Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identifiability and Pauli-setting concentration of metrological signals across scrambler families

Hikaru Wakaura

h.wakaura@qiri.co.jp

Quantum Integrated Research Institute Inc. (Japan)

Taiki Tanimae

Quantum Integrated Research Institute Inc. (Japan)

Short Report

Keywords: Quantum metrology, Information scrambling, Doped Clifford circuits, Matchgate circuits, Quantum Fisher information, Pauli measurements

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11724448/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Identifiability and Pauli-setting concentration of metrological signals across scrambler families

Hikaru Wakaura[1*] and Taiki Tanimae[1]

[1*]QIRI (Quantum Integrated Research Institute Inc.), Tokyo, 107-0061, Japan.

*Corresponding author(s). E-mail(s): h.wakaura@qiri.co.jp;
Contributing authors: t.tanimae@qiri.co.jp;

Abstract

Quantum sensors are increasingly embedded in many-body systems whose dynamics scramble a signal before it can be read out. Decoupling theorems bound how much information survives in a subsystem, and learning theory bounds how hard a decoder is to find; how much of the surviving information is usable with hardware-native readout has not been mapped. We address this with a calibration-assisted protocol: an observer holding a subsystem records the responses of known probe generators passed through the same unknown dynamics, identifies which generator encoded an unlabeled signal, and estimates it with Pauli measurements. For t-doped Clifford scramblers the accessible fraction decides whether information is present, while non-Clifford gates fill the rank of the dictionary response and spread sensitivity over exponentially many settings; the best single-setting fraction decays at 0.142 ± 0.014 per T gate for 8 to 18 qubits. The same inversion recurs for free-fermion scramblers doped with non-Gaussian gates, where we prove that the response rank is at most twice the subsystem size. Clifford circuits of any depth produce none of it, and a U(1) symmetry changes the readout ceiling by a factor of two. These results indicate that readout concentration is governed by the departure from free operator dynamics.

Keywords: Quantum metrology, Information scrambling, Doped Clifford circuits, Matchgate circuits, Quantum Fisher information, Pauli measurements

1 Introduction

Precision measurement is among the most mature quantum technologies [1]. Its frontier is moving from isolated, well-characterized probes toward signals embedded in interacting many-body systems such as quantum simulators, Rydberg arrays, and trapped-ion crystals. In these platforms the dynamics between signal encoding and readout are typically chaotic. Local information spreads, operators grow in support, and the signal becomes distributed over global degrees of freedom. This process is now known as information scrambling [2].

Two research programs have engaged with this situation from opposite directions. Chaos-assisted metrology treats scrambling as a resource: quantum-enhanced sensitivity survives even when only a small fraction of the sensor is accessible [3], or when readout is restricted to simple collective spin measurements [4], and for Haar-random scrambling the average reduced Fisher information recovers its full value once the retained subsystem exceeds half the system [5]. Non-stabilizerness has itself been quantified in spin-squeezing and related metrological protocols [6]. Learning-based readout has produced both encouraging and prohibitive results. Unsupervised quantum autoencoders denoise corrupted quantum data [7], learned decoders retrieve information scrambled by quasi-chaotic dynamics [8, 9], and quantum extreme learning machines estimate states beyond the scrambling time [10]. On the other hand, a no-go theorem shows that variationally learning an unknown scrambler drawn from a 2-design ensemble is exponentially hard with high probability, because any ansatz encounters a barren plateau [11–13]; efficient learning becomes possible again when two-copy Bell sampling is available and the non-Clifford content is low [14, 15].

However, these results conflate two logically distinct questions: whether information about the signal is *present* in the accessible subsystem, and whether it is *concentrated* into a restricted set of readout settings. In this paper we separate the two questions for a concrete operational task, which we call *meta-metrology* (Fig. 1). A calibration stage records the subsystem response to each generator of a known probe dictionary, passed through the same unknown dynamics. An unlabeled target response is then decomposed in this reference library to classify *which* dictionary generator G encoded the signal, and finally the signal value θ is estimated with the best calibrated setting. The protocol is calibration-assisted and assumes probe access; the target data themselves are unlabeled.

