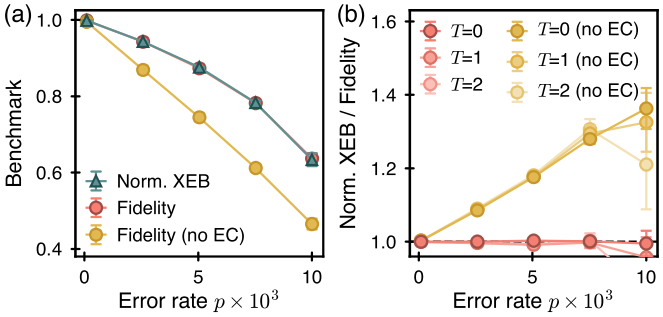

FIG. 8. **Logical XEB and fidelity for an encoded circuit.** We simulate 30 instances of random noisy encoded degree-2 hIQP circuits ($\mathfrak{D} = 4$, $\mathfrak{L} = 2$) with the property that the ideal XEB value $\overline{\chi}_{\text{ideal}} = 3$, taking 10^5 shots per circuit. The circuits are subject to i.i.d. *physical* single-qubit Pauli errors such that the single- and two-qubit gate fidelities are $1 - p/10$ and $1 - p$, respectively. Error bars represent one standard deviation. (a) The average logical fidelity and normalized logical XEB postselected on perfect X-basis stabilizers with and without error correction (EC) performed in the internal state representation. (b) Relative value of the average normalized XEB and fidelity for various X-axis postselection thresholds T.

that from the previous section, necessitates incorporation of an error-correcting procedure in the fidelity definition that leaves a valid code state.

In this procedure, we would typically assume a perfect round of stabilizer measurements and decoding. However, in the sampling protocol we consider, the Z-stabilizers are not measured and a fault-tolerant protocol using the Z-stabilizer information might change the sampling output. Therefore, a meaningful notion of logical fidelity to compare with the XEB should only consider corrections that do not change the logical X basis samples. This is why we define the reference logical fidelity as a state overlap on the level of logical operators after returning the state to the Z-basis code space by performing a virtual (in the simulation code) error correction step that does not modify the XEB score. To achieve this, we use a lookup-table decoder constructed by enumerating low-weight X errors in an $[\![8, 3, 2]\!]$ code.

We simulate random encoded degree-2 hIQP circuits (with the property that the ideal XEB value $\chi_{\text{ideal}} = 3$) at experimentally realistic two-qubit gate noise rates around 99.5% [30]. In Fig. 8 we then compare the logical XEB with the logical fidelity with and without correction of the X errors prior to the X measurement. We observe that for the corrected logical fidelity the correspondence with XEB is tight and consistent with Fig. 7. In contrast, the deviation between XEB and fidelity increases when comparing to the state fidelity without error correction.

These results show that intermediate-size encoded circuits with experimentally relevant noise rates exhibit a close correspondence between logical XEB and an appropriate notion of logical fidelity, i.e., they are in the "healthy regime" of the correspondence between XEB and

fidelity. They also point to some of the inherent subtleties in assessing the performance of encoded circuits where the syndrome qubits necessitate context-dependent definitions of logical process, state preparation, and measurement fidelities.

D. A statistical model for second moments of random IQP circuits

In order to derive the results on the noisy and ideal linear XEB score, we derive in this section a statistical-mechanics mapping for second-moments of degree-2 IQP circuits analogous to those used in the analysis of locally Haar-random circuits [22, 37, 38, 100]. We provide a broad overview of the properties and behavior of our mapping in Fig. 9.

We begin by observing that the ideal and noisy linear XEB score is a second-moment quantity of the IQP circuit. To be precise, we define the *second-moment operator* of a family $\mathcal{I}$ of IQP circuits as the two-copy projector

$$M_2 \coloneqq \mathbb{E}_{C\sim\mathcal{I}} C^{\otimes 2} \left|+^n\right\rangle \left\langle +^n\right|^{\otimes 2} C^{\otimes 2}. \tag{17}$$

We can also allow for noise in one or both copies of the circuit; denote by $\mathcal{N}$ a noise model and by $\rho_C(\mathcal{N})$ the noisy state induced by the application of $C \in \mathcal{I}$ and the noise model. Then the noisy second-moment operator is just given by $M_2(\mathcal{N}, \mathcal{N}') = \mathbb{E}_{C\sim\mathcal{I}} \rho_C(\mathcal{N}) \otimes \rho_C(\mathcal{N}')$. With this definition at hand, we can immediately see that the (noisy) XEB is given by

$$\overline{\chi} = 2^n \sum_x \left\langle x\right|^{\otimes 2} M_2(\mathbb{1}, \mathcal{N}) \left|x\right\rangle^{\otimes 2} - 1, \tag{18}$$

and hence we can derive all its properties from the (noisy) second moment operator. The statistical-mechanics model will describe the evolution of the second-moment operator as a function of the depth of the IQP circuit.

The ideal XEB score of IQP circuits has been computed for uniform degree-2 circuits by Bremner *et al.* [24], while the general moment operator for degree-n circuits has been found by Nechita and Singh [101]. The statistical-mechanics model will allow us to go beyond both settings and analyze arbitrary circuits composed of random degree-2 gates, as well as arbitrary IQP and CNOT gates and (potentially distinct) Pauli noise on each copy. To motivate our approach, we build on the derivation of the ideal XEB score of uniform degree-2 IQP circuits by Bremner *et al.* [24]. Their argument makes use of the fact that the output states of degree-2 IQP circuits can be written as a *polynomial phase state*

$$C\left|+^n\right\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)} \left|x\right\rangle, \tag{19}$$

where $f(x) = \sum_{ij} b_{ij} x_i x_j + \sum_i a_i x_i$ is a degree-2 Boolean polynomial whose coefficients a_i (b_{ij}) are 1 if a degree-1 (degree-2) gate is applied to the qubits labeled by the indices in the circuit C, and 0 otherwise. We can then write the two-copy average over such phase states as

$$M_2 = \frac{1}{2^{2n}} \sum_{z_1,\ldots,z_4} \mathbb{E}_f \left[(-1)^{\sum_i f(z_i)}\right] \left|z_1 z_3\right\rangle \left\langle z_2 z_4\right| \tag{20}$$

$$= \frac{1}{2^{2n}} (\mathbb{1}^{\otimes n} + \mathbb{S}^{\otimes n} + \mathbb{P}^{\otimes n} - 2\mathbb{X}^{\otimes n}), \tag{21}$$

where $\mathbb{S} = \sum_{a,b\in\{0,1\}} \left|ab\right\rangle \left\langle ba\right|$, $\mathbb{P} = \sum_{a,b\in\{0,1\}} \left|aa\right\rangle \left\langle bb\right|$, and $\mathbb{X} = \sum_{a\in\{0,1\}} \left|aa\right\rangle \left\langle aa\right|$. This follows from evaluating conditions under which the averages over the phase $(-1)^{\sum_i f(z_i)}$ does not vanish for any fixed configuration z_1, z_2, z_3, z_4, through the randomness of the coefficients a, b.

The formula (21) is reminiscent of the one for global Haar-random circuits, where we find $M_2 \propto (\mathbb{1}^{\otimes n} + \mathbb{S}^{\otimes n})$, but additional invariances appear in the case of IQP averaging. In fact, one can show that the action of a single-qubit Haar-random gate in the two-copy average is to project onto the space spanned by $\mathbb{1}$ and $\mathbb{S}$. If a quantum circuit contains layers of single-qubit Haar-random gates one can therefore describe the average two-copy evolution as an evolution of these two 'states', which represent the local invariances of the circuit.

The same effect occurs for degree-2 IQP circuits. We find that a good basis for the local state space of degree-2 circuits is given by $\mathcal{S} = \{\mathsf{X}, \mathsf{I}, \mathsf{S}, \mathsf{P}\}$ with

$$\mathsf{X} \coloneqq \left|00\right\rangle \left\langle 00\right| + \left|11\right\rangle \left\langle 11\right|, \tag{22}$$

$$\mathsf{I} \coloneqq \left|01\right\rangle \left\langle 01\right| + \left|10\right\rangle \left\langle 10\right|, \tag{23}$$

$$\mathsf{S} \coloneqq \left|01\right\rangle \left\langle 10\right| + \left|10\right\rangle \left\langle 01\right|, \tag{24}$$

$$\mathsf{P} \coloneqq \left|00\right\rangle \left\langle 11\right| + \left|11\right\rangle \left\langle 00\right|. \tag{25}$$

Intuitively, this is because these states are invariant under swaps between the two copies and contain two $\left|1\right\rangle$ states each so that a Z gate on two copies leaves them invariant. We illustrate the statistical model in Fig. 9 and describe it in detail in Appendix D.

The second-moment operator expressed in the $\mathcal{S}$-basis undergoes simple evolution under random degree-2 gates. Without loss of generality, we can move a layer of random single-qubit Z gates to the beginning of the circuit in all of the models we study. Writing $Z^z = \prod_i Z_i^{z_i}$ we can then write the second moment of the input state

$$\mathbb{E}_{z\in\{0,1\}^n} (Z^z \left|+^n\right\rangle \left\langle +^n\right| Z^z)^{\otimes 2} = \frac{1}{4^n} (\mathsf{I} + \mathsf{S} + \mathsf{P} + \mathsf{X})^{\otimes n}, \tag{26}$$

which can be expressed as a uniform mixture of all states in $\mathcal{S}^{\otimes n}$. If we apply a random degree-2 circuit C to a pair of qubits, we find that a state $P \otimes Q \in \mathcal{S}^{\otimes 2}$ evolves as

$$\mathbb{E}_C C^{\otimes 2} (P \otimes Q) C^{\otimes 2} = \begin{cases} P \otimes Q & \text{if } P = \mathsf{X} \text{ or } Q = \mathsf{X} \text{ or } Q = P \\ 0 & \text{else,} \end{cases} \tag{27}$$

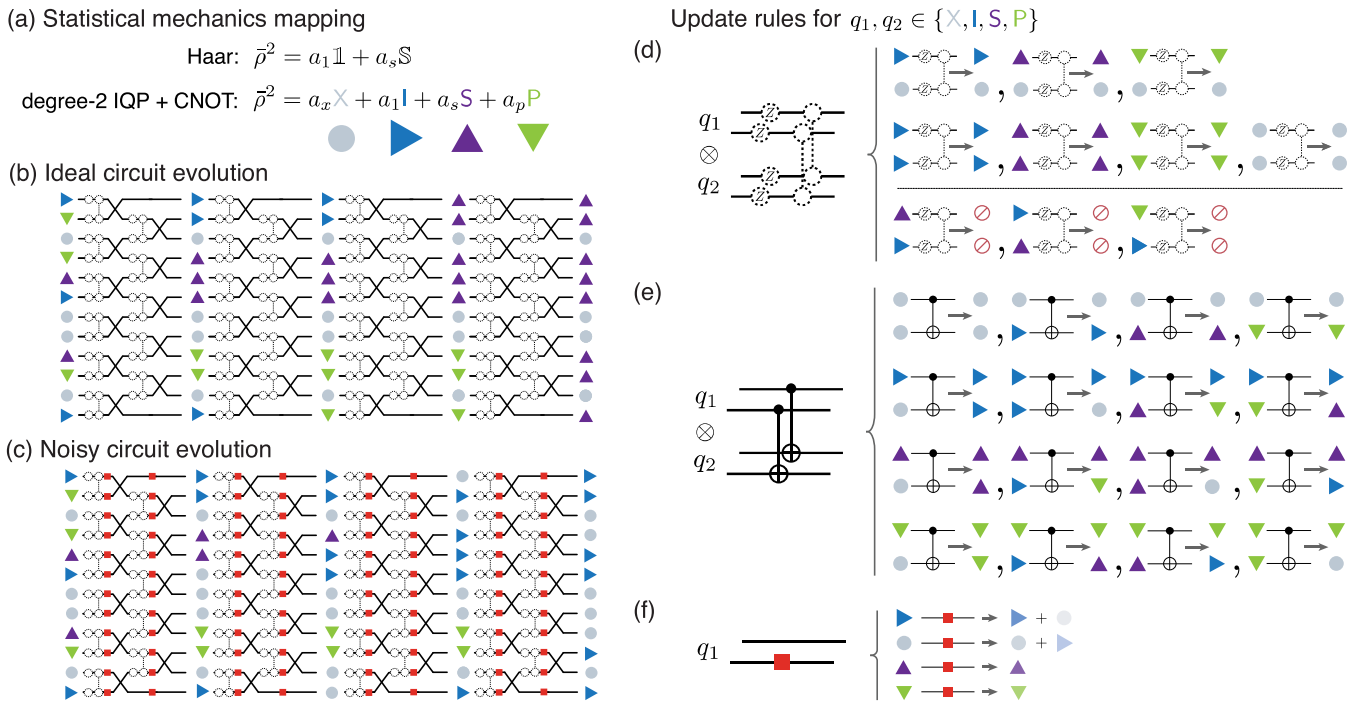

FIG. 9. **Statistical-mechanics mapping of degree-2 second moments.** (a) To analyze the ideal and noisy XEB of degree-2 IQP circuits we map the second moment $\bar{\rho}^2 = \mathbb{E}_C \rho_C \otimes \rho_C$ of IQP + CNOT circuits with random degree-2 circuits on subsets of qubits to the evolution of four classical states $\mathsf{X}, \mathsf{I}, \mathsf{S}, \mathsf{P}$ at each physical site. This mapping is analogous to the mapping used for Haar-random circuits, where the invariant states correspond to the identity and swap operator between the two copies, $\mathbb{1}$ and $\mathbb{S}$, respectively (b) Under a local random circuit in which random degree-2 circuits are applied to random qubits or alternated with layers of SWAP gates the system (depicted here), the system will percolate to states with only one $Q \in \{\mathsf{I}, \mathsf{S}, \mathsf{P}\}$ and X states otherwise, yielding a globally invariant state in $\{Q, X\}^n$ for $Q \in \{\mathsf{I}, \mathsf{S}, \mathsf{P}\}$. The intermediate states shown here represent typical states at the respective circuit layer since under the ideal circuit evolution states only survive or die. (c) Under the noisy circuit evolution, all but the X and I state eventually die out. (d-f) Update rules under random and fixed gates. (d) Subjected to a random IQP circuit a state on two sites either survives (if it contains X or is of the form QQ for $Q \in \{\mathsf{I}, \mathsf{S}, \mathsf{P}\}$) or 'dies' otherwise (crossed out circles). (e) Under a deterministic CNOT gate a state PQ is permuted to a state $P\varepsilon_P(Q)$. (f) To analyze the noisy value of the linear XEB, we need to apply noise to one copy of the circuit. Under X-Y symmetric Pauli noise the amplitude of all states is reduced, while the total amplitude of $\mathsf{X} + \mathsf{I}$ is conserved.

see also Fig. 9(d). In other words, the states of our model can only survive or die under random IQP circuit evolution.

Let us pause here for a moment and consider the evolution of the states. In particular, we would like to understand under which condition the second moments of a sparse random degree-2 circuit converges to its uniform value (21) as a function of depth. It is apparent from Eq. (27) that the X-state plays a special role in the model, and that the model tends towards completely 'polarized states' in which the only appearing states are X and one of $\mathsf{I}, \mathsf{S}, \mathsf{P}$. These states are invariant under the circuit evolution—they are immortal. We can understand the behaviour of the circuit evolution to the completely polarized states $\mathcal{S}_{\text{imm}} = \{\mathsf{X}, \mathsf{I}\}^{\otimes n} \cup \{\mathsf{X}, \mathsf{S}\}^{\otimes n} \cup \{\mathsf{X}, \mathsf{P}\}^{\otimes n}$ in terms of percolation, see Fig. 9(b). As soon as the degree-2 gates form a large connected component of the graph with high probability most states outside of $\mathcal{S}_{\text{imm}}$ will have died so that the second moment approximately recovers the form (21). It is at this point that we expect anticoncentration, since every state contributes exactly 2^{-n} to the overall XEB. Indeed, in Theorem 1 we find a sharp transition as the number of random degree-2 gates exceeds $2.75\, n \log n$ which is precisely the scaling we expect from the percolation argument [102, 103].

Next, we notice that $C^k Z$ gates leave the immortal states in $\mathcal{S}_{\text{imm}}$ invariant. However, they can have a non-trivial effect on states outside of that set that leaves the space spanned by those states. For instance, a CCZ gate maps ISX state to $\mathsf{I} \otimes \mathsf{S} \otimes |00\rangle\langle 00| - \mathsf{I} \otimes \mathsf{S} \otimes |11\rangle\langle 11|$. In our circuit ensembles, we always consider random higher-degree gates in conjunction with uniformly lower-degree gates on the same set of qubits, however. They therefore cannot alter the invariant state. Indeed, even degree-n circuits have the same second moment as degree-2 circuits [101]. Generally, adding additional higher-degree random gates can only speed up anticoncentration.

Furthermore, we find that the action of a CNOT gate

on a string $P \otimes Q$

$$\mathbb{E}_C \mathrm{CNOT}^{\otimes 2}(P \otimes Q)\mathrm{CNOT}^{\otimes 2} = P \otimes \varepsilon_P(Q), \qquad (28)$$

is just a permutation $\varepsilon_P(Q) \in \mathcal{S}$ that depends on the control of the CNOT gate, illustrated in Fig. 9(e); see Lemma 20 in Appendix D 2 for details. In combination with local random degree-D circuits on fixed qubits, CNOT gates thus have a similar effect compared to random degree-D circuits on random subsets of qubits, since they transport 'excitations', i.e., non-X states through the system. We exploit this property when we compute the asymptotic XEB score of block-random IQP and hIQP circuits in Theorem 2.

Finally, consider the effect of noise. Noise is reflected in an asymmetry between the two circuit copies, wherein—in the case of Pauli noise—a Pauli operator is applied somewhere in the circuit on one copy but not the other. We find that Pauli X and Z noise leave the states X, I invariant, and have a nontrivial effect on S, P: X noise permutes $\mathsf{S} \leftrightarrow \mathsf{P}$, while Z noise adds a -1 phase to S and P. If X noise occurs during the circuit evolution, it therefore maps some immortal states to states which will eventually die. Z noise can be moved to the end of the circuit and adds a phase to S and P states. In the XEB, Z noise thus leads to a cancellation of the contributions of states with nontrivial S and P contribution at the end of the circuit. X noise on the other hand leads to a decay of all states outside of the set $\{\mathsf{X}, \mathsf{I}\}^{\otimes n}$. If we therefore consider Pauli noise which is X-Y symmetric, we find a decay of the S, P states and a stochastic permutation of the X, I states, see Fig. 9(f). Under noisy degree-2 circuit evolution, the system thus undergoes a damped percolation to $(\mathsf{X} + \mathsf{I})^n$, while all other states die out, eventually leading to an XEB score of 0, see Fig. 9(c).

In Theorem 5 we compute upper and lower bounds to the noisy XEB of sparse IQP circuits with Pauli noise during the circuit evolution as a function of the circuit depth. To show the theorem, we find a set of long-lived states of the form

$$\{\mathsf{I}, \mathsf{X}\}^{\otimes (n-2)} \otimes \{\mathsf{I}\} \otimes \{\mathsf{S}\}, \qquad (29)$$

i.e., states with a single 'excitation' which slowly decays. It is these states which lead to a deviation between XEB and fidelity for circuits with depth $\ell \in O(n \log n)$.

At the same time, the contribution of those states to the XEB vanishes if we consider a last, fixed layer of CZ gates. To see this, observe that $\mathsf{IS} \xmapsto{\mathrm{CZ}} -\mathsf{IS}$ while $\mathsf{XS} \xmapsto{\mathrm{CZ}} \mathsf{XS}$. Exactly half the states of the form (29) will incur a negative sign at the end of the circuit since a CZ gate is guaranteed to act on the excitation S. In order to lower-bound XEB, we now need to consider states with two excitations, e.g., PPI^{n-2}. This reduces the upper and lower bounds by a factor of 3/2 and 2 in the exponent, respectively.

To summarize, our statistical model can be used to qualitatively understand and quantitatively analyze the behaviour of second-moment properties of IQP circuits. We have done so here for the noisy and ideal values of the XEB score in the hIQP and sparse IQP model, but other architectures can also be considered. An interesting case to study in future work is, for instance, that of geometrically local IQP circuits. A convenient model to study in this case is an IQP + SWAP model wherein degree-2 IQP circuits act on fixed pairs of qubits while the qubits are coupled by SWAP gates acting in a geometrically local fashion. We expect the behaviour in this setting to be quite similar to that of Haar-random circuits, in particular, we expect anticoncentration in logarithmic depth [39, 100].

IV. VALIDATED ADVANTAGE VIA TRANSVERSAL BELL SAMPLING

In the previous section, we have argued that sampling from random degree-D hIQP and sIQP circuits for $D \geq 3$ is classically intractable, and that the linear XEB is a good measure of the output state fidelity in the regime of low noise. Together, these results allow us to perform verified quantum advantage experiments in realistic settings. The drawback of this approach, however, is the fact that estimating the XEB requires evaluating ideal outcome probabilities, which by definition are classically hard to compute in the quantum advantage regime.

This raises the question whether there are more efficient ways to validate the quantum advantage of IQP sampling. In contrast to universal circuits, it is known that the output states of IQP circuits can be verified using the ability to perform (near-)perfect single-qubit measurements for certain families. These include IQP circuits with two-local rotations around arbitrary angles [104, 105] and also general degree-3 circuits which can be thought of as preparing hypergraph states [106, 107]. Such tests require very high-fidelity state preparations, however. Since the stabilizers of IQP circuits can be computed efficiently, it is also possible to perform direct fidelity estimation [108], which is highly sample-efficient, but beyond geometrically local architectures [98] requires potentially long-range entangled measurements whose correctness must be guaranteed. All of these approaches furthermore require different measurements for sampling and verification.

An alternative, unified approach to sampling and verification is to make use of a Bell measurement on two copies of an IQP circuit [41]. For universal random circuits, this gives rise to a sampling task which is classically intractable to simulate while at the same time the output distribution has certain structure that can be exploited for efficient validation of the samples as well as learning properties of the underlying quantum circuit. While this test can be efficiently spoofed by an adversary, it can be used to efficiently validate samples from a noisy quantum device in a 'trusted' laboratory environment.

The output probabilities of a Bell measurement on two

copies of a quantum state $|C\rangle$ are given by

$$P_C(x,z) = \frac{1}{2^n} |\langle C| X_x Z_z |C\rangle|^2, \tag{30}$$

where by $X_x = \prod_{i=1}^n X_i^{x_i}$ and likewise for Z we denote the Pauli X and Z operator on the sites labeled by $x, z \in \{0,1\}^n$, respectively. Since a degree-D IQP circuit C is in the D-th level of the Clifford hierarchy we thus find that the Bell sampling probabilities of degree-D IQP sampling are given by the output probabilities of certain degree-$(D-1)$ IQP circuit C_x that depends on X^x. Concretely, we find

$$P_C(x,z) = \frac{1}{2^n} |\langle +^n| C_x Z_z |+^n\rangle|^2. \tag{31}$$

The gates of C_x derive from the gates in C, observing that for $x \in \{0,1\}^{k+1}$, $\mathrm{C}^k\mathrm{Z} \cdot X_1 \cdot \mathrm{C}^k\mathrm{Z} = X \otimes \mathrm{C}^{k-1}\mathrm{Z}$. In other words, a degree-k gate which meets one or several X operators will reduce to a lower degree gate on the complement sites. All other gates cancel. Thus, the highest-degree gates in C_x are those gates in C which meet a single X operator and are therefore reduced by one degree.

It follows that computing the output probabilities of Bell sampling with degree-3 circuits, which is hard in the standard (X) basis, becomes easy since all amplitudes are just amplitudes of various degree-2 Clifford circuits. However, if we consider Bell sampling from degree-4 circuits, we recover hardness of sampling.

Theorem 7 (Hardness of degree-4 Bell sampling (informal))**.** *Approximately sampling from the output distribution of Bell measurements on uniformly random degree-D IQP circuits for $D \geq 4$ is classically intractable under certain complexity-theoretic assumptions.*

We elaborate the theorem more formally as Theorem 32 in Appendix G.

