

Reconstruction of detector error model for quantum error correction

Cheng Ye(叶澄) and Pan Zhang(张潘)

Citation: Chin. Phys. B, 2026, 35 (**7**): 070301. DOI: 10.1088/1674-1056/ae663a
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Distributed quantum circuit partitioning and teleportation optimization based on a multi-dimensional evaluation strategy

Le Zhang(张乐), Zhijin Guan(管致锦), Shuo Qin(秦硕), Zheng Luo(罗政), Fei Ding(丁飞), and Xueyun Cheng(程学云)
Chin. Phys. B, 2026, 35 (**5**): 050305. DOI: 10.1088/1674-1056/ae1c30

Hierarchical QAOA circuit design framework for distributed quantum computing

Ting-Yu Luo(骆挺宇) and Yu-Xin Deng(邓玉欣)
Chin. Phys. B, 2026, 35 (**4**): 040309. DOI: 10.1088/1674-1056/ae00ae

Generation of cross-cross resonance gates with two fluxonium qubits

Xinpeng Chen(陈鑫鹏) and Zeliang Xiang(项泽亮)
Chin. Phys. B, 2026, 35 (**4**): 040308. DOI: 10.1088/1674-1056/ae13ea

Schemes of IPsec integrated with quantum key distribution

Chun-Hui Zhang(张春辉), Wen-Xuan Zhang(张文轩), Xing-Yu Zhou(周星宇), Yuan Cao(曹原), Jun Wang(汪军), Jian Li(李剑), and Qin Wang(王琴)
Chin. Phys. B, 2026, 35 (**4**): 040305. DOI: 10.1088/1674-1056/ae3301

Distributed Kuperberg's algorithm

Peng-Yu Yang(杨鹏宇), Xin Zhang(张新), and Song Lin(林崧)
Chin. Phys. B, 2026, 35 (**3**): 030303. DOI: 10.1088/1674-1056/adfb58

RAPID COMMUNICATION

Reconstruction of detector error model for quantum error correction

Cheng Ye(叶澄)[1,2] and Pan Zhang(张潘)[1,2,3,4,†]

[1] *CAS Key Laboratory for Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing* 100190, *China*
[2] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing* 100049, *China*
[3] *School of Fundamental Physics and Mathematical Sciences, Hangzhou Institute for Advanced Study, UCAS, Hangzhou* 310024, *China*
[4] *Beijing Academy of Quantum Information Sciences, Beijing* 100193, *China*

(Received 18 March 2026; revised manuscript received 22 April 2026; accepted manuscript online 29 April 2026)

Fault-tolerant quantum computing fundamentally relies on the accurate characterization of circuit-level noise to optimize decoding algorithms. However, extracting complex multi-body error correlations remains challenging. Contemporary greedy inference algorithms can suffer from statistical distortion, discarding true physical mechanisms while introducing many unphysical false positives. Here, we introduce the correlation-analysis-based hypergraph reconstruction (CAHR) algorithm, a globally consistent framework to invert experimental syndrome statistics directly into discrete physical hypergraphs. By coupling exact algebraic correlation equations with a top-down concurrent-pruning strategy, CAHR recovers the fault topology without false positives for both $d = 5$ rotated surface codes and dense 8-body 2D color codes in our benchmark settings. Furthermore, we show that exact continuous parameter extraction in dense codes is limited by a *variance cascade*, where absolute statistical variance accumulates linearly from high- to low-degree mechanisms. This motivates a two-stage inference paradigm: utilizing CAHR to extract the fault topology, followed by continuous probability optimization. This provides a practical approach for characterizing and decoding highly correlated noise in realistic quantum hardware.

Keywords: quantum computing, quantum error correction

PACS: 03.67.Pp, 03.67.Lx, 05.40.Ca

DOI: 10.1088/1674-1056/ae663a **CSTR:** 32038.14.CPB.ae663a

1. Introduction

The realization of fault-tolerant quantum computing relies on quantum error correction (QEC) to actively protect quantum information from environmental decoherence and operational imperfections.[1–3] A central bottleneck in this pursuit is the precise characterization of complex, circuit-level noise to provide high-quality prior information for decoding algorithms.[4] To formalize these intricate dynamics, the detector error model (DEM) provides a concise probabilistic hypergraph representation.[5] While standard two-point graph representations suffice for simple noise, realistic extraction protocols and advanced topological codes — such as surface codes[6,7] and 2D color codes[8] — inherently generate complex, multi-body error correlations.[9] These manifest as high-order hyperedges within the DEM, representing single physical faults that simultaneously trigger multiple detectors.[10]

Systematically reconstructing arbitrary-order hypergraphs directly from experimental syndrome statistics remains a formidable challenge. Recent works have already established exact correlation equations and algebraic inversion routes for DEM estimation.[11–13] Among them, Ref. [13] provides the most direct empirical reference for DEM reconstruction, while Refs. [11,12] primarily focus on correlation analysis and algebraic inversion. However, practical topology discovery still faces a severe combinatorial bottleneck. Existing workflows often rely on order-by-order (greedy) inference or computationally expensive maximum-likelihood optimization, and sequential greedy paradigms are particularly vulnerable to finite-sample contamination: when lower-order correlations are fitted before higher-order mechanisms are resolved, unmodeled high-order physical errors pollute the statistical estimators, distort residual correlations, and can simultaneously generate false positives and suppress genuine hyperedges. Distinct from these topological failures, extracting exact continuous probabilities is further limited by a *variance cascade* — a phenomenon where statistical variance accumulates from high-order parent hyperedges down to low-order child faults, making precise probability calibration increasingly fragile in dense hypergraphs.

In this work, we address this topology-discovery bottleneck by introducing the correlation-analysis-based hypergraph reconstruction (CAHR) algorithm. The mathematical foundation of CAHR relies on the exact analytical equations of arbitrary-order syndrome correlations that have already been

[†]Corresponding author. E-mail: panzhang@itp.ac.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

derived in the recent literature.[11–13] CAHR is a top-down workflow that combines permissive candidate generation with concurrent pruning, allowing the discrete hypergraph topology to be identified before pursuing high-accuracy continuous parameter fitting. An alternative derivation of the same correlation equations within a tensor network detector error model (TNDEM) formalism is provided in the Appendix, where TNDEM serves as a structured derivational framework that interfaces naturally with tensor-network contraction.

We validate CAHR on both $d = 5$ rotated surface codes and the denser 8-body $d = 5$ 2D color codes. The latter is useful because directly comparable DEM-reconstruction results for dense color-code settings remain scarce. Within these benchmarks, CAHR attains exact topology reconstruction with zero false at 5×10^6 and 1.5×10^7 shots, respectively. We further show that supplying the reconstructed topology to belief propagation with ordered statistic decoding (BP–OSD)[14,15] restores convergence toward the ideal logical error rate (LER).

By isolating discrete geometric structure from cascaded probability errors, we therefore advocate a two-stage decoupled inference paradigm. In this view, CAHR is best understood as a stage-1 topology-discovery tool that resolves the discrete hypergraph, while high-accuracy continuous parameter refinement can be delegated to downstream statistical or learning-based methods on the fixed topology.[16,17] This division of labor is useful in dense correlated-noise settings, where topology identification and parameter calibration exhibit different statistical difficulties.

The paper is organized as follows. Section 2 details our algorithm construction, derives the exact analytical relationships, and introduces the top-down double-pruning algorithm. Section 3 investigates the impact of finite-sample fluctuations and the variance cascade, demonstrating end-to-end LER validation on distinct topological codes. Finally, Section 4 discusses the physical implications of our findings and formally outlines the two-stage decoupled inference paradigm.