We find that the two questions are controlled by two different resources. The access fraction $f = |A|/n$ controls whether information is present: for one-shot encoding the reduced Fisher information decouples near $f = 1/2$. The non-stabilizerness, or magic, of the dynamics—quantified by the number t of non-Clifford gates—controls concentration. Clifford dynamics ($t = 0$) leave the response on a single Pauli "needle," so identification is often impossible but estimation is cheap. Adding magic fills the rank of the dictionary response and simultaneously spreads sensitivity over exponentially many incompatible Pauli settings. Between the two limits lies a finite-size window in which identification and Pauli-readout estimation both succeed. We then ask how far this picture extends beyond doped-Clifford circuits, and measure the same quantities for finite-depth brickwork circuits, U(1)-symmetric circuits, and free-fermion circuits doped with non-Gaussian gates. The inversion recurs in the free-fermion family, with

non-Gaussian gates in the role of T gates, although generic free-fermion gates are not Clifford gates; it is absent in Clifford circuits of any depth.

The contributions of this paper are:

1. A formulation of calibration-assisted dictionary identification whose first-order content is a linear response map, with an exact closed form for the ensemble-mean identifiability Gram matrix, $\mathbb{E}\,\Gamma = 2g\,\mathbb{1}$, valid for any unitary 2-design and therefore independent of magic (Lemma 2).
2. An exact solution of the Clifford point in the stabilizer formalism, giving the rank, the visibility probability, and the best-setting efficiency $\eta(0) = \min(1, 2^{s_A}/N)$ per instance, verified in 120 of 120 tested circuits (Proposition 4).
3. Statevector evidence at $n = 8$–18 that the best single-Pauli sensitivity fraction decays as $e^{-\alpha t}$ with $\alpha = 0.142 \pm 0.014$ while the available Fisher information stays near unity, and an operator-level branching benchmark $\ln(4/3)/2$ that agrees with this rate at the sizes studied.
4. Two theorems that locate which parts of this picture are rigorous: the visibility-filter step of the transfer from global operator purity to the accessible spectrum (Proposition 5), and the magic-independence of the mean sensitivity captured by every fixed stabilizer readout family (Proposition 6).
5. An end-to-end finite-shot protocol, with all calibration shots charged, showing that identification is reliable only when the calibrated pool approaches the full Pauli basis, and a reproduction package in which every number in the paper is regenerated by a gated driver.
6. Measurements of the same quantities for three further scrambler families—finite-depth brickwork, U(1)-symmetric, and free-fermion circuits doped with non-Gaussian gates—together with a proof that a free-fermion scrambler confines the response of the one-local dictionary to a space of dimension $2a$ (Proposition 7) and an operator-level benchmark $-\ln(1-2/n)$ for the decay rate per non-Gaussian gate.

Section 2 places the task among existing work. Section 3 defines the model and the three-stage protocol. Section 4 states the theoretical results and what each one does and does not establish. Section 5 describes the numerical procedures in enough detail to be rerun. Section 6 presents the doped-Clifford results and Sec. 7 the other scrambler families, Sec. 8 translates the resource counts into laboratory quantities, Sec. 9 collects the limits of validity in one place, and Secs. 10 and 11 discuss and conclude. Proofs and extended data are in the appendices.

2 Related work

The limits on decoding scrambled quantum information fall into three groups, which we use to organize prior work.

Information ceiling. Whether information is present in a subsystem is governed by decoupling. For Haar-random dynamics an observer who holds the earlier output recovers a message from little more than its own size in additional qubits [2]. One-shot decoupling theorems state the underlying decoupling condition for an arbitrary

channel in terms of smooth conditional entropies [16]. Random circuits whose gates act on arbitrary pairs of qubits decouple at polylogarithmic depth, whereas a one-dimensional geometry requires a depth that grows linearly with the system size [17]. Conservation laws change the conditions for recovery. A conserved charge delays the leakage of information and leaves a remnant [18], a symmetry imposes a lower bound on the recovery error that only coherence in the conserved quantity relaxes [19], and with a U(1) symmetry recovery depends on the input state and on the charge sector [20, 21]. Erasure of the stored early radiation degrades decoding strongly, and decoherence degrades it more weakly [22]. A teleportation-based decoding protocol separates the decay of out-of-time-order correlators caused by scrambling from the decay caused by noise [23]. In time-independent Hamiltonian systems recovery occurs in some chaotic models and not in others [24]. In metrology the corresponding statements are that chaotic dynamics can enhance sensitivity and robustness [25], that quantum Fisher information survives partial access [3, 5], and that interaction-based readout can exploit it [4].