Let us mention here that in contrast to Bell sampling from universal random circuits [41], the fact that degree-D circuits lie in the D-th level of the Clifford hierarchy introduces significant structure in the output distribution of degree-D Bell sampling. In particular, the output distribution 'shatters' into 2^n sectors determined by the x outcome, and for each such sector, the probabilities are the output probabilities of the smaller IQP circuit C_x. In terms of the statistics of uniform degree-D Bell sampling, we thus find $n+1$ sectors depending on the number $|x|$ of nonzero x outcomes which determine the support of C_x. While within each sector the statistics are the same, they differ between the sectors. This is reminiscent of the situation in Gaussian boson sampling, where the distribution falls into distinct sectors determined by the total photon number [109–111]. Similarly, in boson sampling with a low number of modes the statistics of the outcomes differ depending on the 'collision pattern', that is, the number and size of the collision outcomes [112].

In Appendix G, we detail an adapted hardness proof based on Stockmeyer's counting algorithm [113] for this scenario. A key observation in the proof is that the probability weight is concentrated on outcomes (x,z) for which $|x| = n/2 \pm O(\sqrt{n})$, and therefore hardness of computing outcome probabilities still holds for most outcomes.

For verification of the Bell samples we observe that the degreee-D Bell distribution is supported only on outcomes $(x,z) \in \{0,1\}^{2n}$ for which there is no index i such that $1 = x_i = z_i$. In the terminology of Ref. [41], such indices correspond to Y-outcomes or singlets. Let $\mathrm{supp}_Y(x,z) := \{i \in [n] : x_i = z_i = 1\}$ be the Y-support of (x,z) and $\pi_Y(x,z) = |\mathrm{supp}_Y(x,z)| \mod 2$ be its parity. For universal Bell sampling, the output distribution is supported on outcomes with even Y-support $\pi_Y(x,z) = 0$ [41]. Here, we find $|\mathrm{supp}_Y(x,z) = 0|$ for all outcome strings.

In case the state preparation is noisy, it is discussed in Ref. [41] how and in which scenarios the Bell samples may be used to estimate the average fidelity of random two-copy state preparations. The key element of this verification protocol is to estimate the purity

$$\mathrm{tr}\big[\rho^2\big] \approx \frac{1}{M}\big(|\{(x^i,z^i) : \pi_Y(x^i,z^i) = 0\}| - |\{(x^i,z^i) : \pi_Y(x^i,z^i) = 1\}|\big), \tag{32}$$

from a set of M Bell samples $\{(x^i,z^i)\}_i$. Importantly, this estimate can be computed fully efficiently from the Bell samples. The outcomes with odd Y-parity thus indicate errors in the circuit. In degree-D IQP sampling we have an additional mechanism to detect errors which is given by the even Y-parity outcomes. We leave a detailed study of the possibility to exploit this additional structure in the output distribution in order to diagnose errors in degree-D Bell sampling to future work.

Suffice it to stress here that transversal degree-4 Bell sampling using $[\![2^4, 4, 2]\!]$ color codes provides a viable route towards demonstrating efficiently validated quantum advantage. Efficient validation is possible using the additional structure afforded by the Bell samples. The price at which the efficient validation comes is twofold. First, as in universal Bell sampling, we need to double the number of qubits and perform a Bell measurement across the two circuit copies. This is feasible in the setting considered here since the Bell measurement is transversal in stabilizer codes. Second, there is an additional price to be paid in the specific setting of degree-D Bell sampling: As mentiond above, most outcomes (x,z) have support $|x| = n/2 \pm O(\sqrt{n})$. By Eq. (31) this implies that the naive simulation cost scales exponentially in the worst case as $n/2$ compared to n. On the flipside, that additional price is counterbalanced by doubling the amount of error detection that is possible, since the Bell distribution is guaranteed to have zero weight on Y-outcomes.

To summarize, degree-$(D+1)$ Bell sampling using transversal logical circuits as discussed here is a feasible route towards efficiently validated quantum advantage that requires a linear overhead in the number of

qubits compared to standard degree-D sampling. This overhead is given by a factor of four due to the structure of the Bell distribution times a factor of 2 due to the larger codes. This approach can be compared to other means of efficiently verified quantum advantage using structured states such as cluster states [98, 104], where the quadratic overhead can be thought of as a space-time tradeoff. The space-time-overhead might be traded for intermediate *logical* stabilizer measurements in the context of encoded computations. An alternative route towards error-detected or -corrected quantum advantage could therefore be to use three-dimensional cluster states [55].

V. GENERALIZED CODES FOR SCALABLE COMPUTATION

The $[\![8,3,2]\!]$ code is the smallest realization of the 3D color code on a cube-like region with a logical $\overline{\mathrm{CCZ}}$ gate implemented via a transversal T gate [31].[114] Alternatively, the $[\![8,3,2]\!]$ code can be viewed as the $D=3$ instance of the D-hyperoctahedron code [46] with parameters $[\![2^D, D, 2]\!]$ for any $D \geq 2$. In what follows, we present four constructions that lead to codes with similar transversal implementations of a logical $\overline{\mathrm{CCZ}}$ but an increased code distance $d \geq 4$. Before describing them in more detail, we first provide an overview of each code and their advantages and disadvantages.

In the first code construction, we concatenate a two-qubit Z-repetition code into the $[\![2^D, D, 2]\!]$ code to form a $[\![2^{D+1}, D, 4]\!]$ code. The simplicity of this approach makes it suitable for near-term devices; however, this simplicity comes at the cost of performance because the transversal IQP gates now have to be implemented via two-qubit gates. Furthermore, fault tolerance in this code requires some amount of error detection, so it is not a fully fault-tolerant error correction scheme. We compare the performance of this code to the $[\![8,3,2]\!]$ code, as well as IQP circuits implemented via transversal gates on the $[\![15,1,3]\!]$ code. For a physically reasonable noise model, our conclusion is that the $[\![8,3,2]\!]$ code has the best error detection performance, while the $[\![15,1,3]\!]$ code can be used to reduce the postselection overhead of error detection through error correction at low noise rates.

The next family of codes we consider is a direct generalization of the $[\![2^D, D, 2]\!]$ code to a higher distance color code with parameters $[\![2^{2D}, D, 4]\!]$. The qubit overhead increases substantially compared to the $[\![2^{D+1}, D, 4]\!]$ concatenated codes, but the IQP gates can now all be implemented via transversal single-qubit gates. These are the minimal size color codes with distance ≥ 3 we have found with transversally implemented non-Clifford IQP gates, but the downside of the construction is that scaling to higher distance leads to macroscopic stabilizers that compromise fault-tolerance performance.

These considerations lead us to the third family of color codes we consider, which maintain the LDPC property in their check operators as the code distance is scaled. To achieve this favorable property the qubit overhead (asymptotically) increases by a factor of 5 compared to the previous color code construction with macroscopic stabilizers.

The final family of codes we consider are variants of the three-dimensional toric codes. In the case of the standard 3D toric code with certain boundary conditions, one can implement a logical $\overline{\mathrm{CCZ}}$ gate via transversal physical CCZ gates on each code block. Unfortunately, this is a lower-fidelity physical operation than single-qubit rotations. To overcome this constraint, we propose to concatenate $[\![8,3,2]\!]$ codes with three copies of the 3D toric code so that the local $\overline{\mathrm{CCZ}}$ gates can be implemented via single-qubit gates on the $[\![8,3,2]\!]$ code blocks. This method requires 72 qubits to realize a distance-4 code.

A. The $[\![16,3,4]\!]$ and $[\![15,1,3]\!]$ codes

As a CSS code the $[\![8,3,2]\!]$ code has an unbalanced X/Z-distance of $4/2$. Thus, one way to improve the code distance is to increase the Z distance by replacing each qubit by a two-qubit repetition code with the stabilizer group generated by the two-site XX operator. This $[\![16,3,4]\!]$ code remains a CSS code with transversal CNOT gates and X/Z measurements; however, the in-block $\overline{\mathrm{CZ}}$ and $\overline{\mathrm{CCZ}}$ gates must now be implemented via transversally applied two-qubit $e^{i\theta ZZ}$ gates. The distance of a full transversal circuit thus remains 2, but, depending on the noise model, one may find improved performance of this code.

To test whether such an improvement is possible in our transversal IQP sampling circuits, we compare the performance of the $[\![16,3,4]\!]$ code to the $[\![8,3,2]\!]$ code in Fig. 10. To implement error correction in the $[\![16,3,4]\!]$ code we use a decoder that corrects all syndromes generated by weight-1 errors with the associated weight-1 error and discards samples with syndromes outside this set. We use a noise model with perfect state preparation of the logical code blocks, perfect single-qubit gates, and noisy physical two-qubit gates, i.e., after each two-qubit gate we apply i.i.d. depolarizing noise with rate p to each qubit involved in the gate. Figure 10(a) shows a comparison between the error corrected $[\![16,3,4]\!]$ code and error detection in the deviation of the logical XEB from 1. Both codes have an error rate that scales at low-noise rates as p^2, consistent with expectations; however, we notice that the error detected $[\![8,3,2]\!]$ code outperforms the corrected $[\![16,3,4]\!]$ code and matches the performance of the $[\![16,3,4]\!]$ code with full error detection. Figure 10(b) then shows the performance of the codes in terms of the postselection overhead. In this noise model, the $[\![16,3,4]\!]$ code with error correction still has a larger postselection overhead in our decoder than the $[\![8,3,2]\!]$ code. As a result, when the noise is dominated by two-qubit gate errors, we do not find evidence that the $[\![16,3,4]\!]$ code offers improved performance over the $[\![8,3,2]\!]$ code, despite

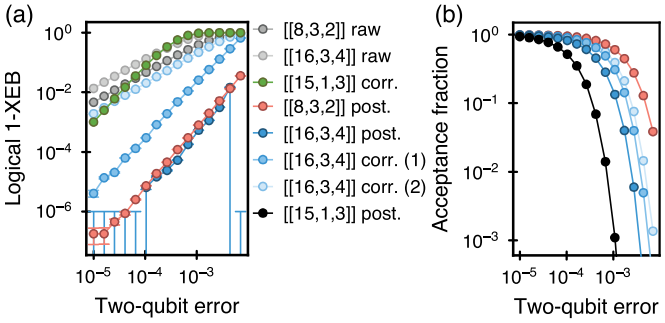

FIG. 10. **Comparison with the $[\![16,3,4]\!]$ and $[\![15,1,3]\!]$ codes.** The same hIQP logical circuit instance ($\mathfrak{D}=4$, $\mathfrak{L}=2$) is executed using the $[\![8,3,2]\!]$, $[\![16,3,4]\!]$, and $[\![15,1,3]\!]$ encodings. The state preparation (encoding circuit) is noiseless, and the transversal part of the circuit is realized with perfect single-qubit gates and two-qubit gates with fidelity $1-p$, implemented as independent single-qubit depolarizing channels on the affected sites. The error bars represent standard deviation of the mean from 10^7 samples and are within markers for most data points. (a) The deviation from 1 of the normalized XEB for various operation modes of the studied encodings. The *raw*, *corr.*, *post.* denote the raw XEB (for a fixed logical operator set), XEB after a correction step, and XEB after postselection, respectively. In the case of the $[\![16,3,4]\!]$, *corr. (1)* corresponds to a decoder that corrects all distance-1 errors and discards all samples with syndromes that it cannot correct while *corr. (2)* implements a sliding-scale postselection protocol that discards a sample if two or more uncorrectable syndromes appear in the entire system. (b) Acceptance fraction for protocols utilizing some degree of postselection.

its improved distance.

To demonstrate an improvement over the $[\![8,3,2]\!]$ code in this noise model, we instead turn to transversal IQP circuits implemented with the punctured Reed-Mueller $[\![15,1,3]\!]$ CSS code that has a logical $\overline{T}$-gate implemented by transversal T-gates. In addition, to transversal T-gates, this code also has transversal logical $\overline{\mathrm{CS}}$ gates implemented via transversal CS gates between code blocks [115]. Thus, we can realize a similar transversal IQP sampling experiment with the $[\![15,1,3]\!]$ code using physical one and two-qubit gates by doing the same CNOT circuit on a hypercube interspersed with block-random $\overline{\mathrm{CS}}$ gates on subsets of $[\![15,1,3]\!]$ code blocks. In this way, we expect to achieve very similar hardness results and circuit families as the degree-3 circuits we studied using the $[\![8,3,2]\!]$ code, see Ref. [25]. In Fig. 10(a), we compare the performance of the error corrected $[\![15,1,3]\!]$ circuit implemented in the same architecture as the $[\![8,3,2]\!]$ code circuit, both with 48 logical qubits. In the comparison both circuits are implemented with only Clifford gates ($\overline{Z}$, $\overline{\mathrm{CZ}}$, $\overline{\mathrm{SWAP}}$ and $\overline{\mathrm{CNOT}}$ gates) to make classical simulations feasible. Although the $[\![15,1,3]\!]$ code removes the postselection overhead, we see that its pseudothreshold behavior for this circuit is below error rates of 10^{-3} and fails to go beyond break-even for the uncorrected $[\![8,3,2]\!]$ circuits until error rates below 10^{-4}. Applying postselection to the $[\![15,1,3]\!]$ circuit data boosts the estimated logical fidelity to 1 for these noise rates and samples, but the postselection overhead is much more severe than the other codes. It is likely that these results could change significantly with repeated rounds of mid-circuit error detection or correction, but that comes with additional qubit overhead. For the present model with error detection/correction implemented only in the final layer of measurements, we find the $[\![8,3,2]\!]$ code to be the highest-performing code.

Finally we remark that there is a closely related code to the $[\![15,1,3]\!]$ code, but with a higher code rate and distance. It is a $[\![64,15,4]\!]$ code where transversal T gates implement a 15-qubit logical $\overline{\mathrm{CCZ}}$ circuit where each logical qubit is involved in 3 distinct $\overline{\mathrm{CCZ}}$ gates [44, Example 6]. Such a code may have promising applications for transversal IQP sampling, but its fault-tolerant properties are not well understood; therefore, we leave the investigation of this and related codes for future work.

B. The 64-qubit code

For any $D \geq 2$, we define the distance-4 D-hyperoctahedron code by taking 2^D D-hypercubes and arranging them to form a D-hypercube; see Fig. 1(g) for an illustration of this construction for $D=2$ and $D=3$. Qubits are placed at the vertices of the resulting lattice, with Z and X stabilizer generators associated with the faces and D-dimensional cells, respectively. The code has parameters $[\![2^{2D}, D, 4]\!]$ and can be interpreted as a D-dimensional color code on a D-hypercube-like region, with D-dimensional cells colored in $D+1$ colors. It has a logical $\overline{\mathrm{C}^{D-1}\mathrm{Z}}$ gate implemented via a transversal $R_D = \mathrm{diag}\left(1, i\pi/2^{D-1}\right)$ gate. In particular, for $D=3$ we obtain the 64-qubit code with a logical $\overline{\mathrm{CCZ}}$ gate.

In order to reduce the weight of stabilizer operators, we can introduce some gauge operators. For the 64-qubit code in Fig. 1(g), whose volumes are colored in four colors, R, G, B and Y, we choose the following gauge group

$$\mathcal{G} = \langle Z_f, X_c, X_{f'} | f \in F, c \in C, f' \in F_{BY} \rangle, \tag{33}$$

where C, F and F_{BY} denote the set of all volumes, faces, and BY faces (shared between B and Y volumes), respectively. The parameters of the resulting code are the same, however we now measure operators of weight at most 8, compared to the original version with one stabilizer of weight 32 and two of weight 16 that were associated with Y and B volumes, respectively.

We remark that, in an analogous way, for any even d we can obtain the distance-d D-hyperoctahedron code by taking an arrangement of D-hypercubes forming a $(d/2) \times \ldots \times (d/2)$ region of the D-hypercubic lattice. The parameters of the code are $[\![d^D, D, d]\!]$. Unfortunately, some of its stabilizers are macroscopic. In particular, for $D=3$ there are $d/2-1$ stabilizers of weight $2d^2$, $d(d-2)/4$ stabilizers of weight $4d$, $(d+4)(d-2)/2$ stabilizers of weight $2d$; weight of other stabilizers is at

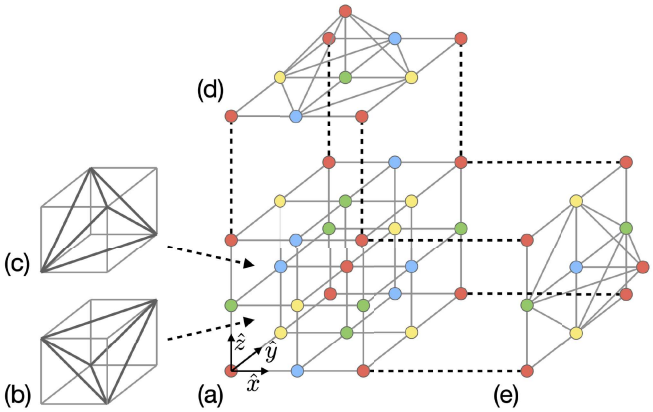

FIG. 11. **Large-distance codes with logical CCZ via transversal single-qubit rotations.** Construction of a lattice $\mathcal{L}$ that facilitates a 3D color code with distance $d = L+2$ and a logical $\overline{\mathrm{CCZ}}$ gate for $L = 2$. Note that we use the dual-lattice picture, where qubits are identified with tetrahedra. (a) A cubic lattice $\mathcal{L}'$ of size $L \times L \times L$, with its vertices colored in four colors, R, G, B and Y. (b)(c) Two possible ways of splitting each cube of $\mathcal{L}'$ into five tetrahedra. We obtain $\mathcal{L}$ from $\mathcal{L}'$ by adding additional tetrahedra along the boundaries of $\mathcal{L}'$. To avoi d clutter, we only illustrate (d) the top and (e) right boundaries.

most 8. From the perspective of fault tolerance we may require that all stabilizers are of constant weight. This motivates a definition of a family of 3D color codes, which we present in the following subsection.

C. A family of 3D color codes

In order to define a 3D color code with even distance $d \geq 4$ and a logical $\overline{\mathrm{CCZ}}$ gate, we need to be able to construct an appropriate lattice supporting this code. Since describing our construction seems easier in the the dual-lattice picture, we will use it in this subsection.

We start with a cubic lattice $\mathcal{L}'$ of size $L \times L \times L$, where $L = d-2$, which will constitute the bulk of the color code lattice; see Fig. 11(a). Each cube of $\mathcal{L}'$ is then split into five tetrahedra in one of two ways [116]; see Fig. 11(b)(c). We require that any two neighboring cubes that share a face are split differently. The vertices of $\mathcal{L}'$ are colored in one of four colors, R, G, B or Y. Then, we obtain $\mathcal{L}$ by adding additional tetrahedra to $\mathcal{L}'$, such that every vertex of color G (respectively, B and Y) on the top or bottom (respectively, left or right, and front of rear) boundary of $\mathcal{L}'$ is covered; see Fig. 11(d)(e). Let $\partial\mathcal{L}_2$ denote the set of triangular faces on the boundary of $\mathcal{L}$. We define $\partial\mathcal{L}_1$ and $\partial\mathcal{L}_0$ to be the sets of outermost edges and vertices bounding the initial lattice $\mathcal{L}'$, respectively. By construction, $|\mathcal{L}| = 5L^3 + 6L^2$, $|\partial\mathcal{L}_2| = 12L^2$, $|\partial\mathcal{L}_2| = 12L^2$, $|\partial\mathcal{L}_1| = 12L$ and $|\partial\mathcal{L}_0| = 8$.

We define a 3D color code associated with the lattice $\mathcal{L}$ as follows. We place one qubit on every tetrahedron of $\mathcal{L}$, triangular face in $\partial\mathcal{L}_2$, edge in $\partial\mathcal{L}_1$ and vertex in $\partial\mathcal{L}_0$, and associate X and Z stabilizers with vertices and edges of $\mathcal{L}$, respectively. This construction is reminiscent of how a 3D tetrahedral color code is defined [117, 118]. The resulting code has parameters $[\![5d^3 - 12d^2 + 16, 3, d]\!]$. Note that for $d = 4$ we require 144 qubits, which is 2.25× more than for the 64-qubit code. While this construction is less qubit efficient (roughly by 5× for large d) than the one from the previous subsection, asymptotically it only has stabilizers of constant weight. By introducing gauge qubits we can further reduce the weight of the largest stabilizer, which is 32, requiring to measure operators of weight at most 6. For more details on the lattice, see Appendix H.

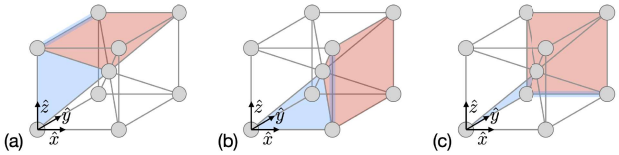

FIG. 12. **3D toric/color codes with transversal CCZ.** Three copies of the 3D toric code with distance two, which admit a transversal logical $\overline{\mathrm{CCZ}}$ gate. To avoid clutter, we only depict some X (red) and Z (blue) stabilizers. The copy in (a) has two X stabilizers (pyramids spanned by the top/bottom faces and the cube center) and eight Z stabilizers (four edges parallel to $\hat{y}$ and four triangular faces spanned by edges parallel to z and the cube center). The copies in (b) and (c) are obtained by rotating the copy in (a).

D. A family of 3D toric/color codes

To construct a family of codes with growing code distance and constant-weight stabilizers we can take three copies of the 3D toric code with open boundary conditions, each encoding one logical qubit and rotated with respect to each other (via a cyclic permutation of the x, y and z axes). The three logical qubits admit a logical $\overline{\mathrm{CCZ}}$ implemented via a transversal CCZ gate [31, 119]. By encoding each triple of qubits supporting a physical CCZ gate into the $[\![8, 3, 2]\!]$ code, we obtain a 3D color code with a logical $\overline{\mathrm{CCZ}}$ implemented via a transversal T gate. The resulting code has parameters $[\![8n, 3, 2d]\!]$, where n and d are the number of qubits and code distance of each copy of the toric code, and typically $n = O(d^3)$.

The smallest realization of this construction uses 72 qubits and has distance four. It is based on the 3D toric code with distance two that we depict in Fig. 12. Analogously, by considering D copies of the D-dimensional toric code with distance two concatenated with the D-hyperocatedron code we obtain the code with parameters $[\![2^D(2^D+1), D, 4]\!]$ and a logical $\overline{\mathrm{C}^{D-1}\mathrm{Z}}$ gate implemented via a transversal $R_D = \mathrm{diag}\left(1, i\pi/2^{D-1}\right)$ gate.

E. Threshold Theorem for Transversal IQP Sampling

An important property of the previous two code constructions is that they are asymptotically low density parity check (LDPC) codes. As shown by Gottesman [26], any quantum LDPC code whose distance diverges at least logarithmically with increasing system size will have a fault-tolerance threshold under a local stochastic noise model. The proof includes the case of transversal gates applied to the code. More precisely under a minimum weight decoder, the theorem implies that for an $[\![n, k, d]\!]$ LDPC code with p below the threshold value p_{th} the logical error rate scales as

$$O[n(p/p_{\text{th}})^{d/2}] \tag{34}$$

as $n \to \infty$. In particular, for the two families of LDPC codes introduced in the previous section, $d = O(n^{1/3})$, which leads to an exponential suppression of the error rate with increasing system size. As a result, the output distribution of random transversal IQP circuits implemented with these codes will converge exponentially close in total variation distance to the ideal noiseless distribution.

A similar threshold behavior in IQP circuit sampling was also studied recently in Ref. [52]. There, the authors consider a different circuit architecture that uses ancilla qubits to compress the IQP circuit to a constant depth. They also consider a different family of 3D color codes with logical $\overline{T}$-gates instead of logical $\overline{\text{CCZ}}$ gates implemented via transversal T-gates.