2. Methodology

In realistic fault-tolerant quantum computing, conventional qubit-level Pauli error models are often insufficient to capture the intricate, correlated fault dynamics introduced by noisy syndrome extraction circuits. To accurately characterize these processes, the detector error model (DEM) is widely adopted.[5,18,19] A DEM formalizes the probabilistic relationship between physical error mechanisms and the triggered detectors across the spacetime circuit. Mathematically, a DEM is represented as a hypergraph $G = (V, E)$, where the vertex set V corresponds to the detectors, and a hyperedge e in set E represents independent physical error mechanisms that simultaneously flip all detectors in set e.

2.1. Analytical solution for arbitrary-order correlations

Building upon a tensor network (TN) formalism of the DEM (with the complete derivation provided in the Appendix), we establish an exact algebraic mapping between experimental observables and physical error probabilities. We define the observable variable $\sigma_i = 1 - 2s_i$, where $s_i \in \{0, 1\}$ is the binary syndrome of detector i. The correlation of a subset of detectors α is governed by the moment equation

$$M_\alpha = \langle \prod_{i \in \alpha} \sigma_i \rangle = \prod_{|e \cap \alpha| \equiv 1 \pmod 2} (1 - 2p_e). \tag{1}$$

This system consists of $2^{|V|} - 1$ equations corresponding to all possible error mechanisms. Rather than relying on logarithmic transformations used in prior heuristic approaches, we preserve the multiplicative structure and introduce a set of recursive helper functions, f_β, to achieve a compact, closed-form inversion:

$$f_\beta = \prod_{\alpha \subseteq \beta} M_\alpha^{(-1)^{|\alpha|-1}}. \tag{2}$$

Using these helper functions, the exact probability of an arbitrary-order hyperedge α is analytically resolved as

$$p_\alpha = \frac{1}{2}\left[1 - \frac{(f_\alpha)^{1/C_\alpha}}{\prod_{\beta \supsetneq \alpha}(1 - 2p_\beta)}\right], \tag{3}$$

where the scaling exponent is $C_\alpha = 2^{|\alpha|-1}$. This closed-form solution is exact for arbitrary-order hyperedges within this formalism, avoiding the need for complex optimization loops and the associated local-minimum issues of numerical fitting procedures.

It is important to note that equivalent systems of equations and their algebraic inversions have already been derived in several recent works.[11–13] Our TN-based treatment is used here as an alternative structured derivation, which is convenient for organizing the correlation equations and connecting them to tensor-network contraction.

2.2. Inference with finite statistics and structural priors

While the original full-rank system yields a unique solution corresponding to the ideal case, practical QEC circuits are constrained by finite sample sizes (N_{shots}) and possess hypergraphs of limited orders where the physical error rates of nonexistent hyperedges are zero. Finite sampling introduces statistical fluctuations, causing the analytical solutions for these non-existent edges to manifest as small, non-zero values.

To address this discrepancy, we modify the solution to maintain consistency with a given topological prior (a candidate hypergraph set E). We compute only the necessary equations for existing edges $\alpha \in E$ and remove the terms of non-

existing edges in the denominator by enforcing a zero probability:

$$p_\alpha = \frac{1}{2} - \frac{1}{2}\frac{(f_\alpha)^{1/C_\alpha}}{\prod_{\beta\supsetneq\alpha,\beta\in E}(1-2p_\beta)}. \tag{4}$$

Note that the helper function still consists of M_α for all α whether in E or not. In previous work, the instance constraining the correlation in pair order has been proved to be consistent with the classical pair correlation analysis.[4]

2.3. Correlation-analysis-based hypergraph reconstruction

In experimental scenarios, the true set of active physical error mechanisms E is *a priori* unknown and must be directly inferred from the measured syndrome data. To systematically reconstruct the discrete hypergraph and suppress the combinatorial explosion of high-order candidates without human bias, we propose the CAHR algorithm via global necessary-condition search and concurrent pruning.

Before detailing the procedure, it is important to clarify our departure from the sequential, order-by-order greedy growth strategies prevalent in recent literature. While the theoretical foundation of greedy hyperedge growth is mathematically rigorous, its layer-by-layer statistical filtering is ultimately excessive for our primary objective. The fundamental bottleneck in arbitrary-order correlation analysis is the combinatorial explosion of potential hyperedges, which scales strongly with the number of detectors $|V|$ and the hyperedge degree k. By enforcing the simple geometric necessary condition of pair-correlation cliques under a reasonable finite-sample regime (where initial false-positive pairs are sufficiently suppressed), we already achieve our core goal: compressing the large combinatorial search space down to a computationally tractable candidate pool. At this stage, further heuristic structural filtering offers only marginal benefits.

Furthermore, given the complex and often unpredictable propagation of finite-sample statistical noise across different correlation orders, we argue that the explicitly calculated physical error probability, p_α, serves as the most reliable metric for hyperedge validation within this framework. As dictated by our exact analytical solution in Eq. (4), the accurate computation of p_α requires the prior knowledge of all its parent hyperedges ($\beta \supsetneq \alpha$). This algebraic dependency necessitates a top-down evaluation sequence. Consequently, our framework deliberately adopts a permissive candidate generation strategy — allowing all potential hyperedges identified by the pair-clique prior to exist initially — and defers the elimination of fluctuation-induced false positives to a *concurrent* pruning executed during the global, top-down probability calculation. The complete reconstruction protocol proceeds as follows:

1. Pair-correlation analysis with initial pruning If high-order hyperedges exist in the underlying DEM, they inevitably project onto the pair-correlation space, yielding an apparent 2-body probability $\tilde{p}_{ij} = \frac{1}{2} - \frac{1}{2}\sqrt{f_{ij}}$. In the physically relevant regime ($p \ll 1$), this mapping is approximated as $\tilde{p}_{ij} \approx p_{ij} + \sum_{\{i,j\}\subsetneq\alpha} p_\alpha$. This leads to a simple geometric necessary condition: any valid high-order hyperedge α must project onto a fully connected clique of positive apparent correlations among its sub-pairs ($\tilde{p}_{ij} > 0$). However, finite-sample statistical fluctuations inevitably generate false-positive pair correlations. Without pruning, these noise-induced edges can interlink and produce a large number of unphysical "ghost cliques." To suppress this effect, we calculate the pair correlations directly from the data and apply a conservative initial pruning threshold $\varepsilon_{\text{pair}}$, removing any edge (i, j) for which $\tilde{p}_{ij} < \varepsilon_{\text{pair}}$.

2. Maximal plausible DEM formulation Treating the pruned $\tilde{p}_{ij}$ elements as an adjacency matrix, we then identify fully connected cliques up to a predefined maximum topological order ($k_{\max}$) using standard network-processing algorithms. Here $k_{\max}$ is chosen from prior knowledge of the syndrome-extraction circuit and code structure, namely, the highest detector-correlation order that a single fault mechanism can generate. For example, we use $k_{\max} = 4$ for the rotated surface code and $k_{\max} = 8$ for the 2D color code. In settings with less prior knowledge, one may choose a larger $k_{\max}$, more conservative $k_{\max}$, at the cost of a larger candidate space and more expensive clique search. The global candidate hypergraph $G_{\text{candidate}}$ is then constructed by including all sub-hyperedges contained within these bounded cliques. This structurally redundant yet geometrically constrained topology rules out regions where high-order hyperedges cannot exist, thereby reducing the combinatorial search space.