Decoder ceiling. Whether a decoder can be found and run is a question of complexity. When the dynamics are known, explicit recovery procedures exist: a decoder for the Hayden–Preskill setting [26], teleportation protocols based on operator spreading [27, 28], and a simplified Petz recovery map that is equivalent to that decoder for Haar-random dynamics [29]; scrambling itself has been verified in trapped-ion circuits [30]. Known dynamics do not make decoding efficient. Recovering information from the radiation of a black hole by operations on the radiation alone is argued to be computationally hard under complexity-theoretic assumptions [31–34]. When the dynamics are unknown, further obstacles appear. Under cryptographic assumptions, pseudorandom states [35] and pseudorandom unitaries [36] cannot be distinguished from Haar-random ones by any efficient observer, and variational learning of an unknown scrambler can encounter barren plateaus [11–13, 37]. Structure restores learnability. A Clifford unitary is identified with a number of queries linear in n when its inverse is also available [38]. For a Clifford circuit doped with t non-Clifford gates, decoders are learnable at a cost that grows exponentially with t [8, 9]; in particular, a Clifford decoder is learned from a number of queries polynomial in n and exponential in t [39]. The states that such circuits prepare are learned in a time polynomial in n and exponential in t [40, 41], with recent extensions based on two-copy Bell sampling [14, 15]. Constant-depth circuits are learned in polynomial time from access to the unitary [42], and fermionic states prepared with t non-Gaussian gates are learned at a cost polynomial in n and exponential in t [43]. Magic also divides states into an entanglement-dominated phase, in which entanglement is estimated and manipulated efficiently, and a magic-dominated phase, in which these tasks are computationally intractable [44], and ensembles of low-magic states can be computationally indistinguishable from ensembles of high-magic states [45]. Quantum extreme learning machines avoid training the dynamics and use a linear readout instead [10].

Readout ceiling. Even when information is present and a decoder exists in principle, the measurements at hand may fail to extract it. This ceiling is the least explored of the three. Restricting measurements to a small subsystem or to low complexity degrades the estimation of time in chaotic many-body systems [46], and random circuits display

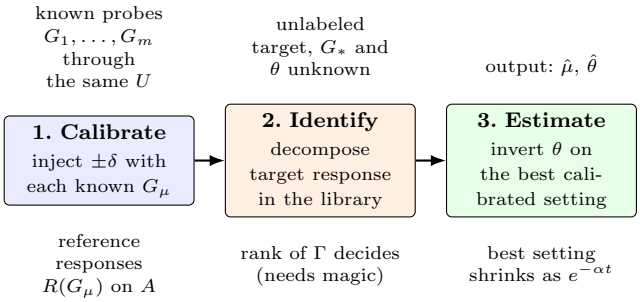

Fig. 1 The meta-metrology protocol. All three stages see only the subsystem A and the same unknown scrambler U. Stage 2 succeeds when the dictionary responses are linearly independent, which magic promotes; stage 3 becomes expensive when magic spreads sensitivity over many Pauli settings.

an intermediate regime, before full scrambling, in which readout response is most useful for learning [47]. The present paper measures this ceiling for Pauli readout across several scrambler families.

All of the works above assume that the encoded observable is known. Our task is also distinct from Hamiltonian learning [48]. There the goal is to reconstruct a Hamiltonian from local measurements of its states; here the intervening unitary is fixed and unknown and is never inverted, and the object identified is the encoding generator, relative to calibrated dictionary responses. The contribution is this calibration-assisted identification problem, its separation from Pauli-setting concentration, and the measurement of that concentration across scrambler families.

3 Model and task

3.1 System, encoding, and observer

We consider n qubits prepared in the product state $\rho_0 = |0\ldots0\rangle\langle0\ldots0|$. A scalar signal θ is imprinted by an unknown Hermitian generator G; throughout, G is a one-local Pauli ($G = X_0$ in the sweeps) and the dictionary consists of one-local Paulis. An unknown scrambling unitary U then acts. In the *one-shot* protocol the observed family is $|\Phi_\theta\rangle = Ue^{-i\theta G}|\psi_0\rangle$. In the *interleaved* protocol the encoding is re-applied between L scrambling blocks, modeling a persistent signal coupled throughout a chaotic evolution. The observer holds a subsystem A of $a = fn$ qubits and can measure the reduced state σ_θ^A.