VI. OUTLOOK

In this work, we have developed and analyzed a hardware-efficient scheme for demonstrating quantum advantage using logical qubits as its fundamental building blocks. At the heart of our scheme is the $[\![2^D, D, 2]\!]$ hypercubic color code, and its transversally realized non-Clifford C^kZ and CNOT gates. Using the transversal gate set of multiple $[\![2^D, D, 2]\!]$ code blocks and physical qubit permutations, we devise a family of degree-D IQP circuits on a hypercube connectivity and show that the family has all desirable features for a demonstration of quantum advantage. At the same time, the hIQP circuit family can be fully parallelized in reconfigurable atom architectures, and a variant thereof has already been demonstrated in recent experiments [27]. Error detection can be performed at the level of the classical samples, allowing to significantly improve their quality as measured by the linear XEB. For efficient validation of hIQP samples, we propose to sample from two copies of a degree-4 hIQP circuit in the transversal Bell basis. Finally, we generalize the $[\![2^D, D, 2]\!]$ code family to arbitrary distances, charting a path towards scalable demonstrations of quantum advantage using repeated error correction.

Our work studies computational complexity in the context of early fault-tolerant quantum computing architectures, using the $[\![8, 3, 2]\!]$ code as its model system. It highlights differences to non-fault-tolerant architectures, where hardware-efficiency obeys quite different constraints [15, 34], and noise behaves in very different ways. The study of noise becomes crucial to assessing the prospects of quantum advantage in a twofold way—in terms of success on the chosen benchmark (the linear XEB in our case) but also in terms of the error reduction capabilities of a chosen code and error correction/detection strategy. For the former, in our study, we analyze in detail the noisy linear XEB of degree-D IQP circuits using numerical simulations and a bespoke statistical-mechanics mapping. For the latter, we compare the performance of error correction in the $[\![16, 3, 4]\!]$ and $[\![15, 1, 3]\!]$ codes and error detection at the end of the circuit using the $[\![8, 3, 2]\!]$ code. Our study shows that, just like for unencoded quantum computations, random quantum circuits remain an important setting for understanding the capabilities of a given architecture in terms of its noise resilience and potential for quantum speedups.

There are a range of interesting directions that arise from the new experimental capabilities of the reconfigurable atom platform in general, and our study of its transversal degree-D IQP circuits in particular. An immediate technical open questions raised by our study of noise in degree-D IQP circuits is what type of transition the XEB undergoes as a function of the noise strength—whether it is a sharp phase transition as for Haar random circuits, or a different type of transition. While we have found that the same two regimes do exist for noisy IQP circuits, we have also presented evidence that the XEB behaves more benignly for random IQP circuits compared to Haar-random circuits. Furthermore, it is an intriguing question how the transition behaves for logical XEB and fidelity in the presence of physical noise in an encoded circuit.

An important question raised by the advent of encoded quantum computations in experiments is how to best make use of limited degree of error detection and correction. In this work, we have focused on a minimal amount of error detection at the very end of the computation, allowing us to eliminate certain errors. Next steps include performing multiple rounds of error detection throughout the circuit execution via stabilizer measurements without feed-forward. In addition, it is important to explore the most efficient ways of using a limited number of measure-and-correct rounds. The availability of the additional information gleaned from mid-circuit stabilizer measurements brings with it the question of how to best make use of that data. While one can perform postselection to improve the overall sample quality, in addition to that, the stabilizer data can yield diagnostic data about circuit performance and even error mechanisms [120–122]. In particular, it has been shown that stabilizer data can be used to learn correctable Pauli errors [123–125] since the stabilizer measurement lets one

measure the Fourier transform of the Pauli noise distribution [126]. Stabilizer data might also be used to infer the average fidelity of transversal gate layers preceding a stabilizer measurement, analogous to cycle benchmarking which includes noise tailoring [127, 128]. This is particularly intriguing in the context of verifying quantum advantage demonstrations, since it may allow us to efficiently estimate the average fidelity of the output state, bypassing the need to resort to the inefficient XEB, or verification schemes that come with a significant resource overhead [41, 98]. Random quantum circuits constitute a well-understood model in which the relation between different benchmarks can be studied. As our work shows, they also provide an ideal playground for questions regarding the nature of early fault tolerance in general.

Beyond sampling circuits, our work can be understood as a case study of circuit compilation in co-design with code and hardware constraints. Optimized towards fast scrambling and simulation complexity, the family of hIQP circuits is highly resource efficient in terms of the chosen code, since it only uses transversal and permutation operations, and in terms of the reconfigurable atom hardware, which allows for parallel long-range gates. One can ask how to compile algorithms with different use cases to the same or similar code and hardware requirements. At the same time, it is interesting to explore which algorithms or sub-routines are most suited towards mostly transversal implementations in high-dimensional color codes with non-Clifford transversal gate sets. For example, it might be possible to encode arithmetic operations using only phase states [129]. From a different perspective, we can think of the transversally prepared states as a resource for more complex algorithms, since these states can be prepared with comparably low errors. In the near term, one can also envision logical-physical quantum circuits. In this scenario high-fidelity transversally prepared states could supply resources that are expensive for physical circuit implementations such as non-local magic and entanglement. These resources could then be exploited by a low-depth universal, physical circuit. Although not a viable path in the long-run, such an approach might allow novel applications on currently available hardware. We hope that the present work will inspire the community to address these and other interesting questions regarding the physics and computational properties of early fault-tolerant quantum computation.

ACKNOWLEDGMENTS

We thank Abhinav Deshpande, Bill Fefferman, Daniel Grier, Jinguo Liu, Brayden Ware, Sepehr Ebadi, Simon Evered, Alexandra Geim, Sophie Li, Tom Manovitz, and Hengyun Zhou for helpful discussions. DH gratefully acknowledges the hospitality of the Simons Institute for the Theory of Computing during Summer 2023 and Spring 2024, where part of this work was conducted. DH acknowledges funding from the US DoD through a QuICS Hartree fellowship. This research was supported in part by NSF QLCI grant OMA-2120757. We acknowledge financial support from IARPA and the Army Research Office, under the Entangled Logical Qubits program (Cooperative Agreement Number W911NF-23-2-0219), the DARPA ONISQ program (grant number W911NF2010021), the DARPA IMPAQT program (grant number HR0011-23-3-0012), the Center for Ultracold Atoms (a NSF Physics Frontiers Center, PHY-1734011), the National Science Foundation (grant number PHY-2012023 and grant number CCF-2313084), the Army Research Office MURI (grant number W911NF-20-1-0082), and QuEra Computing. XG acknowledges support from U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator, NSF PFC grant No. PHYS 2317149 and start-up grants from CU Boulder.

[1] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, PNAS **114**, 7555 (2017).
[2] C. Gidney and M. Ekerå, Quantum **5**, 433 (2021).
[3] L. Clinton, J. Bausch, and T. Cubitt, Nat Commun **12**, 4989 (2021).
[4] D. Litinski, (2023), arXiv:2306.08585.
[5] J. D. Watson, J. Bringewatt, A. F. Shaw, A. M. Childs, A. V. Gorshkov, and Z. Davoudi, (2023), arXiv:2312.05344.
[6] P. Shor, in *Proceedings of 37th Conference on Foundations of Computer Science* (1996) pp. 56–65.
[7] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996).
[8] D. Aharonov and M. Ben-Or, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '97 (Association for Computing Machinery, New York, NY, USA, 1997) pp. 176–188.
[9] When restricting to transversal operations, it can be proven that the implemented gate set in fact has to be discrete [50].
[10] E. Knill, (2004), arXiv:quant-ph/0402171.
[11] E. Knill, (2004), arXiv:quant-ph/0404104.
[12] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).
[13] A. Paetznick and B. W. Reichardt, Phys. Rev. Lett. **111**, 090505 (2013).
[14] J. T. Anderson, G. Duclos-Cianci, and D. Poulin, Phys. Rev. Lett. **113**, 080501 (2014).
[15] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, and others, Nature **574**, 505 (2019).
[16] A. Morvan, B. Villalonga, X. Mi, S. Mandrà, A. Bengtsson, P. V. Klimov, Z. Chen, S. Hong, C. Erickson, I. K. Drozdov, and others, Phase transition in random circuit sampling (2023), arXiv:2304.11119.
[17] M. J. Bremner, R. Jozsa, and D. J. Shepherd, Proceed-

ings of the Royal Society A: Mathematical, Physical and Engineering Sciences **467**, 459 (2010).
[18] S. Aaronson and A. Arkhipov, Th. Comp. **9**, 143 (2013).
[19] D. Hangleiter and J. Eisert, Rev. Mod. Phys. **95**, 035001 (2023).
[20] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Nature Phys **14**, 595 (2018).
[21] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, R. Finkelstein, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale, H. Pichler, F. G. S. L. Brandão, S. Choi, and M. Endres, Nature **613**, 468 (2023).
[22] B. Ware, A. Deshpande, D. Hangleiter, P. Niroula, B. Fefferman, A. V. Gorshkov, and M. J. Gullans, arXiv:2305.04954 (2023).
[23] D. Shepherd and M. J. Bremner, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **465**, 1413 (2009).
[24] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Phys. Rev. Lett. **117**, 080501 (2016).
[25] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Quantum **1**, 8 (2017).
[26] D. Gottesman, arXiv:1310.2984.
[27] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. Bonilla Ataides, N. Maskara, I. Cong, X. Gao, P. Sales Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić, and M. D. Lukin, Nature **626**, 58 (2024).
[28] J. Beugnon, C. Tuchendler, H. Marion, A. Gaëtan, Y. Miroshnychenko, Y. R. P. Sortais, A. M. Lance, M. P. A. Jones, G. Messin, A. Browaeys, and P. Grangier, Nature Phys **3**, 696 (2007).
[29] M. Schlosser, S. Tichelmann, J. Kruse, and G. Birkl, Quantum Inf Process **10**, 907 (2011).
[30] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara, H. Levine, G. Semeghini, M. Greiner, V. Vuletić, and M. D. Lukin, Nature **622**, 268 (2023).
[31] A. Kubica, B. Yoshida, and F. Pastawski, New J. Phys. **17**, 083026 (2015).
[32] D. Bluvstein, H. Levine, G. Semeghini, T. T. Wang, S. Ebadi, M. Kalinowski, A. Keesling, N. Maskara, H. Pichler, M. Greiner, V. Vuletić, and M. D. Lukin, Nature **604**, 451 (2022).
[33] D. Maslov, S. Bravyi, F. Tripier, A. Maksymov, and J. Latone, (2024), arXiv:2402.03211.
[34] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, and others, Science Bulletin **67**, 240 (2022).
[35] D. K. Mark, J. Choi, A. L. Shaw, M. Endres, and S. Choi, Phys. Rev. Lett. **131**, 110601 (2023).
[36] X. Gao, M. Kalinowski, C.-N. Chou, M. D. Lukin, B. Barak, and S. Choi, PRX Quantum **5**, 010334 (2024).
[37] T. Zhou and A. Nahum, Phys. Rev. B **99**, 174205 (2019).
[38] N. Hunter-Jones, (2019), arXiv:1905.12053.
[39] B. Barak, C.-N. Chou, and X. Gao, in *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 185, edited by J. R. Lee (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2021) pp. 30:1–30:20, arXiv:2005.02421.
[40] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Commun. Math. Phys. **405**, 78 (2024).
[41] D. Hangleiter and M. J. Gullans, (2023), arXiv:2306.00083.
[42] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, Nature **627**, 778 (2024).
[43] Q. Xu, J. P. B. Ataides, C. A. Pattison, N. Raveendran, D. Bluvstein, J. Wurtz, B. Vasic, M. D. Lukin, L. Jiang, and H. Zhou, (2023), arXiv:2308.08648.
[44] N. Rengaswamy, R. Calderbank, M. Newman, and H. D. Pfister, IEEE JSAIT **1**, 499 (2020), arXiv:1910.09333.
[45] Y. Wang, S. Simsek, T. M. Gatterman, J. A. Gerber, K. Gilmore, D. Gresh, N. Hewitt, C. V. Horst, M. Matheny, T. Mengle, B. Neyenhuis, and B. Criger, Fault-tolerant one-bit addition with the smallest interesting colour code (2023), arXiv:2309.09893.
[46] M. Vasmer and A. Kubica, PRX Quantum **3**, 030319 (2022).
[47] C. Sheng, X. He, P. Xu, R. Guo, K. Wang, Z. Xiong, M. Liu, J. Wang, and M. Zhan, Phys. Rev. Lett. **121**, 240501 (2018).
[48] H. Levine, D. Bluvstein, A. Keesling, T. T. Wang, S. Ebadi, G. Semeghini, A. Omran, M. Greiner, V. Vuletić, and M. D. Lukin, Phys. Rev. A **105**, 032618 (2022).
[49] E. Campbell, The smallest interesting colour code (2016), accessed: 2023-01-29.
[50] B. Eastin and E. Knill, Phys. Rev. Lett. **102**, 110502 (2009).
[51] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).
[52] L. Paletta, A. Leverrier, A. Sarlette, M. Mirrahimi, and C. Vuillot, (2023), arXiv:2307.10729.
[53] S. Bravyi, D. Gosset, R. König, and M. Tomamichel, Nat. Phys. **16**, 1040 (2020).
[54] R. Mezher, J. Ghalbouni, J. Dgheim, and D. Markham, Phys. Rev. Research **2**, 033444 (2020).
[55] R. Raussendorf, J. Harrington, and K. Goyal, Annals of Physics **321**, 2242 (2006).
[56] T. Jochym-O'Connor, A. Kubica, and T. J. Yoder, Phys. Rev. X **8**, 021047 (2018).
[57] M. Newman and Y. Shi, Quant. Inf. Comp. **18**, 0927 (2018), arXiv:1704.07798.
[58] S. Beigi and P. W. Shor, Quant. Inf. Comp. **10**, 0041 (2010).
[59] B. Zeng, X. Chen, and I. L. Chuang, Phys. Rev. A **77**, 042313 (2008).
[60] B. Eastin and E. Knill, Phys. Rev. Lett. **102**, 110502 (2009).
[61] J. T. Anderson, (2022), arXiv:2212.05398.
[62] S. X. Cui, D. Gottesman, and A. Krishna, Phys. Rev. A **95**, 012329 (2017).
[63] E. Kubischta and I. Teixeira, Phys. Rev. Lett. **131**, 240601 (2023).
[64] E. Kubischta and I. Teixeira, (2023), arXiv:2310.17652.
[65] D. N. Page, Phys. Rev. Lett. **71**, 1291 (1993).
[66] T. Hashizume, G. S. Bentsen, S. Weber, and A. J. Daley, Phys. Rev. Lett. **126**, 200603 (2021).
[67] M. Rossi, M. Huber, D. Bruß, and C. Macchiavello, New J. Phys. **15**, 113022 (2013).
[68] O. Gühne, M. Cuquet, F. E. S. Steinhoff, T. Moroder, M. Rossi, D. Bruß, B. Kraus, and C. Macchiavello, J.

Phys. A: Math. Theor. **47**, 335303 (2014).
[69] D. W. Lyons, D. J. Upchurch, S. N. Walck, and C. D. Yetter, J. Phys. A: Math. Theor. **48**, 095301 (2015).
[70] J. Miller and A. Miyake, npj Quantum Inf **2**, 1 (2016).
[71] Y. Takeuchi, T. Morimae, and M. Hayashi, Sci Rep **9**, 13585 (2019).
[72] M. Gachechiladze, O. Gühne, and A. Miyake, Phys. Rev. A **99**, 052304 (2019).
[73] D. Hangleiter, M. Kliesch, J. Eisert, and C. Gogolin, Phys. Rev. Lett. **122**, 210502 (2019).
[74] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Nat. Phys. **14**, 595 (2018).
[75] S. Aaronson and L. Chen, in *32nd Computational Complexity Conference (CCC 2017)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 79, edited by R. O'Donnell (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2017) pp. 22:1–22:67.
[76] S. Aaronson and S. Gunn, (2020), arXiv:1910.12085.
[77] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Quantum **1**, 8 (2017).
[78] A. M. Dalzell, A. W. Harrow, D. E. Koh, and R. L. L. Placa, Quantum **4**, 264 (2020).
[79] J.-G. Liu, L. Wang, and P. Zhang, Phys. Rev. Lett. **126**, 090506 (2021).
[80] I. L. Markov and Y. Shi, SIAM Journal on Computing **38**, 963 (2008).
[81] J.-G. Liu, X. Gao, M. Cain, M. D. Lukin, and S.-T. Wang, SIAM J. Sci. Comput. **45**, A1239 (2023).
[82] GenericTensorNetworks.jl documentation, accessed 2023-09-27.
[83] S. Bravyi and D. Gosset, Phys. Rev. Lett. **116**, 250501 (2016).
[84] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, Quantum **3**, 181 (2019).
[85] J. Codsi and J. van de Wetering, (2023), arXiv:2212.08609.
[86] D. Shepherd, (2010), arXiv:1005.1744.
[87] M. Schwarz and M. V. den Nest, arXiv:1310.6749 (2013).
[88] H. Pashayan, S. D. Bartlett, and D. Gross, Quantum **4**, 223 (2020).
[89] D. Lokshtanov, R. Paturi, S. Tamaki, R. Williams, and H. Yu, in *Proceedings of the 2017 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Proceedings (Society for Industrial and Applied Mathematics, 2017) pp. 2190–2202.
[90] S. Bravyi, D. Gosset, and Y. Liu, Phys. Rev. Lett. **128**, 220503 (2022).
[91] I. L. Markov, A. Fatima, S. V. Isakov, and S. Boixo, (2018), arXiv:1807.10749.
[92] J. Rajakumar, J. D. Watson, and Y.-K. Liu, (2024), arXiv:2403.14607.
[93] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, Nature Phys **15**, 159 (2019).
[94] A. L. Shaw, Z. Chen, J. Choi, D. K. Mark, P. Scholl, R. Finkelstein, A. Elben, S. Choi, and M. Endres, Nature **628**, 71 (2024).
[95] A. Deshpande, P. Niroula, O. Shtanko, A. V. Gorshkov, B. Fefferman, and M. J. Gullans, PRX Quantum **3**, 040329 (2022).
[96] F. Pan, K. Chen, and P. Zhang, Phys. Rev. Lett. **129**, 090502 (2022).
[97] To see the latter observe that $2^{2n}\mathbb{E}_C p_C(x)p_C(y) - 1 = 0$ for $x \neq y$.
[98] M. Ringbauer, M. Hinsche, T. Feldker, P. K. Faehrmann, J. Bermejo-Vega, C. Edmunds, L. Postler, R. Stricker, C. D. Marciniak, M. Meth, I. Pogorelov, R. Blatt, P. Schindler, J. Eisert, T. Monz, and D. Hangleiter, (2023), arXiv:2307.14424.
[99] S. Kuriyattil, T. Hashizume, G. Bentsen, and A. J. Daley, PRX Quantum **4**, 030325 (2023).
[100] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, PRX Quantum **3**, 010333 (2022).
[101] I. Nechita and S. Singh, Linear Algebra and its Applications **613**, 46 (2021).
[102] S. R. Broadbent and J. M. Hammersley, Mathematical Proceedings of the Cambridge Philosophical Society **53**, 629 (1957).
[103] B. Bollobás and O. Riordan, *Percolation* (Cambridge University Press, Cambridge, 2006).
[104] J. Bermejo-Vega, D. Hangleiter, M. Schwarz, R. Raussendorf, and J. Eisert, Phys. Rev. X **8**, 021010 (2018).
[105] M. Hayashi and Y. Takeuchi, New J. Phys. **21**, 093060 (2019).
[106] T. Morimae, Y. Takeuchi, and M. Hayashi, Phys. Rev. A **96**, 062321 (2017).
[107] H. Zhu and M. Hayashi, Phys. Rev. Appl. **12**, 054047 (2019).
[108] S. T. Flammia and Y.-K. Liu, Physical Review Letters **106**, 230501 (2011).
[109] R. Kruse, C. S. Hamilton, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Phys. Rev. A **100**, 032326 (2019).
[110] A. Deshpande, A. Mehta, T. Vincent, N. Quesada, M. Hinsche, M. Ioannou, L. Madsen, J. Lavoie, H. Qi, J. Eisert, D. Hangleiter, B. Fefferman, and I. Dhand, Science Advances **8**, eabi7894 (2022).
[111] A. Ehrenberg, J. T. Iosue, A. Deshpande, D. Hangleiter, and A. V. Gorshkov, (2024), arXiv:2312.08433.
[112] A. Bouland, D. Brod, I. Datta, B. Fefferman, D. Grier, F. Hernandez, and M. Oszmaniec, (2023), arXiv:2312.00286.
[113] L. Stockmeyer, in *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing* (ACM, 1983) pp. 118–126.
[114] For the discussion in this section, we will emphasize the fact that a gate G is a logical gate using the standard notation $\overline{G}$ in contrast to physically implemented gates.
[115] This follows from the fact that CNOT is transversal for CSS codes since $\mathrm{CS} = \mathrm{CNOT}(\mathbb{1} \otimes T^\dagger)\mathrm{CNOT}(\mathbb{1} \otimes T)$ [130].
[116] B. J. Brown, N. H. Nickerson, and D. E. Browne, Nature Communications **7**, 12302 (2016).
[117] H. Bombín, New Journal of Physics **17**, 083002 (2015).
[118] A. Kubica and M. E. Beverland, Phys. Rev. A **91**, 032330 (2015).
[119] M. Vasmer and D. E. Browne, Physical Review A **100**, 012312 (2019).
[120] Y. Fujiwara, (2014), arXiv:1405.6267.
[121] A. G. Fowler, D. Sank, J. Kelly, R. Barends, and J. M. Martinis, (2014), arXiv:1405.1454.
[122] M.-X. Huo and Y. Li, New J. Phys. **19**, 123032 (2017).
[123] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, Quantum **6**, 809 (2022).
[124] P. Iyer, A. Jain, S. D. Bartlett, and J. Emerson, Phys. Rev. Res. **4**, 043218 (2022).

[125] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, Phys. Rev. Lett. **130**, 200601 (2023).
[126] S. T. Flammia and J. J. Wallman, ACM Transactions on Quantum Computing **1**, 3:1 (2020).
[127] J. J. Wallman and J. Emerson, Phys. Rev. A **94**, 052325 (2016).
[128] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Nat Commun **10**, 5347 (2019).
[129] Y.-F. Wang, Y. Wang, Y.-A. Chen, W. Zhang, T. Zhang, J. Hu, W. Chen, Y. Gu, and Z.-W. Liu, (2023), arXiv:2312.09111.
[130] S. Koutsioumpas, D. Banfield, and A. Kay, (2022), arXiv:2210.14066.
[131] S. S. Bullock and I. L. Markov, Quantum Info. Comput. **4**, 27–47 (2004), arXiv:quant-ph/0303039.
[132] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).
[133] B. Schumacher, (1996), arXiv:quant-ph/9604023.
[134] M. A. Nielsen, (1996), arXiv:quant-ph/9606012.
[135] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).
[136] M. A. Nielsen, Physics Letters A **303**, 249 (2002), arXiv:quant-ph/0205035.
[137] J. J. Wallman and S. T. Flammia, New J. Phys. **16**, 103032 (2014).
[138] C. Gidney, Quantum **5**, 497 (2021).
[139] NetworkX developers, NetworkX version 3.2.1 (2024).
[140] L. Gurobi Optimization, Gurobi optimizer reference manual (2024).

CONTENTS

Appendix A: Quantum channels on phase states

We begin with a brief recap of the properties of IQP computations that will be helpful in the following. In particular, we analyze quantum channels which preserve phase states.

1. Some background on phase states

In its most general form, IQP computation is about the preparation and manipulation of a class of quantum states called *phase states*.

Definition 8. *Given a function $\phi : \mathbb{Z}_2^n \to \mathbb{R}$, a phase state $|\phi\rangle$ on n qubits is*

$$|\phi\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} e^{i\pi\phi(x)} |x\rangle . \tag{A1}$$

We denote the single-qubit $x/y/z$ Pauli operators by $X/Y/Z$ and use the notation $A^a \equiv A^{a_1} \otimes \cdots \otimes A^{a_n}$ for a bit string $a \in \{0,1\}^n$ and a single qubit operator A. A few convenient facts about phase states are

- A pure quantum state $|\psi\rangle$ is a phase state *iff* for every nonzero $a \in \{0,1\}^n$, $\langle\psi| Z^a |\psi\rangle = 0$. The forward direction is clear, and the reverse holds because states satisfying $\langle\psi| Z^a |\psi\rangle = 0$ for all nonzero a are precisely those states with equal magnitude coefficients on all computational basis states, i.e., phase states.