3. Global correlation analysis with instant pruning Finally, we execute the modified global correlation analysis over this candidate hypergraph $G_{\text{candidate}}$. The deliberately loose and redundant conditions of the previous step preserve topological completeness, but they also introduce many non-physical candidate hyperedges. Due to statistical fluctuations, the exact analytical inversion may assign small positive values to such dummy hyperedges, thereby producing false positives. We therefore apply a global pruning rule: immediately after calculating p_α, the hyperedge is discarded if $p_\alpha < \varepsilon_{\text{global}}$. As a result, finite-sample pruning inevitably carries a false-negative risk whenever the inferred probability of a genuine mechanism is pushed below zero or below the chosen non-negative threshold.

Two remarks are important for interpreting this procedure. First, because the pair signal in step 1 represents the projection of higher-order mechanisms, the initial screening should follow the signal scale of the higher-order mechanisms

rather than that of genuine pair mechanisms. In the current implementation, if the projection of a genuine hyperedge is pruned at this stage, then that hyperedge will still be removed in step 3 even if it is retained in the candidate topology.

Second, statistical fluctuations affect not only dummy hyperedges but also genuine weak mechanisms. In addition to producing small positive probabilities for non-physical candidates, finite sampling may also drive the inferred probabilities of genuine weak mechanisms to negative values when the sample size is insufficient. Since physical error rates require non-negative probabilities, only positive pruning thresholds are meaningful in practice.

The operational significance of integrating this pruning directly into the recursive evaluation process is that it maintains algebraic consistency. By resolving the analytical equations in descending order of hyperedge degree, this *concurrent* truncation ensures that once a spurious high-order hyperedge is identified and pruned, its unphysical probability contribution is set to zero before it can propagate further down the hierarchy. In this way, the procedure decouples statistical artifacts from the underlying equation system and mitigates the downward propagation of finite-sample noise into the probability estimates of lower-order constituent edges.

3. Results

3.1. Global topological reconstruction via correlation analysis

To establish the structural exactness of our framework, we first evaluate its capacity to reconstruct the complete physical hypergraphs of distinct topological codes from raw experimental statistics, which are efficiently generated using the high-performance stabilizer simulator Stim.[5] Table 1 summarizes the topological discovery performance. In these experiments, rather than applying a single uniform cutoff, we implement a mechanism-specific dual-threshold strategy: $\varepsilon_{\text{single}}$ for single-detector errors (boundary edges) and $\varepsilon_{\text{multi}}$ for multi-detector errors (correlating hyperedges). This distinction is critical because the anticipated physical error rates inherently differ across these two fundamental classes of fault mechanisms.

Table 1. Exact global hypergraph reconstruction via DEM inference. This table demonstrates the performance of our algorithm on a $d = 5$, $r = 5$ rotated surface code and a highly dense $d = 5$, $r = 4$ 2D color code (both at a physical error rate of $p = 0.001$) across varying finite-sample regimes. To accurately isolate genuine correlations from finite-sample statistical noise, the topologies are evaluated using mechanism-specific pruning thresholds (e.g., $\varepsilon_{\text{multi}} = 3 \times 10^{-5}$ for multi-detector correlations), anchored at $\approx p_{\text{min}}/2$ of their respective error classes. For the surface code, finite-sample fluctuations are sufficiently suppressed at 5×10^6 shots, causing both false-positive and false-negative artifacts to vanish and recovering the ideal topology. For the highly dense 2D color code, which features hyperedges up to degree 8, the same criterion is reached at 1.5×10^7 shots. The execution time is decoupled into CPU core time (for graph generation and clique search) and GPU time (for batched tensor inversion).

Topological code	Analysis shots	Total edges ($\lvert E\rvert$)	False Pos.	False Neg.	CPU time (s)	GPU time (s)
Rotated surface ($d = 5$, $r = 5$)	5×10^4	5262	3944	361	8.76	4.01
	5×10^5	2219	542	2	4.73	0.54
	5×10^6	1679	0	0	7.37	3.17
2D color ($d = 5$, $r = 4$)	1.5×10^5	1106	422	159	2.97	2.97
	1.5×10^6	860	31	14	1.26	1.26
	1.5×10^7	843	0	0	7.79	7.79

Grounded in statistical signal detection theory, both thresholds are set to approximately half of their respective minimum anticipated rates ($p_{\text{min}}^{\text{single}}/2$ and $p_{\text{min}}^{\text{multi}}/2$). Under finite-sample conditions, the inferred correlation $\tilde{p}_\alpha$ behaves as a random variable sampled from a Gaussian-like distribution centered around its true underlying value, p_α^{Ideal}. Because p_α^{Ideal} is constrained to be either 0 (false edges) or a genuine probability $p \geq p_{\text{min}}$ (true physical error mechanisms), the midpoint $\varepsilon \approx p_{\text{min}}/2$ serves as a natural maximum-likelihood decision boundary for each respective error class when the two local hypotheses have comparable finite-sample widths.

This practical choice is further informed by an explicit threshold-sensitivity check. In addition to the working choice $\varepsilon \approx p_{\text{min}}/2$, we also tested the lower and higher cutoffs $p_{\text{min}}/4$ and $3p_{\text{min}}/4$ for both the rotated surface-code and color-code benchmarks; the detailed results are summarized in Tables E1 and E2. These data show that exact reconstruction is not tied to a single finely tuned threshold. At finite statistics, $p_{\text{min}}/4$ admits more false positives, whereas $3p_{\text{min}}/4$ generally gives slightly better overall results than $p_{\text{min}}/2$. The reason is that the rule $p_{\text{min}}/2$ is anchored to the weakest physical mechanism; for most hyperedges, this choice is therefore conservative. Nevertheless, in the main text we retain $p_{\text{min}}/2$ as the reference threshold because it provides a simple, uniform prescription across the full candidate set.

Applying these optimal boundaries to the $d = 5$, $r = 5$ rotated surface code (Table 1, upper half) under a physical error rate (PER) of $p = 0.001$, we observe the direct impact of finite-sample suppression. The inherently sparse and localized nature of the surface code's syndrome extraction circuits yields a manageable combinatorial space. With lower statistics (5×10^4 and 5×10^5 shots), statistical noise distributions are broad, inevitably causing the noise floor to cross the decision boundaries and admit false positives. However, by scaling the analysis to 5×10^6 shots, the noise variance is sufficiently compressed. Consequently, our method successfully infers the

exact hypergraph topology without false positives, and in this benchmark the 50% safety margin is sufficient to avoid truncating genuine weak mechanisms.

Pushing the algorithm to more demanding structures, we analyze the $d = 5, r = 4$ 2D color code at the same PER (Table 1, lower half). The lowest intrinsic probability of the true multi-detector physical hyperedges in the color code is of the same order of magnitude as that in the surface code, allowing us to utilize the identical threshold $\varepsilon_{\text{multi}} = 3 \times 10^{-5}$. However, its highly non-local extraction circuits generate a dense entanglement structure featuring up to 8-body hyperedges. Due to this combinatorial growth, the resulting baseline statistical noise is significantly elevated. Separating this 8-body topology from the noise floor requires three times the sample volume. This requires 1.5×10^7 shots (compared against the 1.5×10^5 and 1.5×10^6 baselines) to ensure that the noises causing false-positive edges and false-negative errors are suppressed below the decision boundaries, yielding accurate topological reconstruction. For context, color-code correlation-analysis studies have reported sample counts as high as 5×10^7,[10] which is consistent with the statistical demands of this setting.