The scrambler is drawn from the t-doped Clifford ensemble: random Clifford layers interleaved with t single-qubit T gates. At $t = 0$ the dynamics are maximally scrambling but classically structured; as t grows they approach Haar-random dynamics. We use t as the magic parameter. Non-stabilizerness can equivalently be quantified by measures such as the stabilizer Rényi entropy [8, 9, 49]; Ref. [49] establishes faithfulness, invariance under Clifford unitaries, and additivity, while monotonicity under general stabilizer-preserving operations is a separate and contested property.

3.2 The three-stage protocol

The protocol has three stages (Fig. 1). (1) *Calibrate.* For each dictionary element G_μ, $\mu = 1, \ldots, m$, inject a known small signal $\theta = \pm\delta$ and record the resulting Pauli expectation values on A. Their finite difference is the reference response of G_μ. (2) *Identify.* Nature encodes an unknown G_* from the dictionary with an unknown θ. The observed subsystem response is decomposed by least squares over the reference library, and the largest coefficient names $\hat{\mu}$. (3) *Estimate.* The calibrated response of $G_{\hat{\mu}}$ selects the single Pauli setting with the largest sensitivity, and θ is inverted linearly on that setting. In the noiseless analysis below, stages 1–2 are summarized by the response map and its Gram matrix, and stage 3 by the best-setting sensitivity fraction. The finite-shot simulation of Sec. 6 runs all three stages explicitly.

3.3 The response map and identifiability

Because the target samples are unlabeled and G is unknown, the first-order observable content of the data is the reduced response $R(G) = \partial_\theta \sigma_\theta^A|_{\theta=0}$. It obeys the structural identity

$$R(G) = -\,i\,\mathrm{Tr}_{\bar{A}}\big(U\,[G, \rho_0]\,U^\dagger\big)\,, \tag{1}$$

which follows by differentiating $\mathrm{Tr}_{\bar{A}}(Ue^{-i\theta G}\rho_0 e^{i\theta G}U^\dagger)$ at $\theta = 0$. Three consequences are immediate (Lemma 1). The response is linear in G and depends on G only through $[G, \rho_0]$. Generators that commute with the initial state are invisible. Unlabeled data determine G only up to a positive factor.

For a dictionary $\mathcal{G} = \mathrm{span}\{G_1, \ldots, G_m\}$ we form the Gram matrix $\Gamma_{\mu\nu} = \langle R(G_\mu), R(G_\nu)\rangle$. We call the dictionary coefficients *structurally identifiable* when Γ has full rank. Full rank guarantees a unique noiseless decomposition. It does not guarantee a stable one: the smallest eigenvalue and the condition number of Γ control finite-shot inversion. We therefore report rank maps as structural results and measure the Gram spectrum separately (Sec. 6). The counting condition $m \leq 4^a - 1$ is necessary but not sufficient.

3.4 The needle metric for Pauli readout

Whether the information present in A can be *extracted* with Pauli readout is not measured by the quantum Fisher information F_Q^A [50] alone. It depends on how that information is distributed over incompatible settings. Expanding the sensitivity over all Pauli observables P_j on A, $s_j = \partial_\theta \langle P_j\rangle$, the total $\sum_j s_j^2$ is basis-invariant and does not decay under scrambling. What collapses is the best single setting. We therefore define

$$\eta_{\mathrm{needle}} = \max_j s_j^2 \,/\, F_Q^A, \tag{2}$$

the fraction of the available Fisher information carried by the best individual Pauli setting. This is a concentration proxy. The exact per-setting Fisher information is $s_j^2/(1 - \langle P_j\rangle^2)$, so η_{needle} omits a variance factor; recomputing the $n = 10$ sweeps with that factor included changes the best-setting fraction by at most 3% at every sampled (a, t) (Appendix C). Writing $p_j = s_j^2/\sum_k s_k^2$ and $\Pi_A = \sum_j p_j^2$ for the purity

of the accessible sensitivity spectrum, one always has $\eta_{\text{needle}} \leq C_{n,a}\sqrt{\Pi_A}$ with $C_{n,a} = \sum_j s_j^2/F_Q^A \leq 2^a$.