- Any function $\phi : \mathbb{Z}_2^n \to \mathbb{R}$ can be expanded in terms of polynomial coefficients $\alpha_a \in \mathbb{R}$ as

$$\phi(x) = \sum_{a\in\{0,1\}^n} \alpha_a x^a, \tag{A2}$$

  where $x^a = x_1^{a_1} \cdots x_n^{a_n}$.

- A phase state $|\phi\rangle$ with polynomial coefficients α_a can be prepared with a unitary operator composed of diagonal Z interactions

$$U_\phi = \prod_{a\in\{0,1\}^n} \exp\left[i\pi\alpha_a \prod_i \left(\frac{\mathbb{1} - Z_i}{2}\right)^{a_i}\right], \tag{A3}$$

  such that $|\phi\rangle = U_\phi |+^n\rangle$ where $|+^n\rangle$ is the n-fold tensor power of the single-qubit $+1$-eigenstate $|+\rangle$ of X.

- If $\phi(x)$ is a degree-D polynomial, i.e., a polynomial with maximum degree D, then U_ϕ can be implemented by a quantum circuit consisting of at most $O(2^D n^D)$ gates: There are $O(n^D)$ terms in the polynomial and the D-local gates $\exp\big(i\pi\alpha_a \prod_i ((\mathbb{1} - Z_i)/2)^{a_i}\big)$ can be decomposed into at most $O(2^D)$ two-qubit gates [131].

2. Equivalence of degree-D IQP and degree-D IQP + CNOT

Let us show the equivalence between degree-D IQP circuits and circuits composed of degree-D IQP gates and CNOT gates. Let us define a binary phase state as a state $|\psi\rangle$ which can be written as

$$|\psi\rangle = \sum_x (-1)^{f(x)} |x\rangle , \tag{A4}$$

where $f : \{0,1\}^n \to \{0,1\}$ is an arbitrary Boolean function. We now observe that a degree-D IQP circuit C_f

acting on n qubits prepares a polynomial binary phase state

$$C_f \left|+\right\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_x (-1)^{f(x)} \left|x\right\rangle, \tag{A5}$$

where $f(x) = \sum_{k=1}^{D} \sum_{j_1^k,\ldots j_k^k} c_{j^k}^k x_{j_1^k} \cdots x_{j_k^k}$ is a Boolean degree-D polynomial whose coefficients $c_{j^k}^k \in \{0,1\}$ are nonzero if a $\mathrm{C}^{k-1}\mathrm{Z}$ gate is applied to the k qubits labeled by $j^k = (j_1^k, \ldots, j_k^k)$. The action of a CNOT gate on a two-qubit basis state $\left|xy\right\rangle$ is given by $\mathrm{CNOT}\left|xy\right\rangle = \left|x, x \oplus y\right\rangle$, where $\oplus$ denotes addition modulo 2. This implies that the action of a circuit $\mathcal{C}$ composed of CNOT gates on a polynomial binary phase state is given by

$$\mathcal{C}C_f \left|+^n\right\rangle = \frac{1}{\sqrt{2^n}} \sum_x (-1)^{f(A^{-1}x)} \left|x\right\rangle = D_{f_A} \left|+^n\right\rangle, \tag{A6}$$

where A is a Boolean invertible matrix defined by the CNOT circuit. A^{-1} preserves the degree of f and hence $f(A^{-1}x)$ is just a different degree-D Boolean polynomial $f_A(x)$ that depends on f and A. Any hIQP circuit which includes IQP and CNOT gates is therefore equivalent to some purely IQP circuit.

3. Review of quantum operations

Let us now briefly review the basic formalism of quantum channel theory.

A *completely positive* (CP) map acts on density matrices ρ as $\mathcal{E}(\rho) = \sum_\mu K_\mu \rho K_\mu^\dagger$, where the linear operators K_μ are called *Kraus operators*. A *quantum channel* (or CPT map) $\mathcal{E}$ is a completely positive map that also preserves the trace, i.e., its Kraus operators satisfy $\sum_\mu K_\mu^\dagger K_\mu = \mathbb{1}$.

Definition 9. *Let P_n be the Pauli group on n qubits, then a Pauli channel is defined by a probability distribution $q(P)$ over $P \in P_n$ such that*

$$\mathcal{E}(\rho) = \sum_{P \in P_n} q(P) P \rho P^\dagger, \tag{A7}$$

i.e., we can take the Kraus operators to be Pauli operators.

To analyze the effects of quantum channels on phase states, we first introduce two (inequivalent) levels of conservation of the phase state property. First, we introduce *phase-state-preserving* maps (Appendix A 4), then *diagonal CP maps* (Appendix A 5).

4. Phase-state-preserving quantum operations

Definition 10. *A CP map $\mathcal{E}$ is a* phase-state-preserving (PSP) *map if, for any mixture of phase states ρ, $\mathcal{E}(\rho)$ can be written as a mixture $\sum_a p_a \left|\phi_a\right\rangle \left\langle\phi_a\right|$ of phase states $\left|\phi_a\right\rangle$ with $p_a \geq 0$.*

Definition 11. *A CP map $\mathcal{E}$ is a* completely-phase-state preserving (CPSP) *map if $\mathcal{E} \otimes \mathbb{1}$ is a PSP map for any trivial extension of $\mathcal{E}$ to a larger Hilbert space.*

We give some examples of channels that fall into different categories according to these definitions:

- The partial trace operation is a CPSP channel. Take a phase state $\left|\phi\right\rangle$ on the joint system $\mathcal{H}_A \otimes \mathcal{H}_B$ with n_A and n_B qubits, respectively. We can expand this state in the computational basis as $\left|\phi\right\rangle = \sum_{x,y} c_{x,y} \left|x\right\rangle \otimes \left|y\right\rangle$. Taking the partial trace over B, we arrive at the state

$$\begin{aligned} \mathrm{Tr}_B(\left|\phi\right\rangle \left\langle\phi\right|) &= \sum_{x,x',y} c_{x,y} c_{x',y}^* \left|x\right\rangle \left\langle x'\right| \\ &= \sum_y \frac{1}{2^{n_B}} \left|\phi_y\right\rangle \left\langle\phi_y\right|. \end{aligned} \tag{A8}$$

  Each state $\left|\phi_y\right\rangle := \sum_x c_{x,y} \left|x\right\rangle$ in this mixture is a phase state because $\left|c_{x,y}\right|^2 = 1/2^{n_A+n_B}$. This argument shows that the partial trace results in a mixture of phase states, proving the PSP condition. The trivial extension of the partial trace operation is also a partial trace, which shows the CPSP property.

- Let $\mathcal{U} = U \cdot U^\dagger$ with some unitary U be a unitary PSP channel on n qubits, then $\mathcal{U}$ is CPSP. To prove the CPSP condition we need to prove $\mathcal{U} \otimes \mathbb{1}$ is PSP acting on the doubled Hilbert space $\mathcal{H} \otimes \mathcal{H}$. Taking a joint state $\left|\phi\right\rangle$ and expanding it as we did in the partial trace discussion above, we can represent the state as the superposition

$$\left|\phi\right\rangle = \sum_z c_z \left|\phi_z\right\rangle \otimes \left|z\right\rangle, \tag{A9}$$

  where $\left|z\right\rangle$ is a computational basis state, $\left|c_z\right|^2 = 1/2^n$ and $\left|\phi_z\right\rangle$ is a phase state. Applying $\mathcal{U} \otimes \mathbb{I}$ to this state will result in a similar superposition of phase states each with equal magnitude coefficients; thus, the resulting state is also a phase state, proving the CPSP condition for $\mathcal{U}$.

- Let U be a circuit composed of CNOT gates and X gates, then $U\left|x\right\rangle = \left|Ax + b\right\rangle$ for some invertible $n \times n$ matrix A over $\mathbb{F}_2$ and a Boolean vector $b \in \mathbb{F}_2^n$. As a result, $U\left|\phi\right\rangle = \left|\phi'\right\rangle$, where $\phi'(x) = \phi[A^{-1}(x + b)]$, so all such U give rise to CPSP channels.

- More generally, let U be a circuit composed of Toffoli gates, CNOT and X gates. Then, U is a CPSP map with the property that $U\left|\phi\right\rangle = \left|\phi \circ f^{-1}\right\rangle$, where $f(x)$ is the invertible Boolean function implemented by the circuit on computational basis states.

- Any mixture of CPSP maps is a CPSP map.

- Any Pauli channel is a CPSP channel.

- A projective measurement in the x-basis is a 1-qubit PSP map, but is not a CPSP map. As a counterexample to the CPSP condition, consider the (unnormalized) two qubit phase state $|00\rangle + |01\rangle + |10\rangle - |11\rangle$ and measure the second qubit in the x-basis. The state after the measurement is $|0+\rangle$ or $|1-\rangle$, neither of which are phase states.
- The amplitude damping channel with Kraus operators $K_0 = (1+\sqrt{1-\gamma})\mathbb{1}/2 + (1-\sqrt{1-\gamma})Z/2$ and $K_1 = (X+iY)\sqrt{\gamma}/2$ is not PSP for any $\gamma > 0$.

From these examples, we can see that the set of operations that preserve the phase state property has a relatively rich structure.

5. Diagonal completely positive maps

In addition to phase-state preservation, another natural class of maps that act in a simple way on phase states are *diagonal* CP maps.

Definition 12. *A CP map $\mathcal{E}$ is diagonal if it has a Kraus decomposition $\{K_\mu\}$ such that every Kraus operator $K_\mu = \sum_a c_a Z^a$ is diagonal in the computational basis.*

We denote a map as a *diagonal-unitary CPSP map* if it has a representation as a mixture of diagonal maps composed with unitary CPSP maps.

6. The z-twirl

In the context of noisy random IQP computations, it is convenient to introduce z-twirling of a CP map.

Definition 13. *The z-twirl of a CP map $\mathcal{E}$ with Kraus operators $\{K_\mu\}$ is the CP map*

$$\mathcal{E}_z(\rho) = \frac{1}{2^n}\sum_{a\mu} Z^a K_\mu Z^a \rho Z^a K_\mu^\dagger Z^a \tag{A10}$$

This twirling operation has the effect of converting any CP map into a diagonal-unitary CPSP map as we now show.

Lemma 14. *Let $\mathcal{E}$ be a CP map; its z-twirl $\mathcal{E}_z$ is a diagonal-unitary CPSP map.*

Proof. Let K_μ be the Kraus operators for $\mathcal{E}$, we expand them in the Pauli basis

$$K_\mu = \frac{1}{2^n}\sum_{ab} c^\mu_{ab} X^b Z^a,\ c^\mu_{ab} = \mathrm{Tr}[K_\mu Z^a X^b]. \tag{A11}$$

With this representation, we can then express the z-twirled map as

$$\mathcal{E}_z(\rho) = \sum_{\mu b} p_{\mu b} X^b K_{\mu b} \rho K_{\mu b}^\dagger X^b, \tag{A12}$$

where $p_{\mu b} = \sum_a |c^\mu_{ab}|^2$ and the Kraus operators $K_{\mu b} = \frac{1}{\sqrt{p_{\mu b}}}\sum_a c^\mu_{ab} Z^a$. The effect of the twirling is to remove the terms on which the X operators act with a different power on each side of the density matrix. This representation of the Kraus maps proves the twirled map is a diagonal-unitary CPSP map because $K_{\mu b}$ is a diagonal map and X^b is a unitary CPSP map.

□

Appendix B: Measuring the quality of noisy encoded computations

Measuring the quality of encoded computations in the absence of full error correction is nontrivial, because physical errors affect encoded qubits quite differently compared to physical qubits. In this section, we take some initial steps towards a theory of *logical gate fidelities* (Appendix B 2). Before we do so, we recap the basics of stabilizer codes and analyze the effective noise channels arising on the logical qubits (Appendix B 1).

1. Stabilizer codes and logical noise channels

Let S be an abelian subgroup of the Pauli group such that $-\mathbb{1} \notin S$. The size of S is 2^{n-k} for an integer k. The code space projector

$$\Pi_0 = \frac{1}{2^{n-k}}\sum_{g\in S} g, \tag{B1}$$

defines an $[\![n,k,d]\!]$ quantum error correcting code where d is the code distance. Taking a minimal size generating set for $S = \langle g_1, \ldots, g_{n-k}\rangle$, we can re-express $\Pi = \prod_{i=1}^{n-k}(\mathbb{1} + g_i)/2$.

For two Pauli operators P and Q, define the scalar commutator as $[[P,Q]] = 0 \in \mathbb{F}_2$ if $[P,Q] = 0$ and $[[P,Q]] = 1 \in \mathbb{F}_2$ if $PQ = -QP$. We can define a canonical basis for the Pauli group with reference to our generating set for S as follows: find a set of operators $\{h_1, \ldots, h_{n-k}\}$ called destabilizers [132] that satisfy $[[h_i, h_j]] = 0$ and $[[g_i, h_j]] = \delta_{ij}$, as well as a set of logical operators $\{\bar{X}_1, \bar{Z}_1, \ldots, \bar{X}_k, \bar{Z}_k\}$ that satisfy $[[\bar{X}_i, h_j]] = [[\bar{X}_i, g_j]] = [[\bar{X}_i, \bar{X}_j]] = 0$, $[[\bar{Z}_i, h_j]] = [[\bar{Z}_i, g_j]] = [[\bar{Z}_i, \bar{Z}_j]] = 0$, and $[[\bar{X}_i, \bar{Z}_j]] = \delta_{ij}$. This canonical basis set can be found efficiently by solving a set of linear equations over $\mathbb{F}_2$.

A logical state $|\bar{\psi}\rangle$ is any state that satisfies $\Pi_0 |\bar{\psi}\rangle = g|\bar{\psi}\rangle = |\bar{\psi}\rangle$ for all $g \in S$. A *syndrome* is a set of measurement outcomes $s \in \{0,1\}^{n-k}$ for the stabilizer generators g_i. The projector onto a given syndrome outcome is

$$\Pi_s = \prod_{i=1}^{n-k} \frac{\mathbb{1} + (-1)^{s_i} g_i}{2}. \tag{B2}$$

We also define Π_{0r} as the projector onto the *r-error subspace*

$$\mathrm{span}\Big(\{P\left|\bar{\psi}\right\rangle : \Pi_0\left|\bar{\psi}\right\rangle = \left|\bar{\psi}\right\rangle,\, P\in P_n,\, \mathrm{wt}(P)\le r\}\Big),$$

where $P=\bigotimes_{i=1}^n P_i$ is a Pauli operator and $\mathrm{wt}(P)=|\{i : P_i\neq \mathbb{1}\}$ is the *weight* of P, i.e., the number of non-identity Paulis in P. The projector Π_{0r} can be interpreted as a projector onto the set of states that have at most r errors. Note, $\Pi_{0n}=\mathbb{1}$, while the smallest m such that $\Pi_{0m}=\mathbb{1}$ is greater than or equal to the code distance, but will generally depend on the code.

In order to formalize the effect of physical noise on the logical subspace it is convenient to introduce a generalization of a quantum channel, called *quantum instrument*.

Definition 15. *A* quantum instrument *is a collection of CP maps* $\{\mathcal{E}_a\}_{a\in A}$ *such that* $\sum_{a\in A}\mathcal{E}_a$ *is a quantum channel.*

A projective measurement $\{\Pi_a\}_{a\in M}$ is an example of a quantum instrument, where a indexes the possible measurement outcomes M.

For a given physical noise channel $\mathcal{E}$, the corresponding *logical noise channel* is the quantum instrument $\{\mathcal{E}_s\}_{s\in\{0,1\}^{n-k}}$ associated with the syndrome outcomes

$$\mathcal{E}_s(\rho)=\Pi_s\mathcal{E}(\rho)\Pi_s=\sum_\mu \Pi_s K_\mu\rho K_\mu^\dagger\Pi_s. \tag{B3}$$

For an initial state in the code space, these are effectively logical channels with the Kraus operators

$$\Pi_s K_\mu\Pi_0 = h^s\Pi_0 h^s K_\mu\Pi_0=\sum_{abc} c^\mu_{abcs} h^s g^c\bar{X}^b\bar{Z}^a\Pi_0, \tag{B4}$$

where $h^s\equiv\prod_{i=1}^{n-k}h_i^{s_i}$ is a product of destabilizers that corrects the syndrome and $c^\mu_{abcs}=\mathrm{Tr}[K_\mu\bar{Z}^a\bar{X}^b g^c h^s]$.

2. Logical gate fidelities

We are now ready to define notions of logical transversal gate fidelities. Due to entanglement that builds up between syndrome degrees of freedom and logical degrees of freedom during the computation, the error model for logical qubits is intrinsically non-Markovian, increasing the relative importance of circuit context dependence in defining gate fidelities. We provide a partial resolution to this challenge through the use of r-error filters that restrict the input states of the logical channel to have at most weight-r errors.

Given a syndrome outcome s, a *decoder* is a map from s to a logical Pauli operator $\bar{P}_s$ chosen to correct the error and a destabilizer $\bar{h}^s$ that fixes the syndrome without affecting the logical degrees of freedom. We define the (r,s)-*filtered average gate fidelity* of a decoded logical operation $\bar{\mathcal{U}}=\bar{U}\cdot\bar{U}^\dagger$ conditioned on a particular syndrome outcome s acting on an r-error state as

$$\begin{aligned} F_{sr}(\mathcal{E},\bar{\mathcal{U}}) &= \frac{1}{p_{sr}}\int d\psi_r\,\langle\psi_r|\,\bar{U}^\dagger\bar{P}_s h^s \\ &\quad\times\mathcal{E}_s(\bar{U}\,|\psi_r\rangle\langle\psi_r|\,\bar{U}^\dagger)h^s\bar{P}_s U\,|\psi_r\rangle\,, \qquad \text{(B5)}\\ p_{sr} &= \mathrm{Tr}[\mathcal{E}_s(\Pi_{0r}/d_r)], \qquad \text{(B6)}\\ d_r &= \mathrm{Tr}[\Pi_{0r}], \qquad \text{(B7)}\end{aligned}$$

where $\mathcal{E}\circ\bar{\mathcal{U}}(\rho)$ is the noisy implementation of the logical gate $\bar{\mathcal{U}}$, h^s maps the states with syndrome s back to the logical states, $d\psi_r$ is the uniform measure on the set of r-error states induced by the Haar measure over logical operators and p_{sr} is the probability of observing syndrome s starting from a maximally mixed state in the r-error space. In the case of error detection, where the optimal decoding strategy is to postselect on the all-0 syndrome outcome and set $\bar{P}_0=\mathbb{1}$, the average gate fidelity is

$$F_{0r}(\mathcal{E},\bar{\mathcal{U}})=\frac{1}{p_{0r}}\int d\psi_r\,\langle\psi_r|\,\bar{U}^\dagger\mathcal{E}_0(\bar{U}\,|\psi_r\rangle\langle\psi_r|\,\bar{U}^\dagger)U\,|\psi_r\rangle\,. \tag{B8}$$

In addition to the average gate fidelity, we also introduce the (r,s)-*filtered entanglement fidelity* [133, 134]

$$F_{esr}(\mathcal{E},\bar{\mathcal{U}})=\frac{1}{p_{sr}}\langle\Phi_r|\,(\mathcal{U}^\dagger\circ\mathcal{R}_s\circ\mathcal{E}_s\circ\mathcal{U})\otimes\mathbb{1}(|\Phi_r\rangle\langle\Phi_r|)\,|\Phi_r\rangle\,, \tag{B9}$$

where $\mathcal{R}_s(\rho)=\bar{P}_s h^s\rho h^s\bar{P}_s$ and $|\Phi_r\rangle=\mathrm{vec}(\Pi_{0r}/d_r)$ is a purification of the density matrix Π_{0r}/d_r. As with their standard versions, we find that the (r,s)-filtered entanglement fidelity is related to the (r,s)-filtered average gate fidelity through the formula [135, 136]

$$F_{sr}=\frac{d_r F_{esr}+1}{d_r+1}. \tag{B10}$$

To see this equality, one follows the simple proof of Nielsen [136]. Similarly, we can use the r-filtered average gate fidelity to upper- and lower-bound an (r,s)-filtered version of the diamond distance (defined analogously with respect to r-error input states and decoding conditioned on syndrome outcome s) of the logical channel from the identity [e.g. 137, Prop. 9].

We can analytically compute the (r,s)-filtered entanglement fidelity as

$$F_{esr}=\frac{1}{p_{sr}d_r^2}\sum_\mu\left|\mathrm{Tr}\big[\Pi_0\bar{P}_s h^s K_\mu\Pi_{0r}\big]\right|^2, \tag{B11}$$

where $\{K_\mu\}$ are a set of Kraus operators for $\mathcal{E}$. For two-qubit logical gates between code blocks (e.g., the transversal CNOT) and two-qubit gate error channels that factorize $\mathcal{E}=\mathcal{E}_1\otimes\mathcal{E}_2$, we can directly extend this formula to two-qubit logical gates

$$F_{ess'rr'}=F^{(1)}_{esr}F^{(2)}_{es'r'}. \tag{B12}$$

These definitions of logical gate fidelities implicitly assume either that r is less than the code distance or that most states in the r-error space are close to a logical state. Otherwise, the r-filtered gate fidelity is dominated by the projection back to the code space. In the case of near-term experiments where accessible code distances are small, which is the focus of this work, it is convenient to define an intermediate notion of logical fidelity that looks at the fidelity of the logical operation on the whole r-error space without error correction and decoding following the gate. The error rate of this fidelity does not decrease with increasing code distance, but provides a useful bound on the fidelity of logical gates in a near-term logical circuit. We thus define the *r-filtered average gate fidelity* and *r-filtered entanglement fidelity* as

$$F_r(\mathcal{E},\bar{\mathcal{U}}) = \frac{d_r F_{er}(\mathcal{E},\bar{\mathcal{U}})+1}{d_r+1}, \tag{B13}$$

$$F_{er}(\mathcal{E},\bar{\mathcal{U}}) = \frac{1}{p_r}\langle\Phi_r|\,(\bar{\mathcal{U}}^\dagger\circ\mathcal{E}\circ\bar{\mathcal{U}})\otimes\mathbb{1}(|\Phi_r\rangle\langle\Phi_r|)\,|\Phi_r\rangle\,, \tag{B14}$$

$$p_r = \frac{1}{d_r}\mathrm{Tr}[\Pi_{0r}\mathcal{E}(\Pi_{0r})], \tag{B15}$$

respectively. In the case of $[\![8,3,2]\!]$ codes, for $r=1$, Π_{01} has support on 144 out of all possible 256 states. For $r\geq 2$, Π_{0r} has support on the whole 8-qubit Hilbert space. Consequently, in this case F_{sr} is negligible for any $r>0$; however, F_r can still be large because of the small size of the code.

In Fig. 4(b) in the main text, we present the r-filtered gate fidelities F_r for transversal CCZ, in-block CNOT, and transversal CNOT gates of the $[\![8,3,2]\!]$ code. In the calculations for the CCZ and transversal CNOT gates we took iid depolarizing noise on each qubit following the application of a physical gate. For the in-block CNOT, which we assume to be implemented via qubit reconfiguration, we modeled the noise as pure Z-dephasing noise on each qubit following the gate. To numerically compute the r-filtered fidelities we directly evaluated the formulas above by expressing them as polynomials in the local noise rate parameter $p(1-p)$. We then used direct calculations on the 2^8 dimensional Hilbert space to evaluate the coefficients of each polynomial.

For a native two-qubit gate error rate of 0.5%, the fidelity of a compiled CCZ gate would be around 3%. We see in Fig. 4(b) that the r-filtered fidelity can readily exceed that for $r=0$, while at high enough single-qubit gate error rates $10^{-3}-10^{-4}$ the transversal CCZ fidelity exceeds the compiled version for $r>0$. In the case of transversal CNOTs shown in Fig. 4(b), the r-filtered fidelity only exceeds the native fidelity for $r=0$. The pseudothreshold for two-qubit CNOT gates is observed to be at gate error rates of 4%. When comparing the fidelity of physical CNOTs implemented on 3 independent pairs, which is most comparable to the logical transversal CNOT gate for the $[\![8,3,2]\!]$ code, the pseudothreshold increases to 14%.