Our evaluations reveal that the requisite sample complexity for exact topological reconstruction is predominantly governed by the local topological density, exhibiting sub-linear scaling with respect to the extensive spacetime volume. For instance, while a baseline of 5×10^6 shots reconstructs the topology of the $d = 5$, $r = 5$ surface code without false positives, scaling the memory circuit to $d = 5, r = 15$ demands only a modest increase to 7×10^6 shots to maintain the same criterion. Pushing the temporal depth further to $d = 5$, $r = 25$ requires only a continued marginal statistical adjustment to 9×10^6 shots. This mild incremental increase is a natural statistical consequence: while the local variance cascade depth remains topologically invariant, the expanded global candidate pool in deeper circuits marginally elevates the probability of rare statistical tail events crossing the decision boundary, thereby necessitating a slightly tighter suppression of the global noise floor. This near-constant temporal scaling behavior, further detailed in the Appendix, supports the tractability of our framework for deep-circuit quantum hardware.

The computational cost of the clique-search stage follows the same mechanism. As already discussed in Section 2, insufficient statistics generate spurious pair correlations, which in turn induce numerous ghost cliques and enlarge the candidate search space. A dedicated CPU-time benchmark for the $r = 5$ rotated surface code across distances $d = 5$, 7, 9, 11 is summarized in Table F1. The data show that clique search becomes a noticeable bottleneck only in the severely undersampled regime, where fluctuation-induced candidates proliferate rapidly with system size. Once the shot count is sufficiently large, however, the clique-search time becomes very short and increases only relatively slowly with code distance. This confirms that the expensive regime is driven primarily by noise-induced candidate proliferation rather than by code distance alone.

3.2. The variance cascade and hierarchical error accumulation

While our inference method serves as a topological prior solver, analyzing the inferred continuous error probabilities reveals a marked statistical divergence driven by the underlying code density. To achieve zero-false-positive hypergraph reconstruction under identical physical error rates and pruning thresholds, the $d = 5$ rotated surface code requires 5×10^6 analysis shots. In contrast, the highly entangled $d = 5$ color code demands 1.5×10^7 shots, three times the statistical volume for the surface code. We interpret this sample-complexity disparity through the lens of a *variance cascade*, a phenomenon rooted in our recursive algebraic formulation, as illustrated in Fig. 1.

Because the exact probability extraction relies on a hierarchical, downward algebraic subtraction, statistical fluctuations originating from finite sampling inherently cascade from parent to child hyperedges. Under a Gaussian approximation for the finite-sample statistics, the absolute variance σ^2 of a child hyperedge additively absorbs the fluctuations of all its encompassing parent hyperedges, scaling approximately as

$$\sigma^2_{\text{child}} \approx \sigma^2_{\text{isolated}} + \sum_{\text{parent}} \sigma^2_{\text{parent}}. \tag{5}$$

Governed by this additive accumulation, the observed absolute error for both the surface code (Fig. 1(a)) and the color code (Fig. 1(b)) visibly inflates as the hyperedge degree decreases. Beyond the physically existing hyperedges, this cascading variance also accumulates onto the analytically inferred probabilities of non-existing candidate edges. Consequently, the densely entangled color code demands a substantially larger statistical ensemble to compress this magnified cascaded noise floor, thereby reliably differentiating unphysical artifacts from genuine multi-body correlations. The same mechanism also increases the chance that analytically inferred probabilities of genuine weak mechanisms become negative at insufficient shot counts, which in turn raises the false-negative rate once physical non-negativity is enforced.

Despite possessing comparable mean absolute errors at lower degrees, the surface code exhibits a substantial population of degree-2 edges whose absolute errors significantly exceed the multi-body noise truncation threshold ($\varepsilon_{\text{multi}}$, denoted by the horizontal dash-dotted lines). By comparison, the absolute errors of multi-degree edges in the color code are almost entirely confined below this threshold. Nevertheless, at the

shot count used in Fig. 1(a), the surface code still reconstructs the exact fault model. This observation points to a stronger topological robustness in the surface code and, correspondingly, a greater fragility of the color-code error model under statistical fluctuations. Analyzing the relative error provides a clearer view of this contrast. Although degree-2 edges in the surface code endure large absolute fluctuations, their relative errors remain significantly lower than those in the color code (Figs. 1(c) and 1(d)). This discrepancy is driven by the variance cascade. By the Gaussian approximation, the absolute variance of high-order parent hyperedges cascades additively down to their constituent child edges. Consequently, multi-degree child edges, which possess smaller intrinsic physical error rates (p_α) in denser topologies, have weaker noise resilience. The large absolute variance inherited from their densely entangled parent hyperedges increases the probability that their analytically inferred error rates are artificially suppressed into the background noise regime or even pushed to negative values. This mechanism indicates that the densely connected color code is more vulnerable to false-negative issues than the surface code, thereby necessitating a larger sample volume to reduce such inferences. This observation is consistent with Table 1, wherein, when constrained to 1% and 10% of the ideal sample sizes, the false negatives among the erroneous inferences are markedly higher in the color code than in the surface code. In other words, the variance cascade implies a larger effective variance in the color-code setting and, therefore, a higher probability of negative inferred rates under insufficient statistics; this may also be one reason why direct correlation analysis was not adopted in the Willow color-code experiment.[20]