4 Theoretical results

This section states each analytical result, says what it means for the protocol, and points to its proof. The limits of each statement are collected in Sec. 9.

4.1 The mean Gram matrix does not depend on magic

Lemma 2. For U drawn from any unitary 2-design and any dictionary whose commutators $[G_\mu, \rho_0]$ are orthonormal,

$$\mathbb{E}_U \Gamma = 2g\, \mathbb{1}_m, \qquad g = \frac{d\,(d_A^2 - 1)}{d_A\,(d^2 - 1)}, \tag{3}$$

with $d = 2^n$ and $d_A = 2^a$. The dictionary $\{X_i, Y_i\}$ satisfies the orthonormality condition on $|0 \ldots 0\rangle$. The proof is a two-copy Weingarten calculation (Appendix A).

Meaning. The multiqubit Clifford group is an exact unitary 3-design [51], so the same mean holds at $t = 0$ and at every t. The mean identifiability matrix therefore cannot distinguish a Clifford scrambler from a Haar-random one. Whatever changes with magic must be a fluctuation effect. Table 2 confirms this: at $n = 6$, $a = 4$ the Clifford and Haar ensembles share the mean $2g = 0.498$, but the Clifford standard deviation is 7.6 times larger.

4.2 Exact solution at the Clifford point

Proposition 4. At $t = 0$, for Pauli dictionary elements and stabilizer initial states, write $\hat{P} = UG_\mu U^\dagger$ and expand ρ_0 in the stabilizer group $\mathcal{S}$ of $U|0 \ldots 0\rangle$. The response is a sum over "partner" stabilizers, $R = 2^{1-a} \sum_{S \in \mathcal{P}} \pm(\hat{P}S)_A$ with $\mathcal{P} = \{S \in \mathcal{S} : \{S, \hat{P}\} = 0,\ \text{supp}(\hat{P}S) \subseteq A\}$. The partner count $N = |\mathcal{P}|$ is the number of solutions of an affine linear system over GF(2) and is computable exactly per instance. Then: (i) the response is nonzero if and only if $N \geq 1$; (ii) the expected rank fraction is $\mathbb{E}[\text{rank}\,\Gamma]/m = P[N \geq 1]$, because visible responses live on distinct Pauli strings and are mutually orthogonal; and (iii) the needle efficiency is $\eta(0) = \min(1, 2^{s_A}/N)$, where 2^{s_A} counts stabilizers supported inside A. The derivation assumes $a < n$ and response support inside the state support (Appendix B).

Meaning. At the Clifford point each dictionary response is either invisible or a single sparse Pauli needle. Identification fails whenever two dictionary elements are both invisible or land on the same string, which is common at $f \lesssim 1/2$. Estimation, when the response is visible, is maximally concentrated. The formula was checked against an independent numerical measurement in all 120 tested instances (Table 2).

4.3 An operator-level branching benchmark

Under Clifford twirls separating successive T gates, each Pauli component of the Heisenberg-evolved generator that anticommutes with the T gate's axis splits into

two equal-weight components. The split occurs with probability $\frac{1}{2}(1 + \frac{1}{4^n-1})$. In the collision-free operator branch tree the purity of the branch-weight distribution therefore obeys $\mathbb{E}[\Pi_t^{\mathrm{op}}] = (3/4)^t\,\Pi_0^{\mathrm{op}}$ up to the stated $O(4^{-n})$ correction, so that

$$\begin{aligned} \sqrt{\mathbb{E}[\Pi_t^{\mathrm{op}}]} &= (3/4)^{t/2}\sqrt{\Pi_0^{\mathrm{op}}}, \\ \alpha_{\mathrm{branch}} &= \tfrac{1}{2}\ln\tfrac{4}{3} = 0.1438\ldots \end{aligned} \tag{4}$$

The mechanism is that of magic spreading in doped circuits [52], and it is consistent with magic-generation studies of T-doped Clifford circuits [53]. Here it supplies a benchmark rate for metrological sensitivity.

Meaning. Equation (4) describes the *global* operator purity. It is a benchmark for the needle rate, not a theorem about it. Two steps separate the two quantities: the restriction of branches to those visible from A, and the state-level partner factor of Proposition 4. The first step is settled by the next result; the second remains open.