Appendix C: Model for fidelity versus acceptance fraction

As we increase the amount of postselection in the encoded circuit, the XEB improves but the number of discarded samples grows as well. In Fig. 4(a) of the main text and in the experimental data (Fig. 5(e)) of Ref. [27], the relationship between the XEB and the fraction of accepted samples reduces to a power law in a certain regime.

Here, we provide a simple model based on the $[\![8,3,2]\!]$ code that reproduces this behavior. While the exponent will differ for a different code, we expect the power-law behaviour to generalize. We consider n logical qubits encoded into $n/3$ blocks of the $[\![8,3,2]\!]$ code, which experience an i.i.d. dephasing channel with rate p on all physical qubits, right before measurement in the X basis. For each block, we can evaluate the probability of a logical Z errors conditioned on whether the single X stabilizer is violated or not; this is enough to recover the power law behavior mentioned above. The probability of having $k\in\{0,\ldots,8\}$ errors in a block is

$$P_k = \binom{8}{k}p^k(1-p)^{8-k}, \tag{C1}$$

and the corresponding probability of violating the stabilizer is

$$\epsilon = \sum_{k\in\{1,3,5,7\}} P_k, \tag{C2}$$

which corresponds to having an odd number of errors. The logical operators within a block are defined up to multiplication with stabilizers, but in practice we choose a particular operator realization, e.g., the three sides of a cube highlighted in Fig. 1(d). Then, for every odd number of errors, we have a 7/8 probability of having at least one logical fault, since for any choice of logical operators, a single physical qubit does not participate in any of them. Similarly, the probability of having a logical fault given an even number of physical errors is $\{0,1,4/5,1,0\}$ for $\{0,2,4,6,8\}$ errors, respectively. Together, this results in probabilities p_l^s of logical faults conditioned on whether a stabilizer is violated ($s=-$) or not ($s=+$), which are given by

$$p_l^- = 7/8, \tag{C3}$$

$$p_l^+ = \frac{1}{1-\epsilon}(P_2+\frac{4}{5}P_4+P_6). \tag{C4}$$

Given $n_b=n/3$ blocks, the probability of having exactly j stabilizers violated is

$$S_j = \binom{n_b}{j}\epsilon^j(1-\epsilon)^{n_b-j}, \tag{C5}$$

and the probability of having no logical error in that situation is

$$F_j = (1-p_l^-)^j(1-p_l^+)^{n_b-j}. \tag{C6}$$

The acceptance fraction for up to m violated stabilizers is then

$$\bar{S}_m = \sum_{j \leq m} S_j, \tag{C7}$$

and the fidelity (i.e., the probability of no logical errors) is

$$\bar{F}_m = \frac{1}{\bar{S}_m} \sum_{j \leq m} S_j F_j, \tag{C8}$$

which results in an approximate power law (when evaluated numerically) for the fidelity $\bar{F}_m$ as a function of acceptance fraction $\bar{S}_m$. This is in good qualitative agreement with circuit-level simulations such as those in Fig. 4(a).

Finally, we provide an analytical formula for the power-law exponent in the low-noise, highly post-selected regime. Assuming a power-law relationship $\bar{F}_m = a\,\bar{S}_m^{\alpha}$, we can estimate α in the post-selected regime from $\bar{F}_1/\bar{F}_0$ and $\bar{S}_1/\bar{S}_0$ using Eqs. (C7) and (C8). To first order in p, these are given by

$$\bar{F}_1/\bar{F}_0 = 1 - 7n_b\,p + O(p^2), \tag{C9}$$

$$\bar{S}_1/\bar{S}_0 = 1 + 8n_b\,p + O(p^2), \tag{C10}$$

which gives $\alpha = -7/8$ and is, to first order, independent of both p and n_b. In Fig. 4(a), we show a power law fit with this value of α, and observe good agreement with simulated data.

Appendix D: Statistical model for second moment quantities of IQP circuits

To characterize random circuits in terms of their output distributions in the noiseless case and their quality of implementation in the presence of noise, it is useful to analyze their *second-moment quantities*. Second-moment quantities include all quantites which can be expressed as average quantities that are linear in two tensor copies of the random circuits applied to a fixed initial state. Most importantly for us, the average collision probability, the cross-entropy benchmark, the average fidelity and the average purity are all second-moment quantities. In this work, we are interested in those quantities for random degree-3 IQP circuits.

In this section, we derive a framework for analysing the second-moment properties of ideal and noisy circuits with some random IQP degree-2 gates whose output is a polynomial binary phase state

$$|\psi\rangle = \sum_x (-1)^{f(x)} |x\rangle\,, \tag{D1}$$

for some degree-2 Boolean polynomial f. This is simpler than analyzing degree-3 circuits directly, and will allow us to make statements about random degree-3 circuits as well, since we can fix the degree-3 part of a circuit and then average over its degree-2 part.

Importantly, our framework will *not* require the IQP circuits to be *globally random* but only *locally random*. Specifically, we consider circuit architectures, in which the first layer is a layer of Hadamard gates, on certain pairs (i,j) of qubits, a random degree-2 IQP circuit $Z_i^u Z_j^v CZ_{ij}^w$ for $u,v,w \in \{0,1\}$ is applied. All other components of the (ideal) circuit are arbitrary gates which map binary phase states to binary phase states. We also allow for noise which preserves binary phase states.

Choose an IQP circuit $D \sim \mathfrak{I}$ from some IQP ensemble $\mathfrak{I}$ satisfying the properties above. Let us then write $\rho_D(\mathcal{N})$ for the mixture of binary phase states generated by D with noise specified by $\mathcal{N}$. The ideal state is given by $\rho_D(\mathbb{1}) = D\,|0\rangle\langle 0|_x\,D$. Our central quantity of interest is now the *second-moment operator*

$$M_2(\mathcal{N},\mathcal{N}') = \mathbb{E}_{D\sim\mathfrak{I}}\left[\rho_D(\mathcal{N}'), \rho_D(\mathcal{N})\right]. \tag{D2}$$

To be most general, we allow different types of noise on the two copies.

The average collision probability $\overline{C}$, average XEB fidelity $\overline{\chi}$, average fidelity $\overline{F}$, and average purity $\overline{P}$ can be written in terms of Eq. (D2) as

$$\overline{C} = \mathrm{tr}\big[\mathbb{X}_x^{\otimes n} M_2(\mathbb{1},\mathbb{1})\big] \tag{D3}$$

$$\overline{\chi} = 2^n\,\mathrm{tr}\big[\mathbb{X}_x^{\otimes n} M_2(\mathbb{1},\mathcal{N})\big] - 1 \tag{D4}$$

$$\overline{F} = \mathrm{tr}\big[\mathbb{S}^{\otimes n} M_2(\mathbb{1},\mathcal{N})\big] \tag{D5}$$

$$\overline{P} = \mathrm{tr}\big[\mathbb{S}^{\otimes n} M_2(\mathcal{N},\mathcal{N})\big] \tag{D6}$$

where we have defined $\mathbb{S} = \sum_{x,y\in\{0,1\}} |xy\rangle_a \langle yx|_a$ as the single-qubit swap operator between the two copies, and $\mathbb{X}_a = \sum_{x\in\{0,1\}} |xx\rangle_a \langle xx|_a$ as the single-qubit 'delta projector' in the a-basis for $a \in \{x,z\}$. Note that while the collision probability involves no noise on either copy in M_2, the average fidelity and average XEB fidelity involve noise on one copy and no noise on the other copy. In contrast, the purity involves noise on both copies.

Let us now derive the formalism. To start out, we consider D chosen uniformly at random from degree-2 IQP circuits. As discussed in Appendix A 1, each such circuit is characterized by a degree-2 polynomial $f(x) = \sum_{ij} b_{ij}x_ix_j + \sum_i a_ix_i$ with uniformly random choices of $a \in \{0,1\}^n, b \in \{0,1\}^{n\times n}$ specifying the locations of the Z and CZ gates.

1. The global second moment

In order to compute $M_2(\mathbb{1},\mathbb{1})$ we build on the proof of anticoncentration of globally random degree-2 IQP circuits by Bremner *et al.* [24] (Lemma 11 in App. F). Let $\mathfrak{I}_n$ be the family of uniformly random degree-2 IQP circuits on n qubits, and $\mathbb{P} = \sum_{xy\in\{0,1\}} |xx\rangle\langle yy|$ the unnormalized Bell state projector.

Lemma 16 (Global projector)**.** *Consider the moment operator M_2 for random degree-2 IQP circuits in $\mathfrak{I}_n$. It holds that*

$$M_2(\mathbb{1},\mathbb{1}) = \frac{1}{2^{2n}}(\mathbb{1}^{\otimes n} + \mathbb{S}^{\otimes n} + \mathbb{P}^{\otimes n} - 2\mathbb{X}_z^{\otimes n}). \tag{D7}$$

Before we prove the lemma, we would like to observe the analogy with the second moment operator of the Haar measure, which is given by

$$\mathbb{E}_{U\sim U(d)} U^{\otimes 2} |0\rangle\langle 0| (U^\dagger)^{\otimes 2} = \frac{1}{2d_{[2]}}(\mathbb{1}^{\otimes n} + \mathbb{S}^{\otimes n}), \tag{D8}$$

where $d_{[2]}$ is the dimension of the symmetric subspace of two d-dimensional systems. While the diagonal action of the full unitary group is only invariant under a global swap of the two copies, giving rise to the swap operator in the moment formula (D8), the diagonal action of random IQP circuits on binary phase states has additional invariances, namely, precisely those captured by the $\mathbb{P}$ and $\mathbb{X}_z$ operators.

Proof. To show Lemma 16, we expand

$$M_2(\mathbb{1},\mathbb{1}) = \frac{1}{2^{2n}} \sum_{z_1,\ldots,z_4} \mathbb{E}_f \left[(-1)^{\sum_i f(z_i)}\right] |z_1 z_3\rangle \langle z_2 z_4|_z , \tag{D9}$$

and observe that

$$\mathbb{E}_f \left[(-1)^{\sum_i f(z_i)}\right] \propto \sum_{a\in\{0,1\}^n} \sum_{b\in\{0,1\}^{n\times n}} (-1)^{\sum_i \left(\sum_{kl} b_{kl} z_{ik} z_{il} + \sum_k a_k z_{ik}\right)}. \tag{D10}$$

Now, we find criteria on the z_i that have to be satisfied for the coefficient (D10) to be nonzero. Following Ref. [24] this average results in three constraints on the sums,

$$\begin{aligned} &(\mathbb{1}) \quad z_1 = z_2 =: x \text{ and } z_3 = z_4 =: y, \quad \text{or} \\ &(\mathbb{S}) \quad z_1 = z_4 =: x \text{ and } z_2 = z_3 =: y, \quad \text{or} \\ &(\mathbb{P}) \quad z_1 = z_3 =: x \text{ and } z_2 = z_4 =: y. \end{aligned} \tag{D11}$$

For the sake of completeness, we recall the derivation here. Consider the k-th bits $z_{1k}, z_{2k}, z_{3k}, z_{4k}$ of the vectors z_i. Then any nonzero term in the exponent will result in the sum to vanish.

Hence, the sum over a_k combined with the linear term $\sum_k a_k z_{ik}$ enforces $z_{1k} = z_{2k} + z_{3k} + z_{4k}$. Substituting this, the sum over b_{kl} now enforces that

$$\begin{aligned} 0 &= z_{1k}z_{1l} + z_{2k}z_{2l} + z_{3k}z_{3l} + z_{4k}z_{4l} \\ &= (z_{2k} + z_{3k} + z_{4k})(z_{2l} + z_{3l} + z_{4l}) \\ &\quad + z_{2k}z_{2l} + z_{3k}z_{3l} + z_{4k}z_{4l} \\ &= z_{2k}(z_{3l} + z_{4l}) + z_{3k}(z_{2l} + z_{4l}) + z_{4k}(z_{2l} + z_{3l}), \end{aligned} \tag{D12}$$

where we removed terms that appeared an even number of times ($z_{2k}z_{2l}$ etc.) since they evaluate to 0. From the pigeonhole principle, we know that at least two of the $\{z_{2k}, z_{3k}, z_{4k}\}$ must be the same.

To show constraint ($\mathbb{1}$), choose $z_{3k} = z_{4k} = y_k$ and rewrite the above condition as,

$$(z_{2k} + y_k)(z_{3l} + z_{4l}) = 0,$$

where we again removed the term $2z_{2l}y_k$. There are now two cases. In the first case, if $z_{2k} \neq y_k$ we immediately obtain $z_{3l} = z_{4l}$ for all l and thus $z_3 = z_4$ and $z_1 = z_2$. Otherwise, if $z_{2k} = y_k$, then we have $z_{1k} = z_{2k} = z_{3k} = z_{4k}$ which is covered by all of the conditions in Eq. (D11). The equality constraint $z_{1l} = z_{2l} = z_{3l} = z_{4l}$ then enforces one of the conditions in Eq. (D11) on the l-th bit. If the z_{il} are not all equal, we can apply the first case to obtain one of the conditions in Eq. (D11) for all bits. Constraints ($\mathbb{S}$) and ($\mathbb{P}$) follow from the same argument starting with the choice $z_{2k} = z_{3k}$ and $z_{2k} = z_{4k}$, respectively.

The only overlap between the conditions is when

$$(\mathbb{X}_z) \quad z_1 = z_2 = z_3 = z_4 = x. \tag{D13}$$

Thus, we can evaluate these three conditions independently and then subtract the over-counted terms corresponding to all four bitstrings being equal.

Each of the conditions ($\mathbb{1}$), ($\mathbb{S}$), ($\mathbb{P}$), ($\mathbb{X}_z$) in Eqs. (D11) and (D13) results in the averaged exponents in Eq. (D10) to be trivially even-valued, since every term will appear in pairs. Thus, the average coefficient evaluates to 1. The statement of the lemma follows from inserting the conditions in Eq. (D9). □

Lemma 16 directly implies the anticoncentration lemma of Bremner *et al.* [24].

Corollary 17 (Anticoncentration of random IQP)**.** *The average collision probability of random degree-2 IQP circuits is given by* $3 \cdot 2^{-n} - 2^{-2n+1}$.

Proof. We apply Lemma 16 to Eq. (D3) to find

$$\overline{C} = \frac{1}{2^{2n}} \operatorname{tr}\left[\mathbb{X}_x^{\otimes n}(\mathbb{1}^{\otimes n} + \mathbb{S}^{\otimes n} + \mathbb{P}^{\otimes n} - 2\mathbb{X}_z^{\otimes n})\right] \tag{D14}$$

$$= \frac{1}{2^{2n}}(2^n + 2^n + 2^n - 2) = 3 \cdot 2^{-n} - 2^{-2n+1}, \tag{D15}$$

where we have used that $\operatorname{tr}[\mathbb{X}_x \mathbb{O}] = \operatorname{tr}[\mathbb{X}_z \mathbb{O}]$ for $\mathbb{O} = \mathbb{1}, \mathbb{S}, \mathbb{P}$ since the Hadamards cancel, and $\operatorname{tr}[\mathbb{X}_z \mathbb{X}_x] = 1$. □

2. Statistical mechanics model for locally random degree-2 circuits

In the following, we will use the properties of IQP circuits with locally restricted random degree-2 parts. To this end, we follow a line of thought developed in Refs. [22, 36, 40]. There, the observation is, that layers

of single-qubit Haar-random gates in the circuit project the corresponding second-moment operator locally onto the space spanned by the $\mathbb{1}$ and $\mathbb{S}$ operators. The effect of any gate between those layers of single-qubit Haar-random gates can thus be projected onto this space. As it turns out, the update rules of the $\mathbb{1}$ and $\mathbb{S}$ operators under such two-qubit gates are quite simple.

We will proceed to derive an analogous model that lets us analyze the second-moment properties of noisy and noiseless IQP circuits with certain non-random gates. The relevant state space, in the most general case, is locally spanned by the invariant subspaces of those random IQP circuits.

Lemma 18 (IQP invariants)**.** *Let* $O = \sum_{x,y,z,w=0}^{2^n-1} q(x,y,z,w) \left|xy\right\rangle \left\langle zw\right|$ *be an operator on two copies of* n *qubits. Then*

$$\begin{aligned}&\mathbb{E}_{D\sim\mathfrak{I}_n}[D^{\otimes 2} O D^{\otimes 2}]\\ &= \sum_{x\neq y=0}^{2^n-1} \big[q(x,y,x,y)\left|xy\right\rangle\left\langle xy\right| + q(x,y,y,x)\left|xy\right\rangle\left\langle yx\right|\\ &+ q(x,x,y,y)\left|xx\right\rangle\left\langle yy\right|\big] + \sum_{x=0}^{2^n-1} q(x,x,x,x)\left|xx\right\rangle\left\langle xx\right|.\end{aligned} \tag{D16}$$

Proof. We follow the same argument as in the proof of Lemma 16. □

Lemma 18 implies that there are $3\cdot(2^{2n}-2^n)+2^n$ linearly independent operators $\left|xy\right\rangle\left\langle xy\right|, \left|xy\right\rangle\left\langle yx\right|, \left|xx\right\rangle\left\langle yy\right|$ for all $x,y\in\{0,1\}^n$ which are invariant under the average over random IQP circuits. The IQP-average projects an input operator O onto the space spanned by these operators.

However, for the circuits that we will be interested in, we can significantly reduce this number. To this end, we consider circuits in which there are layers of minimal units of random IQP circuits on two qubits interleaved with other, non-random gates. This is a toy model of the hIQP circuits, where we apply layers of random IQP circuits on blocks of three qubits interleaved with layers of CNOT gates.

From Lemma 18, it follows that there are $3\cdot(16-4)+4=40$ linearly independent operators on the two qubits. We reduce this number to 16 operators which are written as arbitrary products of 4 linearly independent operators per qubit, exploiting that our circuits are quite restrictive.

Specifically, we consider the mutually orthogonal operators on two copies of a single qubit

$$\begin{aligned}\mathsf{I} &= \mathbb{1} - \mathbb{X}_z && \text{(D17)}\\ &= \left|01\right\rangle\left\langle 01\right| + \left|10\right\rangle\left\langle 10\right|\\ \mathsf{S} &= \mathbb{S} - \mathbb{X}_z && \text{(D18)}\\ &= \left|01\right\rangle\left\langle 10\right| + \left|10\right\rangle\left\langle 01\right|\\ \mathsf{P} &= \mathbb{P} - \mathbb{X}_z && \text{(D19)}\\ &= \left|00\right\rangle\left\langle 11\right| + \left|11\right\rangle\left\langle 00\right|\\ \mathsf{X} &= \mathbb{X}_z && \text{(D20)}\\ &= \left|00\right\rangle\left\langle 00\right| + \left|11\right\rangle\left\langle 11\right|.\end{aligned}$$

We define the local state space to be $\mathcal{S} = \{\mathsf{I},\mathsf{S},\mathsf{P},\mathsf{X}\}$. It will also be useful to define $\mathcal{S}_{\neg\mathsf{X}} = \mathcal{S}\setminus\{\mathsf{X}\}$. The operators $\mathsf{I},\mathsf{S},\mathsf{P},\mathsf{X}$ directly correspond to mutually exclusive constraints $(\mathsf{I}),(\mathsf{S}),(\mathsf{P}),(\mathsf{X})$ analogous to the (non-exclusive) constraints $(\mathbb{1}),(\mathbb{S}),(\mathbb{P}),(\mathbb{X}_z)$ defined in Eqs. (D11) and (D13).

The circuit components we now need to analyze are

1. state preparation of $\left|+^n\right\rangle\left\langle +^n\right|^{\otimes 2}$,
2. layers of random IQP circuits $\prod_{i=1}^{n/2} D_{2i,2i+1}^{\otimes 2}$, where $D_{2i,2i+1}\in\mathfrak{I}_2$ are random degree-2 IQP circuits on two qubits,
3. $\mathrm{CNOT}_{i,j}^{\otimes 2}$ gates between any two sites i and j,
4. X and Z noise, that is operators $X\otimes\mathbb{1}$, $Z\otimes\mathbb{1}$ applied to the two-copy state.

Let us first consider state preparation. We find that under a z-twirl the two-copy state $\left|+^n\right\rangle\left\langle +^n\right|^{\otimes 2}$ can be expressed as

$$\begin{aligned}&\mathbb{E}_{z\in\{0,1\}}(Z^z)^{\otimes 2}\left|+\right\rangle\left\langle +\right|^{\otimes 2}(Z^z)^{\otimes 2} && \text{(D21)}\\ &= \frac{1}{8}\sum_{u,v,w,x,z\in\{0,1\}} Z^z\left|u\right\rangle Z^z\left|v\right\rangle\left\langle w\right| Z^z\left\langle x\right| Z^z && \text{(D22)}\\ &= \frac{1}{4}(\mathsf{I}+\mathsf{S}+\mathsf{P}+\mathsf{X})^{\otimes n}. && \text{(D23)}\end{aligned}$$

In the following analysis we will therefore always assume that the first layer of the circuit is a layer of random Z gates. This is no restriction in all cases we consider since this layer can be absorbed in the random Z gates throughout the circuit.

Lemma 19 (Projection update)**.** *Let* $Q\in\mathcal{S}$*, and* $P,R\in\mathcal{S}_{\neg\mathsf{X}}$*. Then*

$$\begin{aligned}\mathbb{E}_{D\in\mathfrak{I}_2}\left[D^{\otimes 2}(Q\otimes\mathsf{X})D^{\otimes 2}\right] &= Q\otimes\mathsf{X} && \text{(D24)}\\ \mathbb{E}_{D\in\mathfrak{I}_2}\left[D^{\otimes 2}(\mathsf{X}\otimes Q)D^{\otimes 2}\right] &= \mathsf{X}\otimes Q && \text{(D25)}\\ \mathbb{E}_{D\in\mathfrak{I}_2}\left[D^{\otimes 2}(P\otimes R)D^{\otimes 2}\right] &= \delta_{P,R}P\otimes R, && \text{(D26)}\end{aligned}$$

where $\delta_{P,R}=1$ *if* $P=R$ *and* 0 *otherwise.*

Proof. The proof follows immediately from Lemma 18, observing that all operators $\mathsf{X}\otimes Q$ and $Q\otimes\mathsf{X}$ satisfy one of the constraints $(\mathsf{I}),(\mathsf{S}),(\mathsf{P}),(\mathsf{X})$, while $P\otimes R$ for $P\neq R$ does not satisfy any. □

In the next step, we analyze the behaviour of the operators under CNOT operators.

Lemma 20 (CNOT update)**.** *Let $P, Q \in \mathcal{S}$. Then*

$$\mathrm{CNOT}^{\otimes 2}(P \otimes Q)\mathrm{CNOT}^{\otimes 2} = P \otimes \varepsilon_P(Q), \tag{D27}$$

where we define the permutation $\varepsilon_P : \mathcal{S} \to \mathcal{S}$ according to the following table

Q	X	I	S	P
$\varepsilon_{\mathsf{X}}(Q)$	X	I	S	P
$\varepsilon_{\mathsf{I}}(Q)$	I	X	P	S
$\varepsilon_{\mathsf{S}}(Q)$	S	P	X	I
$\varepsilon_{\mathsf{P}}(Q)$	P	S	I	X

.

Notice that for any input $Q \in \mathcal{S}_{\neg\mathsf{X}}$, the permutation ε_Q swaps Q with X as well as the other two states in $\mathcal{S}$, while $\epsilon_{\mathsf{X}}(Q) = Q$. In order to prove Lemma 20, and to show the behaviour under X noise, the following lemma will be convenient.

Lemma 21 (X updates)**.** *Let $P \in \mathcal{S}$. Then*

$$(X \otimes X)P(X \otimes X) = (\mathbb{1} \otimes \mathbb{1})P(X \otimes X) = \varepsilon_{\mathsf{X}}(P) \tag{D28}$$

$$(\mathbb{1} \otimes X)P(\mathbb{1} \otimes X) = (X \otimes \mathbb{1})P(X \otimes \mathbb{1}) = \varepsilon_{\mathsf{I}}(P) \tag{D29}$$

$$(X \otimes \mathbb{1})P(\mathbb{1} \otimes X) = (\mathbb{1} \otimes X)P(X \otimes \mathbb{1}) = \varepsilon_{\mathsf{S}}(P) \tag{D30}$$

$$(X \otimes X)P(\mathbb{1} \otimes \mathbb{1}) = (\mathbb{1} \otimes \mathbb{1})P(X \otimes X) = \varepsilon_{\mathsf{P}}(P) \tag{D31}$$

To prove the lemma, we just expand P in the computational basis.