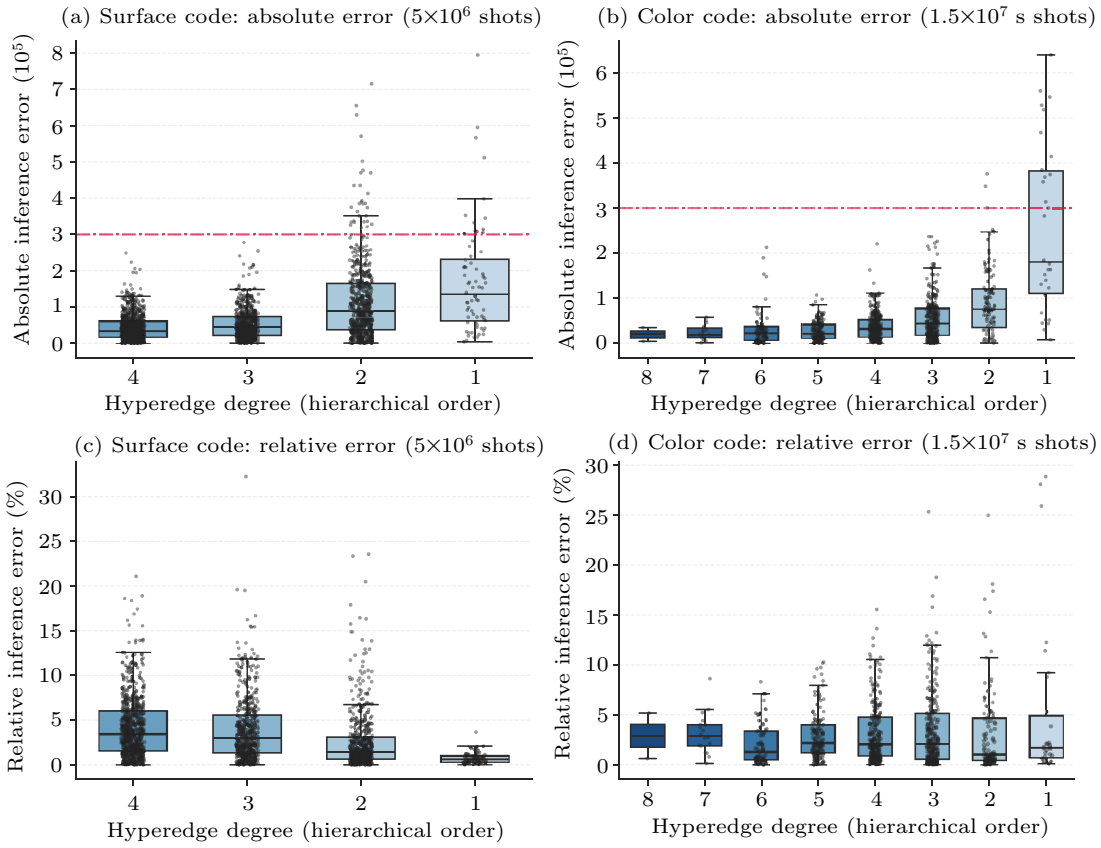


Fig. 1. Hierarchical accumulation of absolute and relative inference errors (the variance cascade). Data is shown at the sample sizes required for zero-false-positive topological reconstruction. (a) Surface code (5×10^6 shots) and (b) color code (1.5×10^7 shots) show a clear inflation of absolute error as hyperedge degree decreases due to algebraic downward subtraction. The horizontal dash-dotted lines denote the multi-body correlation truncation threshold ($\varepsilon_{\text{multi}}$). (c) In the surface code, the relative error is strongly suppressed at lower degrees (degrees 1 and 2) because the intrinsic error probabilities in the denominator are comparatively large. (d) In the color code, the large number of high-degree parent hyperedges drives a pronounced variance cascade; consequently, the relative errors at degrees 1 and 2 remain substantially higher than those of the surface code, with outliers exceeding 20%.

3.3. Operational decoding and the primacy of topology

Subsequently, to evaluate the operational impact of these statistical phenomena, we perform Monte Carlo decoding simulations using belief propagation and ordered statistics decoding (BP-OSD).[21] To ensure rigorous statistical confidence, we collect 200 logical errors for the surface code and 400 logical errors for the color code. We track decoding performance across three scaling regimes: the required shots for perfect reconstruction (100%), alongside 10% and 1% sparse-sampling baselines.

We employ a crucial ablation control — the "Given DEM Topology". In this setup, instead of utilizing the heuristically generated candidate hypergraph, the exact ideal DEM topol-

ogy is fed directly into the third step of our algorithm (Global Correlation Analysis). To isolate the impact of parameter fluctuations, the global pruning threshold is set to zero ($\varepsilon = 0$). This zero-threshold effectively bypasses structural truncation, solely clipping unphysical negative probabilities to zero to prevent the BP message-passing decoder from crashing. Thus, finite statistics are used exclusively to infer the heavily fluctuating continuous error rates over a perfect topological prior.

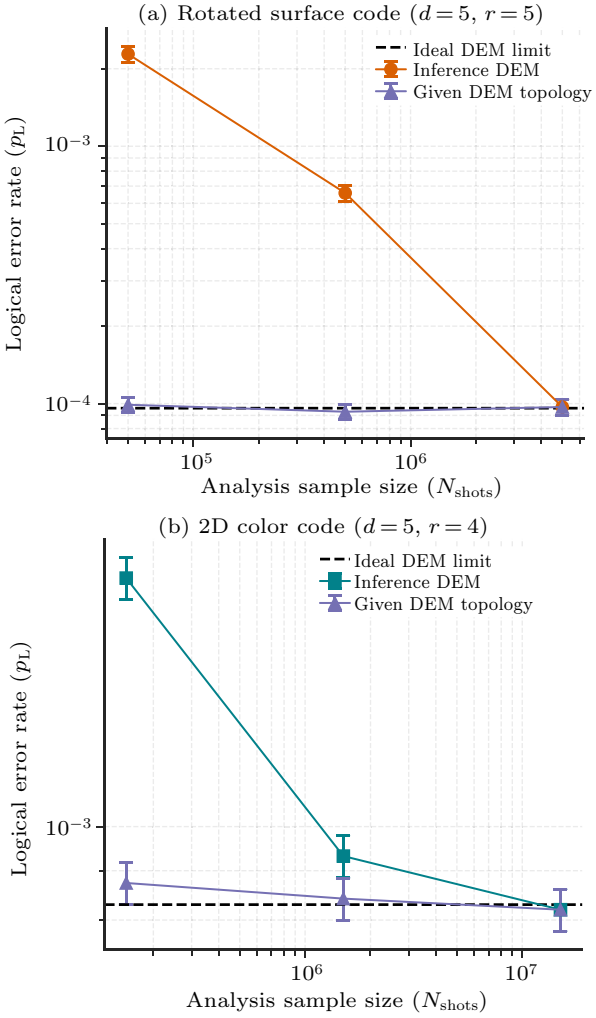


Fig. 2. End-to-end BP-OSD decoding performance and ablation analysis isolating topological vs. parameter sensitivity. Logical error rates (p_L) are evaluated by collecting 200 (surface code) and 400 (color code) logical errors per point. Data spans three regimes corresponding to 1%, 10%, and 100% of the statistics required for DEM reconstruction without false positives. For both the (a) surface code and (b) color code, the unpruned Inference DEM fails at low statistics due to dense false-positive short cycles. By contrast, the Given DEM Topology (where the exact ideal topology is fed directly into the global correlation analysis with the pruning threshold set to zero, $\varepsilon = 0$, solely clipping unphysical negative probabilities) remains much more stable. All curves approach the ideal DEM limit at the corresponding shot threshold (5×10^6 and 1.5×10^7 shots, respectively), indicating that structural exactness is more important for decoder performance than high-precision parameter estimation in this setting.

The decoding results, depicted in Fig. 2, reveal a consistent operational trend across both code families: within the correlation analysis framework, structural exactness (topology) is more important than continuous parameter precision (error rates). For both the surface code (Fig. 2(a)) and the color code (Fig. 2(b)), the Inference DEM suffers substantial logical error rate (LER) inflation at low sampling regimes (1% and 10%). This failure is primarily driven by dense false-positive hyperedges injecting many unphysical short cycles into the Tanner graph and degrading the BP message-passing phase.

By contrast, the Given DEM Topology remains comparatively resilient. Even at the minimal 1% sample size — where the inferred parameters suffer from large finite-sample fluctuations and cascaded relative errors frequently exceeding 200% — the Given DEM Topology LER stays close to the ideal theoretical baseline. For both codes, supplying the decoder with the exact topology paired with highly distorted, 1%-sampled parameters yields a much better LER than utilizing the Inference DEM derived from 10× more data (which still harbors false-positive structural artifacts). Both the Inference DEM and Given DEM Topology LER curves improve monotonically as statistics scale and meet the ideal limit once the 100% shot threshold resolves the exact hypergraph structure.

4. Discussion

In this work, we use a tensor network detector error model (TNDEM) representation as a structured framework for organizing DEM correlation equations and introduce the correlation-analysis-based hypergraph reconstruction (CAHR) algorithm for topology discovery, which is a top-down workflow that combines permissive candidate generation with *concurrent* pruning. Unlike sequential greedy growth methods that prematurely truncate signals and frequently conflate statistical noise with genuine physical correlations, this workflow suppresses unphysical artifacts before they propagate downward and bias lower-order constituent edges.