4.4 The visibility filter preserves the branching rate

Proposition 5. Condition on the branch weights $\{w_b\}$ produced by the doped layers. The final uniform Clifford layer randomizes the identity of each branch, so that branch b is visible from A with probability $q = (4^a - 1)/(4^n - 1)$, and the visibilities of two distinct branches have covariance bounded by $3q^2/(4^a - 1)$. Consequently the filtered purity concentrates at Π_{op}/q once the branch tree is deep enough, with an explicit Chebyshev bound, and the decay rate $\ln(4/3)$ of $\mathbb{E}[\Pi_t^{\mathrm{op}}]$ is preserved. The proof uses the transitivity of the Clifford action on Pauli pairs with fixed commutation relation (Appendix D).

Meaning. Restricting attention to the accessible subsystem changes the prefactor of the purity by $1/q$ but not its rate. The measured A-restricted purity rates, 0.282–0.290, agree with $\ln(4/3) = 0.288$ as this result predicts.

4.5 Fixed readout families have a magic-independent mean

Proposition 6. The same two-copy twirl that proves Lemma 2 gives $\mathbb{E}[s_P s_Q] = c\,\delta_{PQ}$ for non-identity Paulis P, Q on A, with c independent of both P and t. Hence for every set F of Pauli settings fixed *before* the circuit is drawn, the mean fraction of total sensitivity captured by F is $|F|/(4^a - 1)$ at every t. For a maximal commuting (stabilizer) family this fraction is $1/(2^a + 1)$ (Appendix D).

Meaning. No fixed Pauli or stabilizer readout loses sensitivity on average as magic grows. The decay reported below for the best setting, and for a greedily chosen commuting family, is a concentration effect exposed by *adaptive* selection: magic moves sensitivity from a few settings to exponentially many, so the best available setting carries less. This also explains why no shot-count lower bound over arbitrary measurements can exist while F_Q^A persists; the obstruction lives in the class of hardware-native, fixed settings and in the cost of implementing the optimal measurement.

4.6 Free-fermion scramblers: a rank bound

Let $\gamma_{2j+1} = Z_0 \cdots Z_{j-1} X_j$ and $\gamma_{2j+2} = Z_0 \cdots Z_{j-1} Y_j$, $j = 0, \ldots, n-1$, be the Majorana operators of the Jordan–Wigner transformation. A unitary is fermionic Gaussian when $U \gamma_\mu U^\dagger = \sum_\nu O_{\nu\mu} \gamma_\nu$ with $O \in \mathrm{SO}(2n)$; circuits of nearest-neighbor matchgates are of this type [54, 55].

Proposition 7. Let U be a fermionic Gaussian unitary, let ρ_0 be the vacuum $|0 \ldots 0\rangle\langle 0 \ldots 0|$, and let A consist of the first a qubits. Then the responses of the one-local dictionary $\{X_i, Y_i\}_{i=0}^{n-1}$ span a space of dimension at most $2a$, so that

$$\operatorname{rank} \Gamma \leq 2a, \qquad \operatorname{rank} \Gamma / m \leq f. \tag{5}$$

Equality holds whenever the $2a$ operators $[\gamma_\kappa, \rho_A]$, $\kappa \leq 2a$, are linearly independent. The proof (Appendix F) combines the linear action of a Gaussian state on Majorana operators with a parity argument for the modes outside A.

Meaning. However deep it is, a free-fermion scrambler cannot make the one-local dictionary identifiable from a proper subsystem. This is the counterpart of the Clifford point. There the rank is deficient because individual responses are sparse and often invisible; here it is deficient because all $2n$ responses are confined to a $2a$-dimensional space. The bound is saturated in every simulated instance (Sec. 7).

4.7 A degree-chain benchmark for non-Gaussian doping

Gaussian unitaries preserve the weight p_d that a Heisenberg-evolved operator carries on Majorana monomials of degree d. A CZ gate on neighboring modes is a Clifford gate, so it permutes monomials, and it changes their degree by 0 or ± 2. If Haar-random Gaussian unitaries act between successive CZ gates, the expected degree distribution after k gates obeys $p(k+1) = P\, p(k)$ with a transition matrix P fixed by counting local patterns (Appendix F). The signal generator $X_0 = \gamma_1$ has degree one, and a degree-one operator leaves the degree-one sector with probability exactly $2/n$ per CZ gate. Up to a return flow of order n^{-3},

$$p_1(k) \simeq (1 - 2/n)^k, \qquad \alpha_{\mathrm{deg}} = -\ln(1 - 2/n). \tag{6}$$

Meaning. Like Eq. (4), this is an operator-level benchmark and not a theorem about η_{needle}. It differs from the doped-Clifford benchmark in one respect that matters physically: it depends on n. A low-degree operator is sparse in mode space, so a local non-Gaussian gate touches it with probability of order $1/n$, whereas a T gate splits a scrambled Pauli string with probability one half at any size. The per-gate rate of the free-fermion family is therefore expected to fall with system size.