Proof of Lemma 20. Given Lemma 21, the lemma follows immediately from expanding P in the computational basis and observing that every 1-entry in P results in an X applied to the corresponding entry of Q. For each P, the patterns o 1 entries in P, correspond exactly to one of the cases in Eqs. (D28) to (D31). For instance, for $P \otimes Q = \mathsf{I} \otimes Q$, we have

$$\begin{aligned}&\mathrm{CNOT}^{\otimes 2}(\mathsf{I} \otimes Q)\mathrm{CNOT}^{\otimes 2}\\ &= |01\rangle\langle 01|\,(\mathbb{1} \otimes X)P(\mathbb{1} \otimes X) + |10\rangle\langle 10|\,(X \otimes \mathbb{1})P(X \otimes \mathbb{1})\\ &= |01\rangle\langle 01|\,\varepsilon_{\mathsf{I}}(Q) + |10\rangle\langle 10|\,\varepsilon_{\mathsf{I}}(Q) = \mathsf{I} \otimes \varepsilon_{\mathsf{I}}(Q),\end{aligned} \tag{D32}$$

and likewise for the other cases. □

Finally, we need to analyze noise on one copy of the second moment operator. We do so for single-qubit X and Z noise The case of X noise follows immediately from Lemma 21. For Z noise we find

Lemma 22 (Z noise)**.** *We have that*

$$(Z \otimes \mathbb{1})P(Z \otimes \mathbb{1}) = \begin{cases} P & \text{if } P \in \{\mathsf{I}, \mathsf{X}\} \\ -P & \text{if } P \in \{\mathsf{S}, \mathsf{P}\}. \end{cases} \tag{D33}$$

The lemma follows from direct calculation in the computational basis.

One might wonder whether the states in $\mathcal{S}$ remain invariant also under Toffoli gates, since those allow implementing arbitrary permutations. However, this is not the case. To see this, consider the behaviour of the operator $\mathsf{X} \otimes P \otimes Q$ for any $P, Q \in \mathcal{S}$ under two-copy Toffoli gates

$$\begin{aligned}&\mathrm{TOF}^{\otimes 2}(\mathsf{X} \otimes P \otimes Q)\mathrm{TOF}^{\otimes 2} =\\ &\quad |00\rangle\langle 00| \otimes P \otimes Q + |11\rangle\langle 11| \otimes P \otimes \varepsilon_Q(R),\end{aligned} \tag{D34}$$

which cannot be decomposed into a linear combination of operators in $\mathcal{S}^{\otimes 3}$.

Given Lemmas 19 to 22, we can now identify every operator $\sigma \in \mathcal{S}^{\otimes n}$ with a classical state $|\sigma\rangle$. The second moment of a circuit with the above components can then be written as

$$|M_2\rangle = \sum_{\sigma \in \mathcal{S}^{\otimes n}} q(\sigma)\,|\sigma\rangle\,. \tag{D35}$$

Given the second moment $|M_2\rangle$ of a certain random circuit D in terms of the states $|S\rangle$ we can deduce the properties of the same circuit with a local gate or random IQP circuits applied to certain qubits (i, j) as an update $T^E_{ij}\,|M_2\rangle = \sum_S q'(S)\,|S\rangle$, where T^E_{ij} just acts on the tensor copies i and j. We can thus analyze the second moment properties of circuits as an evolution of the distribution q under the local update rules derived above. The general update rule under a single parallel circuit layer with label l is then given by

$$\mathcal{T}(\sigma', \sigma) = \prod_i N^{(i)}_{\sigma'_i, \sigma_i} \prod_{\langle i,j \rangle} T^{(ij)}_{\sigma'_i, \sigma'_j, \sigma_i, \sigma_j}, \tag{D36}$$

where the Z and X noise update matrices N^Z and $N^X \in \{0, 1\}^{4\times 4}$ are defined by the update rules (D33) and (D29), respectively, and the CNOT and random two-qubit IQP circuit update matrices $M^{\mathrm{CNOT}}, M^{\mathfrak{I}_2} \in \{0, 1\}^{16\times 16}$ by Lemmas 20 and 19, respectively.

In order to compute our properties of interest in Eqs. (D3) to (D6), we need to compute the overlap of the resulting state with the $\mathbb{X}_x$ and $\mathbb{S}$ states. Defining $\langle Q|P\rangle = \mathrm{tr}[QP]$, we observe that

$$\langle \mathbb{S}|Q\rangle = \begin{cases} 2 & \text{if } Q \in \{\mathsf{S}, \mathsf{X}\} \\ 0 & \text{if } Q \in \{\mathsf{I}, \mathsf{P}\} \end{cases} \tag{D37}$$

$$\langle \mathbb{X}_x|Q\rangle = 1 \quad \forall Q \in \mathcal{S}. \tag{D38}$$

Appendix E: Anticoncentration of degree-D circuits

Using the rules derived in this section, we can analyze the second-moment properties of arbitrary noiseless circuits with some random IQP degree-2 parts and X, CNOT gates, as well as their behaviour under single-qubit Pauli noise. To start out, let us apply the model

to compute some simple quantities of which we already know.

First, we compute the fidelity of the ideal state with itself, which of course evaluates to one. We begin with the initial state $|+^n\rangle$. We find

$$\mathrm{tr}\big[\mathbb{S}^{\otimes n} M_2(\mathbb{1},\mathbb{1})\big] = \frac{1}{4^n}\sum_{Q\in\mathcal{S}^{\otimes n}}\prod_{i=1}^{n}\mathrm{tr}[Q_i\mathbb{S}]. \tag{E1}$$

But $\mathrm{tr}[Q_i\mathbb{S}] = 0$ unless $Q_i = \mathsf{X}$ or $Q_i = \mathsf{S}$ by Eq. (D37). Hence, only all products of X and S are nonzero, and evaluate equally to 2^n. But there are exactly 2^n such terms and hence

$$\overline{F} = \frac{1}{4^n}\sum_{Q\in\mathcal{S}^{\otimes n}}\prod_{i=1}^{n}\mathrm{tr}[Q_i\mathbb{S}] = \frac{1}{4^n}\sum_{S\in\{\mathsf{X},\mathsf{S}\}^{\otimes n}} 2^n = 1. \tag{E2}$$

Since the states in $\{\mathsf{X},\mathsf{S}\}^{\otimes n}$ are invariant under the two-qubit IQP circuit, and swap and entangling gates just permute them, the fidelity remains unchanged under the action of a circuit.

Next, let us evaluate the XEB score. For XEB every term in $\mathcal{S}^{\otimes n}$ contributes equally with a value 1. For the state $|+^n\rangle$ we thus find

$$\begin{aligned}\overline{\chi} &= 2^n\,\mathrm{tr}\big[M_2(\mathbb{1},\mathbb{1})\mathsf{X}^{\otimes n}\big] - 1 && \text{(E3)}\\ &= \frac{2^n}{4^n}\sum_{Q\in\mathcal{S}^{\otimes n}}\prod_i \mathrm{tr}[Q_i\mathsf{X}] - 1 && \text{(E4)}\\ &= \frac{1}{2^n}4^n - 1 = 2^n - 1, && \text{(E5)}\end{aligned}$$

an exponentially large value.

Applying a uniformly random IQP circuit on the to $|+^n\rangle$, however, should reduces the XEB score to the value found in Corollary 17. To see how this score comes about in our model, we write a uniformly random n-qubit IQP circuit as $n(n-1)/2$ independent random two-qubit IQP circuits $D_{(i,j)}$ on every (unordered) pair of qubits (i,j). This is equivalent to an n-qubit random IQP circuit, since the Z-gates just commute to the end, yielding a uniformly random string. But every two-qubit random IQP circuit $D_{(i,j)}$ acts as a constraint on the states of the model at positions (i,j), namely, any state $S\in\mathcal{S}^{\otimes n}$ with $S_i, S_j \in \mathcal{S}_{\neg\mathsf{X}}$ and $S_i \neq S_j$ is annihilated by the constraint. We thus find

$$\begin{aligned}M_2(\mathbb{1},\mathbb{1}) &= \mathbb{E}_{\{D_{(i,j)}\}}[(\prod_{i,j} D_{(i,j)})\sum_{S\in\mathcal{S}^{\otimes n}} S(\prod_{i,j} D_{(i,j)})]\\ &= \sum_{S\in\{\mathsf{X},\mathsf{I}\}^{\otimes n}\cup\{\mathsf{X},\mathsf{S}\}^{\otimes n}\cup\{\mathsf{X},\mathsf{P}\}^{\otimes n}} S - 2\mathsf{X}^{\otimes n}\\ &= \mathbb{1}^{\otimes n} + \mathbb{S}^{\otimes n} + \mathbb{P}^{\otimes n} - 2\mathbb{X}_z^{\otimes n}, \end{aligned}\tag{E6}$$

which recovers the result of Lemma 19 since the sum over all elements in $\{\mathsf{X},\mathsf{S}\}^{\otimes n}$ equals $\mathbb{S}$.

1. Sparse degree-2 IQP circuits

Let us now consider the first nontrivial example, namely, the following circuit model. Consider a random IQP circuit with ℓ layers such that every layer is given by a random two-qubit IQP circuit $D_{(i,j)}$ on a random qubit pair i,j. We would now like to derive the anti-concentration properties of those circuits as a function of d.

Theorem 23 (Anticoncentration of sparse random IQP)**.** *Consider random IQP circuits on n qubits with uniformly random Z gates and random CZ gates acting on ℓ random pairs. Then*

$$\overline{\chi} \in 2 - 2^{-n+1} + 2^{-\Omega(\log(n))} \quad \text{if } \ell \geq \frac{11}{4}n\log(n) \tag{E7}$$

$$\overline{\chi} \in \Omega(2^n) \quad \text{if } \ell \in O(n). \tag{E8}$$

Proof. For $Q\in\mathcal{S}$, define the Q-Hamming weight of a string $S\in\mathcal{S}$ as

$$n_Q(S) = |\{i : S_i = Q\}|. \tag{E9}$$

We now observe that the action of a circuit layer on the states $|S\rangle$ depends only on the number of pairs $S_i \neq S_j$ such that $S_i, S_j \in \mathcal{S}_{\neg\mathsf{X}}$. Let us call a string S with the property of having k of such pairs a string with *PQ-weight* k, written as $\mathrm{pqw}(S) = k$. We have that

$$\begin{aligned}\mathrm{pqw}(S) &\equiv N(n_{\mathsf{I}}(S), n_{\mathsf{S}}(S), n_{\mathsf{P}}(S))\\ &:= n_{\mathsf{I}}(S)n_{\mathsf{S}}(S) + n_{\mathsf{I}}(S)n_{\mathsf{P}}(S) + n_{\mathsf{S}}(S)n_{\mathsf{P}}(S).\end{aligned}\tag{E10}$$

The probability that a state $S\in\mathcal{S}^{\otimes n}$ is annihilated in a single circuit layer is then given by $\mathrm{pqw}(S)/\binom{n}{2}$ since there are $\binom{n}{2}$ pairs in total.

Let us write the random circuit in layer i with a random degree-2 circuit in a random location as D_i. Let us also define the sets

$$W_n := \{(k,l,m)\in[n]^{\times 3} : k+l+m = n\} \tag{E11}$$

$$W_n^* := W_n \setminus \{(k,l,m) : k = l = 0 \vee k = m = 0 \vee l = m = 0\} \tag{E12}$$

$$\mathcal{S}_n^{(k,l,m)} := \{S\in\mathcal{S}^{\otimes n} : n_{\mathsf{I}}(S) = k, n_{\mathsf{S}}(S) = l, n_{\mathsf{P}}(S) = m\} \tag{E13}$$

We now write

$$\overline{\chi} + 1 = \frac{1}{2^n}\sum_{S\in\mathcal{S}}\mathbb{E}_{\{D_i\}}\,\mathrm{tr}\left[\left(\prod_{i=1}^{\ell} D_i\right) S \left(\prod_{i=1}^{\ell} D_i\right)\mathbb{X}_x\right] \tag{E14}$$

$$= \frac{1}{2^n}\sum_{S\in\mathcal{S}}\left(1 - \frac{\mathrm{pqw}(S)}{\binom{n}{2}}\right)^{\ell} \tag{E15}$$

$$= \frac{1}{2^n}\sum_{(k,l,m)\in W_n}|\mathcal{S}_n^{(k,l,m)}|\left(1 - \frac{N(k,l,m)}{\binom{n}{2}}\right)^{\ell}, \tag{E16}$$

and observe that the strings with generalized Hamming weight in W_n^* contribute exactly the ideal score of $3 - 2^{-n+1}$ to the XEB score so that

$$\overline{\chi} = 2 - \frac{2}{2^n} + \frac{1}{2^n} \sum_{(k,l,m)\in W_n^*} |\mathcal{S}_n^{(k,l,m)}| \left(1 - \frac{N(k,l,m)}{\binom{n}{2}}\right)^{\ell}. \tag{E17}$$

Let us now bound the terms in Eq. (E17). To this end, we observe that

$$|W_n^*| = n^3 - 3n \tag{E18}$$

$$|\mathcal{S}_n^{(k,l,m)}| = \binom{n}{k}\binom{n-k}{l}\binom{n-k-l}{m}. \tag{E19}$$

Using Stirling's formula for $k = pn$ we find the exact scaling of the binomial coefficients to be

$$\binom{n}{pn} \in \left(\frac{1}{2\pi np(1-p)}\right)^{\frac{1}{2}} 2^{nH(p)+\Theta(1/n)}, \tag{E20}$$

where $H(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function (log has base 2). We thus find that

$$\overline{\chi} - 2 + 2^{-n+1} \leq \frac{n^{7/2}}{(2\pi)^{3/2}} \times \tag{E21}$$
$$\times \max_{(k,l,m)\in W_n^*} 2^{-\left[n + \frac{\ell}{\binom{n}{2}} N(k,l,m) - n\left[H(\frac{k}{n}) + H(\frac{l}{n}) + H(\frac{m}{n})\right] - \Theta(\frac{1}{n})\right]}.$$

Now, since 2^{-x} is monotonously decreasing in x we want to find the minimum exponent

$$\min_{(k,l,m)\in W_n^*} \Bigg\{ n + \frac{\ell}{\binom{n}{2}} N(k,l,m) - n\left[H\left(\frac{k}{n}\right) + H\left(\frac{l}{n}\right) + H\left(\frac{m}{n}\right)\right] - \Theta\left(\frac{1}{n}\right) \Bigg\}. \tag{E22}$$

To show the claim, observe that the entropy function $H(p) \geq 1$ and attains this value at $p = 1/2$. This implies that $n[H(\epsilon) + H(\delta) + H(\theta)] \leq 3n$. More generally, let $k/n = \epsilon, l/n = \delta, m/n = \theta$ for arbitrary constants $\epsilon, \delta, \theta \in (0,1)$. For this choice, $2\ell N(k,l,m)/(n(n-1)) \geq 2\ell(\epsilon\delta + \epsilon\theta + \delta\theta)$. For sufficiently large ℓ we therefore have $2\ell N(\epsilon,\delta,\theta) \geq n[H(\epsilon) + H(\delta) + H(\theta)]$ for any choice of $\epsilon, \delta, \theta \in \Theta(1)$. One might think $\ell = Cn$ for a large enough constant C is sufficient, but observe that for any constant C, we can find constants ϵ, δ such that $2C\epsilon\delta < H(\epsilon) + H(\delta)$. Thus, we need to choose $\ell/n = f(n)$ to follow an arbitrary monotonously increasing function $f(n)$ that diverges with n. Otherwise, the corresponding terms in Eq. (E17) will diverge exponentially, showing the claim (E8).

The same argument holds if only two of k, l, m are chosen proportional to n, and the third one sublinear.

Let $k/n \leq 1/f(n)$ for an arbitrary montonously increasing function f that diverges with n. We have

$$\begin{aligned} H(k/n) &\leq \frac{1}{f(n)}\log f(n) - \left[1 - \frac{1}{f(n)}\right]\log\left[1 - \frac{1}{f(n)}\right] \\ &= -\log\left[1 - \frac{1}{f(n)}\right] + \frac{1}{f(n)}\log[f(n) - 1] \quad \text{(E23)} \\ &\leq \frac{1}{f(n)}\log[f(n)]. \end{aligned}$$

Consider $k, l, m \in o(n)$. In this case, by the bound (E23) the entropy-terms decrease as $nH(k/n) \leq no(1)$, which is asymptotically dominated by the linearly growing term.

Now, consider $k = \Theta(n), l = m = o(n)$—the only remaining case. Let us first fix a choice $k = n/2, l = m = 1$. In this case $H(k/n) = 1$, $H(l/n) = H(m/n) \leq \log(n)/n$ and $N(k,l,m) = n + 1$ so that by the bound (E23) the right hand side of Eq. (E21) is at most

$$2^{(7/2+2)\log n - 2\ell/n + \Theta(1/n))}. \tag{E24}$$

This bound decays for $\ell > (11/4) \cdot n\log n$.

We can vary the parameters around the bound by increasing or decreasing k, and increasing l, m. Let us fix $k = n/2$ and increase $l, m > 1$. In this case, $n(H(k/n) + H(l/n) + H(m/n)) \leq n + (l+m)\log(n)$ and $N(k,l,m) = (l+m)n/2 + lm$, in which case for $d > 11/4$ the bound (E21) decays. Next, let us decrease $k/n \leq 1/2 - \epsilon$ for constant $\epsilon > 0$. Since the entropy function is symmetric about $1/2$ and $N(k,l,m)$ increases in k, this is the worst case. Now, $H(k/n) \leq 1 - \epsilon^2$, and hence $n - nH(k/n) \geq n\epsilon^2$ which dominates all other terms. □

2. Coupled block-random IQP circuits

Let us now consider the block-random model of IQP in which we compose uniformly random degree-2 IQP circuits in blocks of m qubits with fixed entangling gates between the blocks. In the hIQP circuits, considered in the main text, there are $m = 3$ qubits per each of 2^D blocks, where D is the hypercube dimension. The blocks are then entangled by CNOT gates with random direction according to the hypercube connectivity. Crucially, while the direction of the CNOTs on every edge of the hypercube is random, $b = 3$ CNOT gates are applied *in parallel* on every edge, that is, all three gates have the same direction, and pair up with the same qubit in every circuit layer. More precisely, the block-random-IQP model is defined by a block size b, a connectivity graph $G = (V, E)$ with $b := |V|$ vertices, and an edge-coloring $C \subset \mathcal{P}(E)$, where $\mathcal{P}(E)$ is the power set of E. A *block-random IQP circuit with hyperdepth* $\mathfrak{L}$ *on* G then proceeds as follows

1. Prepare $|+^{mb}\rangle$.
2. In every hyperlayer $l = 1, \ldots, \mathfrak{L}$, for every color $c \in C$:

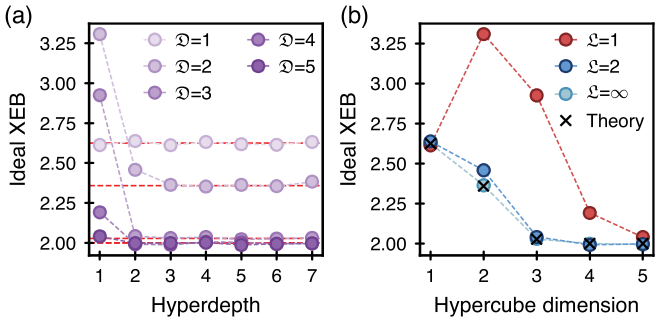

FIG. 13. **Ideal XEB values of random hIQP circuits without in-block CNOTs.** For each data point, we average $\sim 10^6$ random degree-2 hIQP circuits. (a) Convergence of the ideal XEB score of hIQP circuits without in-block CNOT gates to the asymptotic value (dashed lines) $\overline{\chi}_{\mathrm{hIQP}}(\mathfrak{D})$ given in Eq. (E25) as a function of hyperdepth for various $\mathfrak{D}$ (circles). (b) Convergence of the ideal XEB score of hIQP circuits to the uniform value as a function of hypercube dimension $\mathfrak{D}$ for various hyperdepths $\mathfrak{L}$ (circles).

i. apply a uniformly random degree-2 IQP circuit in every block.

ii apply (potentially randomized) transversal two-qubit gates between the blocks connected by the edges in c,

3. apply a uniformly random degree-2 IQP circuit in every block.

4. Measure in the X-basis.

In the hIQP circuits we also apply in-block uniformly random CNOT gates, but for the sake of completeness we will compute the properties of the block-random model first before turning to hIQP. We will consider different types of transversal two-qubit gates between blocks. The first type is what we call *randomized parallel CNOT* gates. A randomized parallel CNOT gate applied on an edge of the block-graph is given by $(\mathrm{CNOT}^{\otimes m})^d(\mathrm{NOTC}^{\otimes m})^{1-d}$ depending on a random choice of direction $d \in \{0,1\}$, where $\mathrm{NOTC} = \mathbb{S}\,\mathrm{CNOT}\,\mathbb{S}$. A *parallel randomized CNOT* gate, on the other hand, is specified by $d \in \{0,1\}^n$ and given by $\prod_{i=1}^n \mathrm{CNOT}^{d_i}\mathrm{NOTC}^{1-d_i}$. A *parallel randomized SWAP* gate, is given by $\prod_{i=1}^m \mathbb{S}^{d_i}$.

The following theorem gives the ideal XEB score for such circuits in the limit of a large number of circuit layers.

Theorem 24. *Let G be a connected graph on b vertices. The average XEB of block-random degree-2 IQP circuits acting on b m-qubit blocks connected by G and hyperdepth $\mathfrak{L}$ satisfies*

$$\overline{\chi} \xrightarrow{\mathfrak{L}\to\infty} \overline{\chi}_{\mathrm{hIQP}}(\mathfrak{D}) := 2 - 2^{-mb+1} + \frac{(2^m-1)(4^b - 3\cdot 2^b + 2)}{2^{mb}} \tag{E25}$$

if the two-qubit gates are randomized parallel CNOT gates, and

$$\overline{\chi} \xrightarrow{\mathfrak{L}\to\infty} 2 - 2^{-mb+1} + \frac{3(4^b - 3\cdot 2^b + 2)}{2^{mb}} \tag{E26}$$

if the two-qubit gates are parallel randomized CNOT or parallel randomized SWAP gates.

In either case, if a uniform CNOT circuit is applied to every block before the inter-block gates, we have

$$\overline{\chi} \xrightarrow{\mathfrak{L}\to\infty} 2 - 2^{-mb+1} \tag{E27}$$

We show the convergence to the asymptotic value for the case of parallel randomized CNOT gates in Fig. 13.

Proof. To show the theorem we proceed in two steps. First, we show that the limits are lower bounds. Second, we show that they are attained in the limit of high depth. Our key tool will again be the statistical-mechanics mapping described above.

Let us first consider the case of randomized parallel CNOT gates, and consider two strings $P, Q \in \mathcal{S}^m$ (we omit the tensor product in the following, understanding sp to be the string composed of $s, p \in \mathcal{S}$). Let $U = \mathrm{CNOT}^n$. Then it follows from Lemma 20 that under the action of U the string $PQ \mapsto PS$, where $S_i = Q_i$ iff $P_i = \mathsf{X}$ and $S_i \neq Q_i$ if $P_i \in \mathcal{S}_{\neg\mathsf{X}}$. Recall also that a string P is annihilated by a random degree-2 circuit if there are at least two indices i, j such that $P_i \neq P_j \in \mathcal{S}_{\neg\mathsf{X}}$, and remains if all $P_i \in \{\mathsf{X}, Q\}$ for a $Q \in \mathcal{S}_{\neg\mathsf{X}}$.