We validated this framework through end-to-end topological reconstruction and BP-OSD decoding simulations on both $d = 5$ rotated surface codes and highly dense $d = 5$ color codes. Relative to the recent DEM-estimation literature, the surface-code experiments provide a direct comparison point to reconstruction-oriented studies such as Ref. [13], while the dense color-code setting addresses a regime where closely matched DEM-reconstruction references remain limited. Our results show that the required sample complexity is governed predominantly by local topological density rather than extensive spacetime volume, with mild growth from $r = 5$ to $r = 25$. The decoding ablation, comparing the Inference DEM topology against the exact ideal topology, further shows that within this correlation-analysis framework, structural exactness of the topology is operationally more important than high-precision continuous parameter estimation when finite-sample noise is present.

In the preceding experiments, we used $\varepsilon \approx p_{\min}/2$ as the working threshold. However, $p_{\min}$ need not be known exactly

for the method to remain useful. As shown by the threshold-sensitivity results in Tables E1 and E2, varying the cutoff from $p_{\min}/4$ to $3p_{\min}/4$ does not qualitatively change the reconstruction behavior: thresholds in the neighborhood of $p_{\min}$, especially slightly larger ones, remain acceptable in practice. Therefore, in blind experimental settings where $p_{\min}$ is not directly available, a reasonable estimate can still serve as a useful threshold scale. In practice, if reliable circuit and noise-model priors are available, the compiled DEM can provide such an estimate. Otherwise, data-driven procedures such as subsampling-based estimation of the smallest stably resolved nonzero mechanism or conservative noise-floor diagnostics, as used in Ref. [13], can serve as fallback strategies.

However, our comparative investigation also exposes a fundamental boundary for purely correlation-analytical inversions. In densely entangled systems like the 2D color code, the downward hierarchical subtraction inevitably drives a pronounced variance cascade, making exact analytical parameter estimation highly susceptible to statistical fluctuations and therefore less suitable as a standalone continuous estimator. This observation clarifies the methodological role of CAHR in a two-stage decoupled inference paradigm: CAHR is most naturally used as a stage-1 topology-discovery tool that resolves the discrete hypergraph and compresses the combinatorial search space, after which stage-2 methods — such as machine-learning-based decoders, reinforcement learning, or maximum-likelihood estimation — can optimize continuous parameters on the fixed structure.

Finally, we acknowledge the current theoretical boundaries of our framework, which is fundamentally built upon the Pauli error assumption. The present formulation does not provide a controlled, explicit model for general non-Pauli noise, and we have not carried out a quantitative CAHR benchmark under explicit non-Pauli noise models; accordingly, we do not claim quantitative performance conclusions for that setting. Non–Pauli mechanisms — such as leakage[22] or coherent errors[23] — therefore require further theoretical generalization for direct modeling within this tensor-network formalism. At the same time, if non-Pauli contributions are present in data without being deliberately separated out, their syndrome-level signatures may still be partially captured by correlation analysis and reflected in the inferred DEM parameters. From the tensor-network perspective, extending TNDEM to faithfully include non-Clifford processes or leakage, and obtaining similar correlation equations, remains an important open problem. Looking ahead, exploring the tensor-network representation for computational and memory optimization, dynamic noise environments, and integration with complementary techniques such as process tomography would further broaden the scope of the framework.

Appendix A: Tensor network detector error model

This appendix maps the circuit-level DEM into a tensor-network formalism, which we refer to as the tensor network detector error model (TNDEM). The purpose of this construction is to provide a structured derivation of the correlation equations used in the main text and to make their connection to tensor-network contraction explicit. While the application of tensor networks to quantum decoding has been explored in contexts such as maximum-likelihood decoding,[24] we use TNDEM here primarily as a derivational and organizational framework for circuit-level noise characterization.

To construct the TNDEM from a general DEM hypergraph $G = (V, E)$, we have two fundamental types of local tensors.

Detector tensors (XOR): Each detector $d_i \in V$ is represented by a k-index XOR tensor, where k is the number of incident error mechanisms. Physically, this tensor enforces the local parity-check constraint, performing a modulo-2 sum (XOR) over all incoming error signals to determine the final classical parity of the detector. Within standard tensor network literature, this is mathematically equivalent to a COPY tensor conjugated by Walsh–Hadamard transforms on all its legs.

Error tensors (diagonal): Each physical error mechanism $e \in E$ is represented by a diagonal probability tensor. This tensor encodes the binary prior distribution of the fault: the error mechanism triggers all its incident detectors simultaneously with probability p_e, and remains dormant with probability $1 - p_e$.

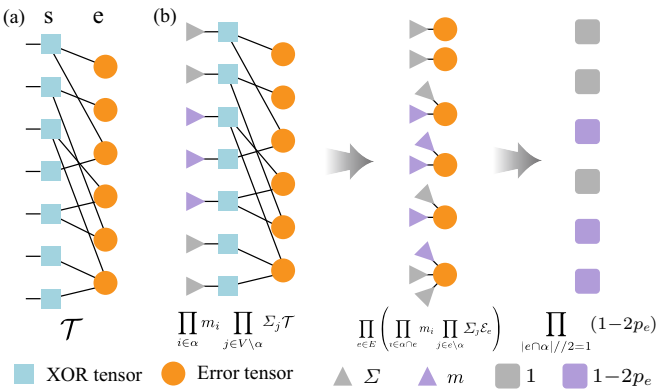


Fig. A1. Tensor network formulation of the detector error model. (a) Graphical representation of the TNDEM architecture. Each detector corresponds to an XOR tensor, which enforces a logical parity constraint for propagating Pauli errors. Independent physical error mechanisms are represented by diagonal tensors, parameterized by their fault probability p_e and identity probability $1 - p_e$. (b) Visualization of the exact tensor contraction corresponding to Eq. (B1). By projecting the detector indices onto specific boundary conditions, the global network contraction naturally decouples into a product of independent scalar contributions.

The complete TNDEM is assembled by contracting the shared indices between the error tensors and the detector tensors, following the topology of the DEM hypergraph.

Appendix B: Derivation of moment equations from the TNDEM

Unlike phenomenological approaches that construct statistical observables heuristically, the analytical power of the TNDEM framework is deeply rooted in its native tensor algebraic properties. In this section, we demonstrate that the exact system of equations governing the hypergraph inference emerges naturally from the fundamental topological rules of the network, specifically, the phase-copying property of XOR tensors. The statistical physics interpretation of these equations is then revealed *a posteriori*.