5 Numerical methods

5.1 Simulation pipelines

All results are exact statevector computations; sampling noise enters only where a measurement budget is modeled explicitly. Three independent pipelines are used. A

dense pipeline handles $n = 8$. A matrix-free pipeline, which applies Clifford layers and T gates without forming unitaries, handles $n = 10$–18 and is cross-validated against the dense pipeline at $n = 8$. A symplectic branch-tree simulator tracks the Pauli branches of the Heisenberg-evolved generator at $n = 100$–1000 without any state vector; it is validated against dense matrices at $n = 6$ with exact agreement per realization. The branch-tree simulator measures only the global operator purity of Eq. (4); it cannot evaluate η_{needle} or F_Q^A, which require the state.

5.2 Ensemble and sampling grid

Each cell of the sampling grid is specified by the system size n, the access size a (equivalently $f = a/n$), the number t of T gates, the protocol (one-shot or interleaved with L blocks), and the circuit seed. For each seed a fresh doped-Clifford circuit is drawn. Table 1 lists the sweeps. Unless stated otherwise, every cell averages eight random circuits; the $n = 10$ rank study uses sixteen, and the $n = 18$ sweep uses four. Each circuit is generated from a seeded random number stream, so every cell can be regenerated bit for bit.

5.3 Quantities recorded

For each cell we record the following. The quantum Fisher information of the full state, F_Q, and of the reduced state on A, F_Q^A, computed from the symmetric logarithmic derivative. The Pauli sensitivities $s_j = \partial_\theta \langle P_j \rangle$ for all $4^a - 1$ non-identity Paulis on A, obtained by finite differences at $\theta = \pm\delta$ with $\delta = 10^{-4}$ in the exact sweeps. From these, the needle fraction η_{needle} of Eq. (2), the effective number of contributing settings $m_{\text{eff}} = (\sum_j s_j^2)^2 / \sum_j s_j^4$, and the sensitivity-spectrum purity Π_A. For the dictionary $\{X_i, Y_i\}_{i=1}^n$ ($m = 2n$) we record the full spectrum of the Gram matrix Γ; the rank is counted with an absolute eigenvalue threshold of $2g \times 10^{-6}$, anchored to the analytic mean, and any result that depends on this threshold is labeled a structural-rank result. For the commuting-family analysis we greedily assemble a mutually commuting set of

Table 1 Sampling grid of the statevector sweeps. t is drawn from $\{0, 1, 2, 4, 6, 8, 12, 16, 24, 32, 48, 64\}$; the subset varies with n ($t \le 32$ at $n = 8$, $t \le 48$ at $n = 16$–18). "int." denotes the interleaved protocol with L blocks. The $n = 8$ sweep supplies the phase map, $n = 10$ the window, the finite-shot study, and (with 16 circuits per cell) the rank study, $n = 8$–18 the decay fits, and the last row the Gram-spectrum study. Branch-tree sweeps at $n = 100$ and 1000 use 24 and 8 realizations with $t \le 32$ and $t \le 24$.

n	a	protocol	circuits/cell
8	2, 4, 5, 6	one-shot; int. L=4	8
10	3, 5, 6, 8	one-shot; int. L=2, 4, 8	8 (16)
12	3, 6, 8, 9	one-shot; int. L=4	8
14	8, 10	one-shot	8
16	10, 12	one-shot	8
18	10, 12	one-shot	4
8–14	$n/2$, $\approx 3n/4$	one-shot	8

high-sensitivity Paulis and evaluate the classical Fisher information of the probability distribution in its joint eigenbasis.