Let the X-*type of a state* $S \in \mathcal{S}^m$ be the locations $T_{\mathsf{X}}(S) := \{i \in [m] : S_i = \mathsf{X}\}$ at which it has an X state. Let the s-*type of a state* $S \in \{\mathsf{X},\mathsf{I}\}^m \cup \{\mathsf{X},\mathsf{S}\}^m \cup \{\mathsf{X},\mathsf{P}\}^m$ be the non-Xstring $T_s(S) \in \{\mathsf{I},\mathsf{S},\mathsf{P}\}$. We observe that the state $S = S^1S^2$ of two neighbouring blocks coupled by a randomized parallel CNOT gate will survive iff $T_s(S^1) = T_s(S^2)$ or $T_{\mathsf{X}}(S^1) = T_{\mathsf{X}}(S^2)$. Hence, the only configurations $S = S^1 \cdots S^b$ of block-states $S^i \in \mathcal{S}^n$ which are guaranteed to survive arbitrarily many circuit rounds are given by

1. states which are immortal under global random IQP circuits, that is, states $S \in \mathcal{S}_{\mathrm{imm}} := \{\mathsf{X},\mathsf{I}\}^{mb} \cup \{\mathsf{X},\mathsf{S}\}^{mb} \cup \{\mathsf{X},\mathsf{P}\}^{mb}$ which have the same s-type in every block, and

2. states $S \in \mathcal{S}_{t_{\mathsf{X}}} := \{Q : T_{\mathsf{X}}(Q^i) \in \{t_{\mathsf{X}}, [m]\}\,\forall i \in [b]\}$ with the same or trivial X-type on all blocks for some $t_{\mathsf{X}} \subset [m]$.

Let us now count those strings. We will do so by removing the *trivial string* (X^{mb}) in all calculations and adding it back at the end. We have

$$|\mathcal{S}_{\mathrm{imm}} \setminus \{\mathsf{X}^{mb}\}| = 3\cdot(2^{mb}-1) \tag{E28}$$

$$|\mathcal{S}_{t_{\mathsf{X}}} \setminus \{\mathsf{X}^{mb}\}| = 4^b - 1 \tag{E29}$$

$$\{t_{\mathsf{X}} \subset [m]\} \setminus \{[m]\} = 2^m - 1 \tag{E30}$$

$$\left(\mathcal{S}_{\mathrm{imm}} \cap \bigcup_{t_{\mathsf{X}}\subset[m]} (\mathcal{S}_{t_{\mathsf{X}}} \setminus \{\mathsf{X}^{mb}\})\right) = 3(2^b-1)(2^m-1). \tag{E31}$$

This implies that the number of surviving strings is given by at least

$$3\cdot(2^{mb}-1)+(4^b-1)(2^m-1)-3(2^b-1)(2^m-1)+1 \\ =(4^b-3\cdot 2^b+2)(2^m-1)+3\cdot 2^{mb}-2. \quad \text{(E32)}$$

Now we show that all strings which are not contained in $\mathcal{S}_{\text{surv}}=\mathcal{S}_{\text{imm}}\cup\bigcup_{t_{\mathsf{X}}\subset[m]}\mathcal{S}_{t_{\mathsf{X}}}$ will die under the circuit evolution. To see this, we take a string $Q\in\mathcal{S}_{\text{die}}=\mathcal{S}^{md}\setminus\mathcal{S}_{\text{surv}}$. By definition, there will be two substrings $Q^i,Q^j\in\mathcal{S}^n$ which have a different X-type, and contain two elements $Q^i_k\neq Q^j_l\in\mathcal{S}_{\neg\mathsf{X}}$ such that $k\in T_{\mathsf{X}}(Q^i), k\notin T_{\mathsf{X}}(Q^j)$, and vice versa for l. We need to show that this state of affairs cannot survive. Wlog. let $Q^i_k=\mathsf{I}, Q^j_l=\mathsf{S}$. First, we show that the circuit evolution of such a string will never leave $\mathcal{S}_{\text{die}}$. Consider a gate between Q^i and a neighbouring substring Q^o with a different X-type. If Q^i_k is the target of a CNOT-gate, which is controlled by X, it is unchanged. If it is the control of a CNOT-gate, it flips its target to I, changing the X-type of the target substring to a type with increased cardinality, and likewise for Q^j_l. At infinite depth, there will be a sequence of CNOT-gates in a circuit layer such that at then end of the sequence there will be a block with substring containing both I and S. Then, the string will die.

To show Eq. (E26), we observe that strings with the same *singleton* X*-types*, i.e., X-type with cardinality $|t_{\mathsf{X}}|=m-1$, are invariant under the circuit evolution, while all other strings will die by an analogous argument.

To show Eq. (E27), we observe that a uniform CNOT circuit is a uniform permutation on the set of X-types. Hence, for every string in which there are two blocks with distinct s-type, there will be a sequence of in-block CNOT gates and parallel CNOT gates such that at the end of the sequence there will be to neighbouring blocks i,j whose substrings $S^i,S^j\in\mathcal{S}^b$ have both different X-type and different s-type. Then the string will die. The only surviving strings are therefore immortal strings, and we have $|\mathcal{S}_{\text{imm}}|=3\cdot 2^{mb}-2$. □

Appendix F: Noisy IQP circuit dynamics

In this section, we study the XEB, fidelity and purity in noisy IQP circuits whose noiseless version is composed of unitary CPSP channels. The goal of this study is to identify conditions under which the XEB can be used as an estimator of fidelity.

We first consider the case of general noise models at low-noise rates and present results under increasing levels of assumptions about the noise and the random circuit ensemble. We then specialize to locally random degree-2 IQP circuits with SWAP gates and mid-circuit iid Pauli noise in order to understand the behaviour of the transversal circuits in color codes. Here, we can use the statistical mechanics mapping derived in Appendix D, in order to compute upper and lower bounds on the XEB. At very low noise rates, we find that XEB is a good fidelity proxy in noisy random IQP circuits, while at high noise rates XEB ceases to approximate fidelity, similar to the behavior observed recently in Haar-random circuits [16, 22, 36].

1. General noise at low error rates

To begin, we consider a general IQP circuit $D^{(\ell)}=D_dD_{\ell-1}\cdots D_1$ composed of ℓ layers of unitary CPSP channels with corresponding conjugate action $\mathcal{D}_i=D_i\cdot D_i^\dagger$ and $\mathcal{D}^{(\ell)}=D^{(\ell)}\cdot(D^{(\ell)})^\dagger$. We model the noisy circuit in terms of the noiseless circuit and a sequence of quantum instruments $\{\mathcal{E}_{i\mu_i}\}_{\mu_i}$ that implements a larger quantum instrument with elements

$$\mathcal{D}^{(\ell)}_\mu=\mathcal{E}_{\ell\mu_\ell}\circ\mathcal{D}_\ell\circ\cdots\circ\mathcal{E}_{1\mu_1}\circ\mathcal{D}_1\circ\mathcal{E}_{0\mu_0}, \quad \text{(F1)}$$

where $\mu=(\mu_0,\ldots,\mu_\ell)$ indexes the measurement outcomes in the quantum instruments. Writing as above $|x\rangle_z$ as the z-basis state with label $x\in\{0,1\}^n$ and $|x\rangle_x=H^{\otimes n}|x\rangle_z$ as the x-basis state with label x the fidelity F_μ and XEB χ_μ of the noisy circuit conditioned on syndrome outcomes μ are defined with respect to the noiseless circuit as

$$F_\mu=\frac{1}{p_\mu}\langle 0|_x(\mathcal{D}^{(\ell)})^\dagger\circ\mathcal{D}^{(\ell)}_\mu(|0\rangle_x\langle 0|_x)|0\rangle_x, \quad \text{(F2)}$$

$$\chi_\mu=\frac{2^n}{p_\mu}\sum_x|\langle x|_x D^{(\ell)}|0\rangle_x|^2\langle x|_x\mathcal{D}^{(\ell)}_\mu(|0\rangle_x\langle 0|_x)|x\rangle_x-1, \quad \text{(F3)}$$

where $p_\mu=\text{Tr}[\mathcal{D}^{(\ell)}_\mu(|0\rangle_x\langle 0|_x)]$ is the probability of measurement outcome $\mu=(\mu_0,\ldots,\mu_\ell)$.

We begin by considering a general model in which noise events are constrained to occur only at the beginning of the circuit, and a fully random, noiseless IQP circuit follows the noisy evolution. We think of this model as a proxy for very low noise rates.

Theorem 25. *Let $(\mathcal{E}_{i\mu_i})_{i=1,\ldots,\ell}$ be a sequence of quantum instruments such that $\mathcal{E}_{\ell\mu_\ell}$ is diagonal, $\mathcal{D}^{(\ell)}=\mathcal{D}_\ell\cdots\mathcal{D}_1$ a ℓ-layer sequence of unitary CPSP channels, $\mathcal{D}^{(\ell)}_\mu$ the associated quantum instrument for the noisy evolution, and D_ℓ a random degree-2 IQP circuit. Then*

$$\mathbb{E}_\ell[\chi_\mu]=2\mathbb{E}_\ell[F_\mu]-1/2^{n-1}, \quad \text{(F4)}$$

where $\mathbb{E}_\ell$ denotes the average over the random gates in D_ℓ.

Proof. We let μ be a fixed set of measurement outcomes. Since $\mathcal{E}_{\ell\mu_\ell}$ is diagonal (so it commutes with D_ℓ) and D_ℓ is real, we can write

$$\mathcal{D}^{(\ell)}_\mu(|0\rangle_x\langle 0|_x)=\sum_{xy}c_{x\mu}c^*_{y\mu}D_\ell|x\rangle_x\langle y|_xD_\ell, \quad \text{(F5)}$$

for a set of complex coefficients $c_{x\mu}$. We utilize the properties of phase states described in Appendix A to write

$$\langle x|_x D_t |y\rangle_x = \frac{1}{2^n}\sum_z (-1)^{f_t(z)+(x+y)\cdot z}, \tag{F6}$$

where $(\cdot)$ represents the usual bitstring scalar product and f_t is a degree-3 polynomial defined by the circuit in layer t.

We can then write a formula for F_μ and χ_μ as

$$\begin{aligned} F_\mu &= \sum_{xyz_i} \frac{c_{x\mu}c^*_{y\mu}}{p_\mu 4^n}(-1)^{F_\ell(z_1)+F_\ell(z_2)+z_1\cdot x+z_2\cdot y}, \\ \chi_\mu &= \sum_{x'xyz_i} \frac{c_{x\mu}c^*_{y\mu}}{p_\mu 8^n}(-1)^{F_\ell(z_1)+F_\ell(z_2)+z_3\cdot x+z_4\cdot y} \\ &\qquad \times (-1)^{\sum_i f_\ell(z_i)+x'\cdot z_i} - 1 \\ &= \sum_{xyz_i} \frac{c_{x\mu}c^*_{y\mu}}{p_\mu 4^n}(-1)^{F_\ell(z_1)+F_\ell(z_2)+x\cdot z_3+y\cdot z_4} \\ &\qquad \times (-1)^{\sum_i f_\ell(z_i)}\delta_{z_4,z_1+z_2+z_3} - 1, \end{aligned}$$

where $p_\mu = \sum_x |c_{x\mu}|^2$, $F_t(z) = \sum_{i=1}^{t-1} f_i(z)$ is the cumulative polynomial of the noiseless circuit from time 0 to $t-1$, and in the last line we used the sum over x' to impose the constraint $z_1+z_2+z_3+z_4 = 0 \mod 2$. When averaging over the last layer of gates, we notice that fidelity is independent of this averaging, while XEB requires the evaluation of

$$\mathbb{E}_\ell\left[(-1)^{\sum_i f_\ell(z_i)}\delta_{z_4,z_1+z_2+z_3}\right]. \tag{F7}$$

To evaluate the circuit-average $\mathbb{E}_\ell$ over the degree-2 part of f_ℓ, we follow the same arguments as in the proof of Lemma 16 (following Bremner *et al.* [24]), yielding the constraints $(\mathbb{1}), (\mathbb{S}), (\mathbb{P})$ in Eq. (D11) with overlap $(\mathbb{X}_z)$ in Eq. (D13) on the bit string quadruples (z_1, z_2, z_3, z_4).

Using these simplifications of the allowed quadruples, we arrive at a simplified formula for the conditional XEB

$$\begin{aligned} \mathbb{E}_\ell[\chi_\mu] &= 2\sum_{xyz_i} \frac{c_{x\mu}c^*_{y\mu}}{4^n p_\mu}(-1)^{F_\ell(z_1)+F_\ell(z_2)+x\cdot z_1+y\cdot z_2} \\ &\quad + \sum_{xyz} \frac{c_{x\mu}c^*_{y\mu}}{2^n p_\mu}(-1)^{x\cdot z+y\cdot z} - 1 \\ &\quad - \frac{1}{2^{n-1}}\sum_{xyz} \frac{c_{x\mu}c^*_{y\mu}}{2^n p_\mu}(-1)^{x\cdot z+y\cdot z} \\ &= 2F_\mu - \frac{1}{2^{n-1}}. \end{aligned}$$

This identity completes the proof because $\mathbb{E}_\ell[F_\mu] = F_\mu$. □

Theorem 25 establishes the equivalence between fidelity and XEB under a broad class of noise models. Crucially, it also allows for noise models that are not quantum channels, but conditioned on measurement outcomes. Such noise arises at the level of logical qubits when studying the fidelity conditioned on a particular syndrome outcome, as is done in error detection. However, the theorem makes the crucial assumption that the measurement error channel is diagonal and that the final layer of gates is a uniformly random degree-2 IQP circuit. In Theorem 23, we saw that using long-range connectivity, it is possible to achieve such an ensemble in depth $\log n$.

To see why this assumption is necessary, we now show that choosing the noise adversarially breaks the correspondence between fidelity and XEB.

Theorem 26. *Consider a noisy IQP circuit $\mathcal{E}_\ell \circ \mathcal{D}_\ell \circ \cdots \circ \mathcal{E}_1 \circ \mathcal{D}_1 \circ \mathcal{E}_0$ with $\mathcal{E}_\ell$ the bit-flip channel on the first qubit $\mathcal{E}_\ell(\rho) = (1-p)\rho + pX_1\rho X_1$ with $0 < p \le 1/2$, D_ℓ a random degree-2 IQP circuit, and $\mathcal{E}_i = \mathbb{1}$, $D_i = \mathbb{1}$for $i < \ell$. For this noisy IQP circuit, we have the identities $\mathbb{E}_\ell[\chi] = 1$ and $\mathbb{E}_\ell[F] = (1-p) + p/2^n$.*

Proof. The x-errors in the final layer have no effect on the XEB, which is why $\mathbb{E}_\ell[\chi_\mu] = 1$ in this noise model (i.e., the average XEB is equal to its noiseless value). For the fidelity, we can make use of the formula

$$F = (1-p) + \frac{p}{4^n}\sum_{z_1,z_2}(-1)^{f_\ell(z_1)+f_\ell(z_1+b)+f_\ell(z_2)+f_\ell(z_2+b)}, \tag{F8}$$

where $b = (1, 0, \ldots, 0)$ is a shift arising from the single-site X operator. Averaging the argument in the sum over the random degree-2 IQP circuit ensemble, for it to be non-zero we arrive at the constraint that $z_1 = z_2$ or $b = 0$. Since $b \neq 0$, we only get a contribution from the 2^n terms in the sum for which $z_1 = z_2$, which completes the proof. □

This breakdown of the correspondence between fidelity and XEB for the adversarial noise is reminiscent of what happens in noisy Haar random circuits [36].

We now show that it is sufficient to consider the z-twirl of the noise channels when comparing the average fidelity and XEB.

Theorem 27. *Consider a noisy IQP circuit with diagonal gate layers $D_i = Z^{a_i}\tilde{D}_i$ containing independent Pauli-Z gates $Z^{a_i} = \prod_j Z_j^{a_{i,j}}$ with uniformly chosen $a_{i,j} \leftarrow \{0,1\}$, then the circuit averaged fidelity and circuit averaged XEB are equal to their values with $\mathcal{E}_{i\mu_i}$ replaced by its z-twirl $\mathcal{E}_{zi\mu_i}$ for $0 < i < \ell$.*

Proof. The proof is analogous to the arguments that show one can twirl the noise into a depolarizing channel in Haar random circuits or Clifford randomized benchmarking. For each CP map $\mathcal{E}_{i\mu_i}$ we take a Kraus decomposition with Kraus operators $K_{i\mu_i\nu_i}$. Define the new random variables

$$b_t = a_1 + \cdots + a_t. \tag{F9}$$

Now, we can write the noisy state as

$$\mathcal{D}_\mu^{(\ell)}(|0\rangle_x \langle 0|_x) = \sum_\nu K_{\mu\nu} |0\rangle_x \langle 0|_x K_{\mu\nu}^\dagger, \tag{F10}$$

$$K_{\mu\nu} = K_{\ell\mu_\ell\nu_\ell} Z^{b_\ell} \tilde{D}_\ell \times \left[\prod_{i=1}^{\ell-1} Z^{b_{\ell-i}} K_{\ell-i\mu_{\ell-i}\nu_{\ell-i}} Z^{b_{\ell-i}} \tilde{D}_{\ell-i}\right] K_{0\mu_0\nu_0}, \tag{F11}$$

where $\nu = (\nu_0, \ldots, \nu_\ell)$ indexes the Kraus operators. The random variables b_i are independent uniformly random bit strings. When we average the noisy state over b_i for $i < \ell$ we can therefore apply the twirled channel. In the fidelity and XEB, the noiseless part of the circuit depends only on the random variable b_ℓ. As a result, we can average over b_i for $i < \ell$ independently of what appears in the noiseless circuit. This simplification allows us to replace $\mathcal{E}_{i\mu_i}$ by its z-twirl for all $0 < i < \ell$ when evaluating average XEB and average fidelity. □

The proof can be readily extended to the case where the D_i are unitary CPSP channels and not just diagonal gates by including random Z gates at both the end and beginning of each layer.

Recall that the z-twirl of a channel has Kraus operators of the form

$$X^b K_{\nu b},\ K_{\nu b} = \sum_a c_{ab}^\nu Z^a,$$

where a and b run over bitstrings of length n and c_{ab}^ν are some complex coefficients. Interestingly, for any mixture of phase states ρ the probability $p_{\mu b}$ that a particular bit-flip error b occurr is independent of the initial state

$$p_{\mu b} = \sum_\nu \mathrm{Tr}[K_{\mu\nu b}^\dagger K_{\mu\nu b}\rho] = \sum_{a,\nu} |c_{ab}^{\mu\nu}|^2. \tag{F12}$$

This identity holds because all phase states have zero expectation value of all Z-type operators. When the state-preparation errors are given by a PSP channel, we can thus think of bit flip errors of the z-twirled channel as occurring in the circuit at a rate that is independent of the current quantum state in the circuit. As a result, it is natural to consider noise models satisfying the following independence property.

Definition 28. *A z-twirled CP map $\mathcal{E}_\mu$ has independent x-noise if the probability of a bit-flip error $p_{\mu b}$ on phase states satisfies*

$$p_{\mu b} = (1 - p_\mu)^{n-|b|} p_\mu^{|b|}, \tag{F13}$$

for some parameter $0 \le p_\mu \le 1$, where $|b|$ is the Hamming weight of b (i.e., the number of nonzero entries in b).

The parameter p_μ thus quantifies the effective local bit-flip rate of $\mathcal{E}_\mu$ applied to phase states.

We now make some observations about noisy IQP circuits with z-twirled noise with independent x-noise bit-flip rates:

- When the ensemble of D converges to random degree-2 IQP circuits and the local bit-flip rate in the μ-th layer $p_\mu = 0$ for $\mu > 0$, then $\mathbb{E}[\chi] = 2\mathbb{E}[F] - 1/2^{n-1}$. This follows from the fact that without bit-flip noise the noise channels commute with D. As a result, we can bring the full channel into a form satisfying the conditions of Theorem 25.
- Take bit-flip noise rates $p_i = c/n \log n$ for some $c > 0$ and a noiseless circuit ensemble with ideal XEB value bounded by $2 - 1/2^{n-1} + o(1)$ in depth $\Omega(\log n)$, e.g., the one of Theorem 23. For sufficiently small c, with probability $1 - O(c)$ over noise realizations, the errors occur in the first layers of the circuit, with $\Omega(\log n)$ noiseless layers before the measurement. As a result, for sufficiently small local bit-flip noise rates $c/n \log n$, we can follow the proof of Theorem 25 and find that $\mathbb{E}[\chi] = 2\mathbb{E}[F] - 1/2^{n-1} + O(c) + o(1)$ for depths $\ell \in \Omega(\log n)$.

This second result is analogous to the white-noise approximation in noisy Haar random circuits established rigorously by Dalzell *et al.* [40]. It implies that at sufficiently low-noise rates, there is an exact correspondence between fidelity and XEB. Moreover, since fidelity will typically decay exponentially in the circuit size, the XEB will exhibit a similar decay at these noise rates. A central question is how this behavior changes as the noise rate increases to larger values c/n for increasing $c > 0$. In noisy Haar random circuits, there is a phase transition in XEB where it goes from a decay linear in the circuit volume to one that depends on the depth ℓ and not n [16, 22, 36].

2. Upper and lower bounds on XEB

To more directly analyze the behavior of XEB in noisy IQP circuits at large noise rates, we specialize in this subsection to the random degree-2 circuit model with SWAP gates. Specifically, we consider circuits in which in every layer of the circuit, we first apply a 2-local random degree-2 IQP circuit to a random pair of qubits, and then apply the noise channel to all qubits. Let us call this model the *local random degree-2 IQP model*.

To organize the calculations it is helpful to represent the initial state as

$$\frac{1}{4^n}(\mathsf{I} + \mathsf{S} + \mathsf{P} + \mathsf{X})^n = \frac{1}{4^n}\sum_k \binom{n}{k} (\mathsf{I} + \mathsf{X})^{n-k} (\mathsf{S} + \mathsf{P})^k, \tag{F14}$$

where we have removed the spatial dependence of the strings due to the permutation symmetry of the initial state and our circuits, and we write SP as shorthand for $S \otimes P$ for states $S, P \in \{\mathsf{I}, \mathsf{S}, \mathsf{P}, \mathsf{X}\}$. We consider the noise to be an X-Y-*symmetric* Pauli channel, i.e., a Pauli channel with symmetric X/Y noise rates $q_\perp/2 = p_x =$

p_y. It is convenient think of the total Pauli noise channel as the action of a pure Z noise channel with rate $p_z/(1-q_\perp)$ followed by a symmetric Pauli-X/Y noise channel with X/Y-noise rates $q_\perp/2$. Defining $q = q_\perp + 2p_z$, the combined noise channel will act on S/P states as

$$\mathsf{S}/\mathsf{P} \to (1-q)\mathsf{S}/\mathsf{P}, \tag{F15}$$
$$\mathsf{I}/\mathsf{X} \to (1-q_\perp)\mathsf{I}/\mathsf{X} + q_\perp \mathsf{X}/\mathsf{I}, \tag{F16}$$

i.e., it leads to a pure damping of the string depending on the S and P weight, while it flips each X and I component of the string at rate $q_\perp$. We can commute all the z noise in the circuit to the initial state, which adds a decay term

$$(\mathsf{S}+\mathsf{P})^k \to (1-q)^{\ell k}(\mathsf{S}+\mathsf{P})^k. \tag{F17}$$

As a result, the only non-trivial effect of the noise in the statistical mechanics model is from the X/Y-noise acting on the I/X operators of the strings. Solving for the full dynamics of this model is not obviously analytically tractable. Instead we prove an upper and lower bound on $\overline{\chi}$ that illustrate the two regimes of XEB and are tight with respect to the scaling with depth.

Theorem 29. *The average XEB for the local random degree-2 IQP model on n qubits with ℓ gates and ℓ layers of X-Y-symmetric Pauli noise is bounded as*

$$\overline{\chi} \geq 2^{1-n}[(1-q_\perp)^\ell + (1-q)^\ell]^n \tag{F18}$$
$$+ 2^{-2\ell/n}[(1-q)(1-q_\perp)]^\ell, \tag{F19}$$
$$\overline{\chi} \leq 2^{1-n}[1+(1-q)^\ell]^n + p(n)2^{-2\ell/n}, \tag{F20}$$

where $p(n) = n^{7/2}/(2\pi)^{3/2}$.