To extract physical error parameters from experimental syndrome data, one must establish correlation equations linking these microscopic parameters to the expectation values of the macroscopic syndromes. The topological structure of the TNDEM naturally yields a highly structured system of equations amenable to exact analytical inversion. Exploiting the COPY property of the XOR tensor in the dual basis, projecting a detector index onto either the sum state $\Sigma = (1,1)$ or the difference state $m = (1,-1)$ completely factorizes the tensor into a product of independent vectors. Consequently, for an N-detector system, enumerating all 2^N possible combinations of Σ and m boundary projections on the TNDEM yields a complete set of 2^N linearly independent equations. Denoting $\alpha \subseteq V$ as the subset of detectors projected onto m, with the remaining $V \setminus \alpha$ projected onto Σ, the network contraction strictly evaluates to

$$\begin{aligned}\prod_{i\in\alpha} m_i \prod_{j\in V\setminus\alpha} \Sigma_j \mathcal{T} &= \prod_{e\in E}\left(\prod_{i\in\alpha\cap e} m_i \prod_{j\in e\setminus\alpha}\Sigma_j \mathcal{E}_e\right)\\ &= \prod_{e\in E}\left[1-p_e+(-1)^{|e\cap\alpha|}p_e\right]\\ &= \prod_{|e\cap\alpha| \bmod 2=1}(1-2p_e). \end{aligned} \tag{B1}$$

This multiplicative system of equations can be rigorously linearized by taking the logarithm, rendering it highly tractable. Before proceeding to the analytical solution, we must elucidate the physical significance of contracting the boundary indices with these specific projectors. The projection state Σ effectively traces out (marginalizes) the corresponding detector index. Conversely, the projection state $m = (1,-1)$ assigns a statistical weight of $+1$ to the false (0) syndrome state and -1 to the true (1) syndrome state. This motivates the definition of a spin-like syndrome variable $\sigma_i = 1-2s_i$, where $s_i \in \{0,1\}$ is the binary value of detector d_i. Under this algebraic mapping, projecting the subset of detectors α onto m mathematically corresponds to computing the physical expectation value of the multi-detector parity observable $\prod_{i\in\alpha}\sigma_i$. Denoting this exact multi-body correlation moment as $M_\alpha = \langle\prod_{i\in\alpha}\sigma_i\rangle$, the tensor contraction rigorously reduces to

$$M_\alpha = \prod_{|e\cap\alpha| \bmod 2=1}(1-2p_e). \tag{B2}$$

The empty set trivially yields $M_\emptyset = \langle 1\rangle = 1$. The remaining 2^N-1 independent equations fully constrain all 2^N-1 possible hyperedge configurations within the fully connected DEM.

Appendix C: Exact analytical solution of the moment equations

In this section, we provide the exact algebraic derivation demonstrating how the helper function $f_\beta = \prod_{\alpha\subseteq\beta} M_\alpha^{(-1)^{|\alpha|-1}}$ systematically isolates the physical probability of specific multi-body error mechanisms.

Substituting the expression for M_α into the definition of f_β yields

$$\begin{aligned} f_\beta &= \prod_{\alpha\subseteq\beta}\prod_{e\in E,|\alpha\cap e| \bmod 2=1}(1-2p_e)^{(-1)^{|\alpha|-1}}\\ &= \prod_{e\in E}(1-2p_e)^{\sum_{\alpha\subseteq\beta}(|\alpha\cap e| \bmod 2)(-1)^{|\alpha|-1}}. \end{aligned} \tag{C1}$$

We denote the exponent of the term $(1-2p_e)$ as $C_{\beta,e}$. By replacing the modulo-2 condition with an algebraic indicator function based on powers of -1, the exponent can be elegantly rewritten as

$$\begin{aligned} C_{\beta,e} &= \sum_{\alpha\subseteq\beta}(|\alpha\cap e| \bmod 2)(-1)^{|\alpha|-1}\\ &= \sum_{\alpha\subseteq\beta}\frac{1-(-1)^{|\alpha\cap e|}}{2}(-1)^{|\alpha\cap e|-1}(-1)^{|\alpha\setminus e|}\\ &= \sum_{\alpha\subseteq\beta}\frac{1-(-1)^{|\alpha\cap e|}}{2}(-1)^{|\alpha\setminus e|}. \end{aligned} \tag{C2}$$

This expression can be reshaped by partitioning the subset enumeration $\alpha\subseteq\beta$ into two parts: one from $e\cap\beta$ and the other from $\beta\setminus e$. Let $n = |e\cap\beta|$ and $m = |\beta\setminus e|$, then the sum decomposes as

$$C_{\beta,e} = \sum_{i=0}^{n}\binom{n}{i}\frac{1-(-1)^i}{2}\sum_{j=0}^{m}\binom{m}{j}(-1)^j, \tag{C3}$$

where i counts the elements in $\alpha\cap e$ and j counts the elements in $\alpha\setminus e$. By the binomial theorem, the trailing sum evaluates to $\sum_{j=0}^{m}\binom{m}{j}(-1)^j = (1-1)^m$. This summation is identically zero except when $m=0$, which eliminates the summation entirely. Therefore, a non-zero exponent strictly requires $|\beta\setminus e| = 0$, imposing the geometric condition $\beta\subseteq e$. Under this condition, $n = |\beta| > 0$, the trailing sum reduces to 1, and the remaining part simplifies to

$$\begin{aligned} C_{\beta,e} &= \sum_{i=0}^{n}\binom{n}{i}\frac{1-(-1)^i}{2}\\ &= 2^{n-1} = 2^{|\beta|-1}. \end{aligned} \tag{C4}$$

Remarkably, the resulting exponent depends exclusively on the cardinality of the subset β, entirely independent of the specific geometric structure of e. Thus, denoting the exponent simply as C_β, we arrive at the exact analytical inversion formula

$$f_\beta = \prod_{e\in E, \beta\subseteq e} (1-2p_e)^{C_\beta}, \tag{C5}$$

$$C_\beta = 2^{|\beta|-1}. \tag{C6}$$

Appendix D: Scalability with temporal depth and constant sample complexity

A critical requirement for any practical fault-tolerant noise characterization protocol is its scalability with respect to the temporal depth (number of QEC rounds, r) of the quantum circuit. A naive expectation might assume that as the spacetime volume of the circuit expands, the required sample size for exact topological reconstruction must scale proportionally to suppress the increasing combinatorial space of potential false-positive mechanisms.

However, the CAHR exhibits favorable sub-linear scaling with respect to the temporal depth of the code. As demonstrated in Table D1, we evaluate the $d = 5$ rotated surface code across different temporal depths: $r = 5$, $r = 15$, and $r = 25$. While the baseline $r = 5$ circuit achieves zero-false topological reconstruction utilizing 5×10^6 shots, scaling the memory circuit to $r = 25$, which contains nearly six times the total number of true physical hyperedges, requires only a marginal statistical adjustment to 9×10^6 shots under identical pruning thresholds.

Table D1. Scaling of sample complexity for perfect reconstruction with temporal depth. Perfect reconstruction performance on the $d = 5$ rotated surface code across varying QEC rounds (r). $|E|$ denotes the total number of physical hyperedges, while no false-positive or false-negative appears. Pruning thresholds are fixed at $\varepsilon_{\text{single}} = 1.6\times10^{-4}$ and $\varepsilon_{\text{multi}} = 3\times10^{-5}$ that $\approx p_{\min}/2$ across all depths.

Round r	Shots	$\lvert E\rvert$	CPU time (s)	GPU time (s)
5	5×10^6	1679	8.76	4.01
15	7×10^6	5759	20.07	14.20
25	9×10^6	9839	34.29	31.58

This mild, sub-linear scaling behavior with respect to r is a direct consequence of the localized nature of the *variance cascade*. As established in Section 3, statistical variance accumulates strictly downwards from high-degree parent hyperedges to their constituent sub-edges. Extending the number of measurement rounds merely replicates the local extraction circuit topology along the time axis; it increases the extensive volume of the hypergraph but strictly preserves the intensive local topological density (i.e., the maximum hyperedge degree remains 4). Consequently, the maximal length of the hierarchical downward subtraction chain remains constant, rendering the local variance cascade depth topologically invariant with r. The slight increase in required statistics (from 5×10^6 to 9×10^6 shots) is purely a natural statistical consequence of the extensive scaling: the expanded global candidate pool in deeper circuits marginally elevates the probability of rare statistical tail events crossing the decision boundary, thereby necessitating a slightly tighter suppression of the global noise floor. This confirms that the CAHR is highly scalable and uniquely suited for characterizing the deep-circuit architectures required for practical quantum memory.

Appendix E: Threshold sensitivity of the practical pruning rule

To complement the concise discussion in the main text, we provide here the detailed threshold-sensitivity data used to assess the practical robustness of the rule $\varepsilon \approx p_{\min}/2$. We evaluate three representative thresholds, $p_{\min}/4$, $p_{\min}/2$, and $3p_{\min}/4$, for both the $d = 5$, $r = 5$ rotated surface code and the $d = 5$, $r = 4$ 2D color code. The purpose of these tables is not to identify a unique optimal cutoff, but rather to show how the false-positive/false-negative trade-off shifts under finite statistics when the threshold is varied around the practical reference scale. In each table, the two axes correspond to shot count and threshold choice, and each entry is reported as an ordered pair (FN, FP).

Table E1. Threshold sensitivity for the $d = 5$, $r = 5$ rotated surface code. The two axes correspond to shot count and threshold choice, and each entry is reported as (FN, FP). Lower thresholds suppress FN more aggressively but admit substantially more FP, while higher thresholds reduce FP at the cost of larger FN at intermediate shot counts.

Shots	$p_{\min}/4$	$p_{\min}/2$	$3p_{\min}/4$
5×10^4	(546, 13270)	(361, 3944)	(323, 2071)
5×10^5	(3, 1716)	(2, 542)	(30, 170)
5×10^6	(0, 114)	(0, 0)	(0, 0)

Table E2. Threshold sensitivity for the $d = 5$, $r = 4$ 2D color code. The same threshold sweep as in Table E1, now for the denser color-code benchmark. Each entry is reported as (FN, FP). The data shows the same qualitative trade-off, but with a sharper sensitivity to finite statistics due to the denser correlated-noise structure.

Shots	$p_{\min}/4$	$p_{\min}/2$	$3p_{\min}/4$
1.5×10^5	(164, 1122)	(156, 600)	(172, 375)
1.5×10^6	(8, 228)	(11, 47)	(23, 25)
1.5×10^7	(0, 20)	(0, 0)	(0, 0)

Several trends are immediate from Tables E1 and E2. First, $p_{\min}/4$ consistently amplifies false positives in both benchmarks, especially in the low- and intermediate-shot regimes, because the weaker cutoff admits a larger fraction of the finite-sample noise floor. Second, $3p_{\min}/4$ more strongly suppresses false positives, but at intermediate shot counts it more readily truncates genuine weak mechanisms and therefore raises the false-negative count. Third, $p_{\min}/2$ provides the most balanced FN/FP trade-off under finite statistics, which is why we adopt it as the practical threshold in the main text. At sufficiently large shot counts, both $p_{\min}/2$ and $3p_{\min}/4$ reach

zero FN and zero FP, whereas $p_{\min}/4$ can still retain residual false positives because its decision boundary remains too close to the finite-sample noise floor.

Appendix F: Clique-search CPU time at different code distances

We next quantify the computational cost of the clique-search stage discussed qualitatively in Section 2. Table F1 reports the CPU core time of the clique-search step for the rotated surface code at $r = 5$, across code distances $d = 5$, 7, 9, 11 and shot counts 5×10^4, 5×10^5, and 5×10^6. All times are reported in seconds and correspond to the average CPU core time over 5 independent runs on cores of an AMD EPYC 9654 96-core processor.

Table F1. CPU time of clique search for the rotated surface code at different distances. The table reports the average CPU core time (in seconds) of the clique-search stage over 5 runs, for $r = 5$ circuits at different code distances and shot counts. The strong low-shot growth reflects the proliferation of ghost cliques induced by fluctuation-generated pair correlations.

Shots	$d = 5$	$d = 7$	$d = 9$	$d = 11$
5×10^4	1.011428	9.504719	53.575279	255.009007
5×10^5	0.143893	0.216576	0.810720	2.085380
5×10^6	0.018353	0.074927	0.146087	0.234581

The table supports two main conclusions. First, clique search becomes a noticeable computational bottleneck only when the shot count is severely insufficient. In that regime, fluctuation-induced pair correlations create many ghost cliques, and the candidate search space expands rapidly with system size. This is precisely the quantitative manifestation of the qualitative mechanism already described in Section 2. Second, once the shot count is large enough to suppress most spurious pair correlations, the clique-search time becomes very short and grows only relatively slowly with code distance.

Acknowledgements

A Python implementation of our algorithm is available at Ref. [25]. We are grateful to Dong-Yang Feng, Han-Yan Cao and Feng Pan for helpful discussions. The work is supported by the National Natural Science Foundation of China (Grant Nos. 12325501 and 12447101).

References

[1] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* 10th ed (Cambridge University Press)
[2] Terhal B M 2015 *Rev. Mod. Phys.* **87** 307
[3] Campbell E T, Terhal B M and Vuillot C 2017 *Nature* **549** 172
[4] Spitz S, Tarasinski B, Beenakker C W J and O'Brien T E 2018 *Advanced Quantum Technologies* **1** 1800012
[5] Gidney C 2021 *Quantum* **5** 497
[6] Dennis E, Kitaev A, Landahl A and Preskill J 2002 *Journal of Mathematical Physics* **43** 4452
[7] Fowler A G, Mariantoni M, Martinis J M and Cleland A N 2012 *Phys. Rev. A* **86** 032324
[8] Bombin H and Martin-Delgado M A 2006 *Phys. Rev. Lett.* **97** 180501
[9] Harper R, Flammia S T and Wallman J J 2020 *Nat. Phys.* **16** 1184
[10] Takou E and Brown K R 2025 *Phys. Rev. A* **112** 052414
[11] Remm A, Lacroix N, Bödeker L, Genois E, Hellings C, Swiadek F, Norris G J, Eichler C, Blais A, Müller M, Krinner S and Wallraff A 2026 *Phys. Rev. Res.* **8** 013044
[12] Blume-Kohout R and Young K C 2025 arXiv:2504.14643
[13] Arms K E, McHugh M J, Nyhan J E, Reus W F and Ulrich J L 2025 arXiv:2512.10814
[14] Panteleev P and Kalachev G 2021 *Quantum* **5** 585
[15] Roffe J, White D R, Burton S and Campbell E 2020 *Phys. Rev. Res.* **2** 043423
[16] Sivak V, Newman M and Klimov P 2024 *Phys. Rev. Lett.* **133** 150603
[17] Cao H, Feng D, Ye C and Pan F 2026 arXiv:2602.19722
[18] Higgott O and Gidney C 2025 *Quantum* **9** 1600
[19] McEwen M, Bacon D and Gidney C 2023 *Quantum* **7** 1172
[20] Lacroix N, Bourassa A, Heras F J H, Zhang L M, Bausch J, Senior A W, Edlich T, Shutty N, Sivak V, Bengtsson A, *et al.* 2025 *Nature* **645** 614
[21] Higgott O 2024 *stimbposd*
[22] McEwen M, Kafri D, Chen Z, Atalaya J, Satzinger K J, Quintana C, *et al.* 2021 *Nat. Commun.* **12** 1761
[23] Darmawan A S and Poulin D 2017 *Phys. Rev. Lett.* **119** 040502
[24] Piveteau C, Chubb C T and Renes J M 2024 *PRX Quantum* **5** 040303
[25] Ye C 2025 *Source code for Correlation-Analysis-based Hypergraph Reconstruction (CAHR)*