Proof. We start by proving the lower bound. Recall that the initial state is a uniform superposition over all $\mathsf{I}, \mathsf{S}, \mathsf{P}, \mathsf{X}$ strings and XEB is computed by summing up the contributions of each of these strings independent of their composition. As a result, we can lower bound the XEB by neglecting the transitions from X noise that change the string composition. The lower bound then follows by computing the contribution to XEB from the strings

$$(\mathsf{I}+\mathsf{X})^n + (\mathsf{S}+\mathsf{X})^n + (\mathsf{P}+\mathsf{X})^n - 2\mathsf{X}^n + (\mathsf{I}+\mathsf{X})^{n-2}\mathsf{IS}. \tag{F21}$$

The total XEB amplitude of the strings $(\mathsf{S}+\mathsf{X})^n$ and $(\mathsf{P}+\mathsf{X})^n$ decays as $2^{-n}[(1-q_\perp)^\ell + (1-q)^\ell]^n$. On the other hand, the total weight of the string with one S operator decays at least as $2^{-2\ell/n}[(1-q)(1-q_\perp)]^\ell$. To see this, observe that the state $(\mathsf{I}+\mathsf{X})$ is an eigenstate of the X decay, and that the PQ-weight of the string $P = \mathsf{SI}^{n-1}$ is given by $\mathrm{pqw}(P) = n-1$.

To compute the XEB contribution $2^{-n}\operatorname{tr}[\mathbb{X}_z S]$ of $S = (\mathsf{I}+\mathsf{S})^{n-2}\mathsf{IS}$, we split the contributions from the noise-induced decay and the circuit-induced decay. In total, S consists of 2^{n-2} individual strings. Following the argument above, the noise-induced decay given by $[(1-q_\perp)(1-q)]^\ell$. The circuit-induced decay is given by the last term in Eq. (E17), $(1-\mathrm{pqw}(S)/\binom{n}{2})^\ell \geq 2^{-2\ell/n-2}$ which follows from $(1-2/n)^\ell \xrightarrow{n\to\infty} e^{-2\ell/n}$.

To compute the upper bound we recall the decomposition of the initial state in Eq. (F14). The X noise leaves the $\mathsf{I}+\mathsf{X}$ components unaffected, while a gate can transform $(\mathsf{I}+\mathsf{X})\mathsf{S} \to \mathsf{XS}$, i.e., it removes the X eigenstate and replaces it by the X state. We now observe that X noise acting on this site can only increase the PQ-weight of this string, which increases the decay rate of the string. As a result, let us change the model such that gates acting on $\mathsf{I}(\mathsf{S}/\mathsf{P})$ pairs do not lead to a decay, while maintaining the rule that gates acting on $(\mathsf{I}+\mathsf{X})(\mathsf{S}/\mathsf{P})$ pairs still send it to $\mathsf{X}(\mathsf{S}/\mathsf{P})$. This change in the model will only increase the XEB so that we can upper bound the XEB using it. The model is then particularly simple to solve because the only effect of the noise is to lead to the decay of S/P components at rate $(1-q)$. To derive the upper bound we then apply this decay process to the strings

$$(\mathsf{I}+\mathsf{X})^n + (\mathsf{S}+\mathsf{X})^n + (\mathsf{P}+\mathsf{X})^n. \tag{F22}$$

We can upper-bound the decay of the contribution from the other strings by following the calculation in Theorem 23. □

The lower bound on XEB shows directly that the XEB has two regimes depending on the strength of the noise. When $(p_x+p_y+p_z) < 2\log 2/n^2$, then the XEB is dominated by the exponential decay that is linear in the circuit volume $n\ell$. Whereas for larger noise rates, the XEB decay is dominated by the decay that is linear in the effective depth ℓ/n (recall that one layer of gates requires $\ell = n/2$ in this model). This sharp change in behavior of the XEB is exactly analogous to what has been found in noisy Haar random circuits [16, 22, 36]. In both cases, it indicates that the XEB ceases to be good proxy for the fidelity, which always decays at a rate linear in the circuit volume until it reaches a value near $1/2^n$. The upper bound in the XEB shows that the two scaling regimes observed in the lower bound are actually tight at low and large enough noise rates.

3. Adding fixed gates

Let us now show that the addition of one fixed gate layer to the circuit can shift the location of the transition in the scaling of the XEB decay by a factor of 3/2 in the depth.

Lemma 30. *The average XEB for the local random degree-2 IQP model on an even number n of qubits with ℓ gates and ℓ layers of X-Y-symmetric Pauli noise with a last layer of parallel CZ gates between neighbouring qubits*

is bounded as

$$\overline{\chi} \geq 2^{1-n}[(1-q_\perp)^\ell + (1-q)^\ell]^n \tag{F23}$$

$$+ 2^{-4\ell/n}[(1-q)(1-q_\perp)]^{2\ell}. \tag{F24}$$

$$\overline{\chi} \leq 2^{1-n}[1+(1-q)^\ell]^n + p(n)2^{-3\ell/n} \tag{F25}$$

Furthermore, the total weight of all strings with a single excitation, i.e., strings of the form $S \in \mathcal{S}_{\neg\mathsf{X}} \times \mathcal{S}_{imm}$, is given by 0.

Proof. Let us first show the final statement. Wlog. consider a string $S = \mathsf{ISI}^{n-2}$ and a CZ gate acting on the first two qubits. Then $S \overset{CZ}{\longmapsto} -S$, and hence the weight of the string is -1. But there is another string with the same weight given by $S' = \mathsf{IXI}^{n-2}$. This string evolves as $S' \overset{CZ}{\longmapsto} S'$ under the last circuit layer and hence contributes with $+1$ so their contributions cancel.

The lower bound follows from the argument in the proof of Theorem 29 but now applied to the strings $\mathsf{PPI}(\mathsf{I}+\mathsf{X})^{n-3}$. These strings are invariant under the final CZ layer since $\mathsf{PP} \overset{CZ}{\longmapsto} \mathsf{PP}$. Their maximum PQ-weight is given by $2(n-2)$ and hence their circuit-induced decay is lower-bounded by $(1-4(n-2)/n(n-1)^\ell \geq 2^{-4\ell/n-2}$. Meanwhile, their noise-induced decay rate doubles.

For the upper bound, we argue that the second term in the upper bound (F20) in the proof of Theorem 1, $p(n)2^{-2\ell/n}$ decreases by adding a last layer of parallel CZ gates. We consider strings with $n_\mathsf{I}(S) = 1, n_\mathsf{S}(S) = 1, n_\mathsf{P}(S) = n/2$ of the form $S = \mathsf{ISP}^{n/2}\mathsf{X}^{n/2-2}$, letting $n/2$ be even. Then $S \overset{CZ}{\longmapsto} -S$, and there is another string $S' = \mathsf{IXP}^{n/2}\mathsf{SX}^{n/2-3}$ obtained from S by a single swap of operators between the non-X part of the string and the X-part of the string, for which $S' \overset{CZ}{\longmapsto} S'$. Hence, the contributions of all strings of this type cancel out, leaving us with strings S with $(n_\mathsf{I}(S), n_\mathsf{S}(S), n_\mathsf{P}(S)) = (2, 1, n/2)$ giving total PQ-weight $\mathrm{pqw}(S) = 3n/2 + 2$ saturating the upper bound. Using these strings, the second term in Eq. (E22) increases to $\sim 3\ell/n$ and hence the upper bound follows, shifting the transition by a factor of 3/2. □

It is clear that the argument can be repeated any number of times, applying CZ gates on new sets of pairs. Eventually, as we apply it n times, we retrieve the volume-like decay of Theorem 25.

Appendix G: Hardness of degree-4 Bell sampling

Let us in this section analyze in more detail the output distribution of uniform degree-D IQP circuits measured in the Bell basis. We will argue that sampling from these circuits is classically intractable unless the polynomial hierarchy collapses. The argument will be somewhat more intricate compared to sampling in the standard-basis.

Recall from the main text that we can write the output probability of a Bell measurement on a degree-D circuit C as

$$P_C(x,z) = \frac{1}{2^n} |\langle +^n| CX_xZ_zC |+^n\rangle|^2 \tag{G1}$$

$$= \frac{1}{2^n} |\langle +^n| C_xZ_z |+^n\rangle|^2, \tag{G2}$$

where $C_x = CX_xC$ is the degree-$D-1$ circuit induced by the X outcomes and the circuit C. Denote by $\mathrm{supp}(x) := \{i \in [n] : x_i = 1\}$ the support of the string x. Then the support of the circuit C_x, i.e., the qubits on which it acts nontrivially, is given by $\mathrm{supp}(C_x) = \mathrm{supp}(x)^c$, since the one-locations of x remove qubits from the support of C.

Let us observe a few properties of this output distribution.

Lemma 31. *We have the following facts about the output probabilities $P_C(x,z)$ of Bell sampling from degree-D IQP circuits.*

i. For all (x,z) such that there is an $i \in [n]$ with $x_i = z_i = 1$, $P_C(x,z) = 0$.

ii. The marginal probability $P_C(x) = \sum_z P_C(x,z) = 2^{-n}$.

Proof. To see claim (i), observe that the zero-entries of x determine $\mathrm{supp}(C_x)$. So for all $i \in [n]$ for which $x_i = 1$, C_x acts trivially, and hence, $\langle +^n| C_xZ_i |+^n\rangle = \langle +| Z |+\rangle \langle +^{n-1}| (C_x)|_{[n]\setminus\{i\}} |+^{n-1}\rangle = 0$.

To see claim (ii), we simply compute $P_C(x) = \sum_z P_C(x,z) = 2^{-n}\sum_z |\langle z| H^{\otimes n}C_xH^{\otimes n} |0^n\rangle|^2 = 2^{-n}$ by the normalization of the output distribution of the degree-$D-1$ circuit C_x. □

The output distribution of degree-D Bell sampling therefore 'shatters' into separate sectors for each fixed outcome x of the X measurement. All of those sectors are equally likely, and the output distribution within each sector is given by the output distribution of the induced IQP circuit C_x acting nontrivially on $n - |x|$ qubits.

Let us now consider a uniformly random degree-D circuit C. Then for every x, C_x is a uniformly random degree-$(D-1)$ circuit acting on $\mathrm{supp}(C_x) = \mathrm{supp}(x)^c$. To see this in each x-sector, fix the indices of the gates in C corresponding to the $x_i = 1$ outcomes and observe that the marginal distribution on the remaining indices remains uniform. This motivates grouping the x-sectors further according to the size of their support, since the statistics of the output distribution in each sector are now solely determined by the size of the uniformly random IQP circuit in that sector. Hence, there are $n+1$ distinct sectors $S_k := \{x : |x| = n-k, z \in \{0,1\}^k\} \subset \{0,1\}^{2n}$ oon which P_C has nontrivial support, and the statistics of the outcomes in each sector are the same.

This state of affairs is reminiscent of that in boson sampling. For instance, in Gaussian boson sampling the outcomes fall into distinct photon number sectors [109, 110], and in boson sampling with a linear number of modes there are sectors corresponding to the collision patterns of the outputs [112]. This analogy will also guide us in

the following argument for hardness of sampling from degree-D Bell sampling with $D \geq 4$.

Theorem 32 (Hardness of degree-D Bell sampling)**.** *Sampling from the Bell sampling output distribution $P_C : \{0,1\}^{2n} \to [0,1]$ from uniformly random degree-D IQP circuits for $D \geq 4$ up to constant total-variation distance is classically intractable unless the polynomial hierarchy collapses and the approximate average-case hardness conjecture 33 is false.*

Conjecture 33 ([24])**.** *Approximating the output probabilities of uniformly random degree-D IQP circuits for $D \geq 3$ up to relative error $1/4$ is #P-hard.*

Since the output distribution of random degree-D Bell sampling shatters into $n+1$ sectors, we need to consider those sectors individually. Observe, for instance, that the output probabilities of the $k = 1$ sector are very easy to compute. However, that sector has negligible probability weight as the following lemma shows.

Lemma 34. *The probability weight of P_C outside of the sectors S_k with $k \in [n/2 - O(\sqrt{n}), n/2 + O(\sqrt{n})]$ is negligible.*

Proof. We use (ii) of Lemma 31 to see that

$$P_C(S_k) = \binom{n}{k}/2^n, \tag{G3}$$

and letting $p(k,n) = \sqrt{2\pi(n/2+k)(1/2-k/n)}$

$$P_C(S_{n/2+k}) = \binom{n}{n/2+k}/2^n \tag{G4}$$

$$= \frac{1}{2^n p(k,n)} 2^{nH(1/2+k/n)+\Theta(1/n)} \tag{G5}$$

$$= \frac{1}{p(k,n)} 2^{-2k^2/(n\ln 2)+\Theta(1/n)}, \tag{G6}$$

since $H(1/2+\epsilon) = 1 - 2\epsilon^2/\ln 2 + O(\epsilon^3)$. Therefore for $k \in n/2 + \omega(\sqrt{n})$, we have $p(k,n)P_C(S_{n/2+k}) \in o(1)$ and we have

$$\sum_{k \in n/2 \pm \tilde{\omega}(\sqrt{n})} P_C(S_k) \in o(1). \tag{G7}$$

Note that we have chosen $k \in n/2 + \tilde{\omega}(\sqrt{n})$ since the sum adds a factor of $n - O(\sqrt{n})$ which needs to be accounted for with a $\log \sqrt{n}$ shift in the exponent of Eq. (G6). □

Lemma 34 implies that the dominant contributions to the output distribution are the sectors $k \in n/2 \pm \tilde{O}(\sqrt{n}) \subset \Theta(n)$. For each of these sectors approximating the output probabilities is a #P-hard task in the worst case since the respective degree-$D-1$ circuits on k qubits are arbitrary [24]. Likewise, we inherit the approximate-average case hardness conjecture 33 from Bremner *et al.* [24].

Next, we show the hiding property in each sector. The hiding property is an important ingredient in the reduction from computing probabilities to sampling. The hiding property asserts that the distribution over input circuits C is invariant under a 'hiding procedure'. This procedure removes the dependency on a particular outcome x by hiding that outcome in the probability of obtaining a different outcome y of a different random circuit C_y. This allows us to restrict our attention to the distribution over circuits of a fixed outcome.

Lemma 35 (Hiding)**.** *For each $k \in [n]$, the Bell output distribution $P_C \restriction_{S_k}$ of a uniform degree-D circuit C satisfies the hiding property.*

Proof. First, we observe that the hiding property trivially holds with respect to the z outcomes in every k-sector. We can translate one z outcome to a different outcome z' by applying $Z_{z'+z}$ gates to the IQP circuit. But this leaves the distribution over circuits invariant since Z gates are applied with uniform probability. Next, consider the x-outcomes in every sector. These are all related by a permutation. Therefore, we can obtain an outcome (x,z) of a degree-D circuit C by from an (x',z') outcome of a circuit $C(x',z')$ which is obtained from C by permuting the labels of the gates according to a permutation Π such that $x' = \Pi x$, and then applying $Z_{z'+z}$. □

Finally, we consider anticoncentration of the Bell sampling distribution from uniform degree-D IQP circuits with $D \geq 3$. Again, we find that, restricted to each k-sector, anticoncentration holds. To this end, we compare the squared first moment (given by the squared uniform distribution) in every sector with the second moment in the sector.

The correct uniform distribution on the sample space is given by sampling a uniformly random $x \in \{0,1\}^n$ and then a uniformly random $z \in \{0,1\}^k$. This translates into the uniform distribution in every k-sector as

$$\mathcal{U}_k = P_C(S_k) \cdot \frac{1}{|S_k|} = \frac{\binom{n}{k}}{2^n} \frac{1}{2^k \binom{n}{k}} = 2^{-nk}. \tag{G8}$$

Lemma 36 (Anticoncentration of degree-D Bell sampling)**.** *The second moment of the outcome probabilities of uniform degree-D Bell sampling for $D \geq 3$ with $(x,z) \in S_k$ satisfies*

$$\mathcal{U}_k^{-2} \cdot \mathbb{E}_C P_C(x,z)^2 = 3 - 2^{-k+1}, \tag{G9}$$

Proof. Let uIQP(k,D) be the family of uniform degree-D circuits on k qubits. Furthermore let $z_x \in \{0,1\}^k$ be the restriction of z to supp(x).

$$\mathbb{E}_C P_C(x,z)^2 = \frac{1}{2^{2n}} \mathbb{E}_{C \sim \mathrm{uIQP}(k,D-1)} |\langle +^n| C |z_x\rangle|^4 \tag{G10}$$

$$= \frac{1}{2^{2n}} \frac{3 - 2^{-k+1}}{2^{2k}}. \tag{G11}$$

□

Proof of Theorem 32. To prove the theorem, we follow the route for hardness arguments of Gaussian boson sampling, see e.g. [110]. We apply the argument due to Stockmeyer individually in every sector S_k. In each sector, the hiding argument holds, as well as anticoncentration for the postselected distribution. Hence, we can follow argue for hardness of sampling for the distributions in each sector, see Ref. [19] for details. For the outcomes with $k \in n/2 + O(\sqrt{n})$, we have worst-case hardness as well as conjectured average case hardness (Conjecture 33) due to Ref. [24]. Moreover, by Lemma 34 these outcomes dominate the probability distribution.

Now suppose there was an efficient sampling algorithm $\mathcal{A}$ that samples from a distribution Q_C which satisfies $d_{\mathrm{TVD}}(Q_C - P_C) \leq \epsilon$. Then we can use Stockmeyer's algorithm on input $\mathcal{A}, (x, z)$ to compute any outcome probability up to constant relative error, since in the worst case the error on the full distribution is concentrated in the sector $S_k \ni (x, z)$. □

Appendix H: 3D color code lattices

In this appendix, we provide a detailed description of the lattices that give rise to the family of 3D color codes with a growing code distance and a logical $\overline{\mathrm{CCZ}}$ gate implemented via a transversal T gate.

We start by defining $\mathcal{L}'$ to be a region of a cubic lattice with vertices corresponding to the coordinates $(x, y, z) \in \mathbb{Z}_L^3$, where $L \equiv 0 \mod 2$. We color each vertex (x, y, z) of $\mathcal{L}'$ as follows: red iff $x \equiv y \equiv z \mod 2$; green iff $x \equiv y \equiv z + 1 \mod 2$; yellow iff $x \equiv y + 1 \equiv z \mod 2$; blue iff $x + 1 \equiv y \equiv z \mod 2$; see Fig. 11(c) for an illustration. Each cube in $\mathcal{L}'$ is split into five tetrahedra spanned by the vertices of $\mathcal{L}'$. Namely, for the cube centered at $\left(x + \frac{1}{2}, y + \frac{1}{2}, z + \frac{1}{2}\right)$ if $x + y \equiv z \mod 2$, then we split it accordingly to Fig. 11(d), where one of the tetrahedra is spanned by the vertices $(x + 1, y, z)$,$(x, y + 1, z)$,$(x, y, z + 1)$ and $(x + 1, y + 1, z + 1)$; otherwise, we split it accordingly to Fig. 11(e), where one of the tetrahedra is spanned by the vertices (x, y, z),$(x + 1, y + 1, z)$,$(x + 1, y, z + 1)$ and $(x, y + 1, z + 1)$.

We obtain $\mathcal{L}$ from $\mathcal{L}'$ by including additional red vertices and tetrahedra. Namely, for every green vertex (x, y, z) at the top $z = L$ or bottom $z = 0$ boundaries of $\mathcal{L}'$ we add a vertex at either $(x, y, L + 1)$ or $(x, y, -1)$; for every yellow vertex (x, y, z) at the front $y = 0$ or rear $y = L$ boundaries of $\mathcal{L}'$ we add a vertex at either $(x, -1, z)$ or $(x, L + 1, z)$; for every blue vertex (x, y, z) at the left $x = 0$ or right $x = L$ boundaries of $\mathcal{L}'$ we add a vertex at either $(-1, y, z)$ or $(L + 1, y, z)$. Subsequently, for each newly added red vertex we add four tetrahedra spanned that vertex and neighboring vertices belonging to the boundary of $\mathcal{L}'$, as illustrated in Fig. 11(f,g).

Appendix I: Numerical simulation techniques

In this appendix, we outline the numerical methods used to produce Figs. 5 to 8 and 10 in the main text.

1. XEB and fidelity of Clifford circuits

We simulate all Clifford circuits using the `stim` package [138]. To evaluate the XEB, we perform Gaussian elimination on the noiseless stabilizers of the target state, effectively finding the subset of stabilizer generators that are diagonal in the measurement basis. For n qubits and k diagonal stabilizers, every non-zero bitstring probability is 2^{k-n} and, thus, the ideal XEB of the circuit is

$$\chi_{\mathrm{ideal}} = 2^k - 1, \tag{I1}$$

which we use to obtain Fig. 5(a) by sampling random hIQP circuits.

For noisy simulati ons, we employ two separate approaches. In Fig. 4(a) and Fig. 10, we use `stim`'s mega-sampling capabilities to quickly obtain large numbers of noisy bitstrings whose probabilities are then evaluated based on diagonal, pre-computed logical-circuit stabilizers. In Fig. 7 and Fig. 8, we need to also evaluate fidelity (either encoded or not) so we perform multiple tableau simulations, one for each noise realizations; the fidelity is 0 if any of the circuit-level stabilizers is violated, and 1 otherwise. In Fig. 8, we additionally calculate the corrected (EC) logical fidelity where we perform error correction in the Z basis (directly in the tableau) before evaluating the fidelity; we use a simple lookup-table decoder constructed by enumerating up to weight-2 physical errors.

Since the numerical studies performed here are based on sampling various circuits and noise realizations, the accessible system sizes and noise rates are severely limited due to the uncertainty associated with a finite sample size. A more sophisticated numerical study based on statistical-mechanics models would enable, e.g., detailed studies of the phase transition in the XEB-to-fidelity ratio.

2. TVD of IQP-circuit connectivity

A degree-2 IQP circuit is described by the polynomial function

$$f(x) = \sum_{ij} b_{ij} x_i x_j + \sum_j a_i x_i, \tag{I2}$$

as discussed in Eq. (19). Thus, we can capture the circuit connectivity via an $n \times n$ matrix whose off-diagonal and diagonal terms are given by the b_{ij} and a_i variables, respectively. We look at averaged eigenvalue distribution of such adjacency matrices to compare various circuit ensembles and calculate the total variation distance (TVD) of the histograms as the figure of merit.

For a uniformly random IQP circuit, b_{ij} and a_i are drawn from a random distribution with uniform probability 1/2, whereas for the hIQP circuits they are sampled as described in Fig. 3(a). The CNOT gate layers are incorporated through a frame transformation on the adjacency matrix A,

$$A \to M^{\mathrm{T}} A\, M, \tag{I3}$$

which is then added modulo 2. The CNOT matrix M is constructed based on the CNOT gate action that transforms bitstrings as

$$\mathrm{CNOT}(x_1, x_2) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \tag{I4}$$

which is embedded for many CNOTs in parallel in the larger M matrix.

In Fig. 5(b), we average over 10^6 uniform and hIQP circuit realizations for each data point.

3. Estimating classical hardness

In Fig. 6, we compute the treewidth and minimal vertex cover of the effective IQP hypergraphs realized by random hyperdepth-$\mathfrak{L}$ hIQP circuits for various hypercube dimensions $\mathfrak{D}$. To do this, for each pair $(\mathfrak{L}, \mathfrak{D}) \in [2]\times[5]$ we sample 100 random hIQP circuits and compute the effective IQP hypergraph using the relation (A6). The result is a hypergraph on $3 \cdot 2^{\mathfrak{D}}$ vertices.

To compute the treewidth of this hypergraph, we translate the hypergraph to a simple graph G by adding a 3-clique for every degree-3 hyperedge. We then use the `treewidth_min_fill_in` heuristic implemented in the python `networkx` package to compute the treewidth [139].

To compute the minimum vertex cover of the degree-3 part of the hypergraph, we write the optimization problem as an integer linear program ILP

$$\min_{x\in\{0,1\}^n} \sum_{i=1}^{n} x_i \tag{I5}$$
$$\text{subject to } x_{e_1} + x_{e_2} + x_{e_3} \geq 1 \;\forall e \in E_3$$

where $E_3 \subset [n]^{\times 3}$ denotes the set of degree-3 hyperedges of the IQP graph. We then use the `Gurobi Optimizer`, a state-of-the-art solver to solve this ILP [140].