5.4 Fitting procedure and uncertainties

The needle traces are normalized by their $t = 0$ value and fitted on a logarithmic scale to

$$\eta_{\text{needle}}(t) = (1 - c)\, e^{-\alpha t} + c, \tag{7}$$

with $\alpha \geq 0$ and a floor $c \in [10^{-6}, 1]$, using nonlinear least squares over all sampled t. Only cells with $f > 1/2$ enter the decay analysis. Three uncertainties are reported and should not be confused. (i) The standard error of each fit, which is conditional on the sampled circuits. (ii) The mean and standard deviation across the twelve fitted traces, 0.142 ± 0.014, which is a descriptive summary across sizes and access fractions. (iii) A nonparametric bootstrap over circuit realizations (2000 resamples, fixed seed), which gives per-fit and grand-mean confidence intervals over the circuit ensemble. In addition, every fit is repeated with the depth window restricted to $t \leq 16$, 24, and 32 to quantify the systematic dependence on the window (Appendix C).

5.5 Finite-shot protocol

The end-to-end simulation runs the three stages of Fig. 1 with two noise sources: finite sampling (S shots per Pauli setting, Gaussian approximation to the binomial) and readout bit-flip error ($\langle P\rangle \to (1 - 2q_{\text{ro}})^{w(P)}\langle P\rangle$ for a weight-w Pauli, flip probability q_{ro} per qubit). Calibration injects $\theta = \pm\delta$ with $\delta = 0.1$ for each of the $m = 20$ generators at $n = 10$ and estimates all Pauli expectations on A from S shots each. Identification decomposes the noisy target response by least squares and declares the largest coefficient. Estimation uses the identified generator's best calibrated setting to invert $\theta_{\text{sig}} = 0.05$. Each configuration is repeated for eight circuits and twenty noise trials. A second experiment restricts both tasks to K Pauli settings drawn at random and charges every shot: the total budget is $B = SK(2m+2)+S$, counting calibration (m generators, two θ points, K settings), signal acquisition, and estimation.

5.6 Other scrambler families

The three additional families are simulated with a separate statevector pipeline that applies arbitrary two-qubit gates and computes the response exactly instead of by finite differences. For $G^2 = 1$ and a computational-basis initial state, $\partial_\theta \rho_A = -i\,\mathrm{Tr}_{\bar{A}}(|u_G\rangle\langle u_0| - |u_0\rangle\langle u_G|)$ with $u_0 = U|\psi_0\rangle$ and $u_G = UG|\psi_0\rangle$. The observables are those of the doped-Clifford study.

(i) *Brickwork.* One-dimensional brickwork circuits with open boundaries and depth $D \leq 32$, built from Haar-random or uniformly random Clifford two-qubit gates, act on $|0\ldots0\rangle$. The subsystem is either the block of a qubits that contains the signal qubit ("near") or the block at the opposite end of the chain ("far"). (ii) *U(1).* The same brickwork is built from Haar-random gates that commute with $Z_i + Z_{i+1}$ and compared with unconstrained Haar gates, both acting on the Néel state $|0101\ldots\rangle$ up

to depth 128. (iii) *Free fermion.* Blocks of $2n$ brickwork layers of random nearest-neighbor matchgates are separated by single CZ gates on a random neighboring pair, with up to $k = 32$ CZ gates, acting on $|0 \dots 0\rangle$; A is the first a qubits.

All three families are run at $n = 10$ (16 circuits per cell; $a = 5, 6, 8$) and $n = 12$ (8 circuits per cell; $a = 6, 7, 10$). The free-fermion family is also run at $n = 8$ (16 circuits; $a = 4, 5, 6$) and $n = 14$ (8 circuits; $a = 7, 8$). A global Haar-random unitary provides the deep-circuit reference at each size. Decay rates are fitted with Eq. (7) to the circuit-averaged trace, with 95% intervals from a bootstrap over circuits (400 resamples). For the free-fermion family we also report the slope of $\log(\eta_{\text{needle}} - c)$ over $k \leq 8$, where c is the mean of the two largest-k points; this removes the finite-size floor, which is larger at small n. The pipeline carries five self-checks: agreement of the exact response with a finite difference, the Haar value of Eq. (3), conservation of charge and of parity by the respective gate sets, Gaussianity of matchgate circuits, and the value ± 2 of every nonzero Pauli sensitivity in Clifford circuits.

5.7 Verification and reproduction

Three verification suites run before any data are quoted. The Weingarten constant of Eq. (3) is checked against 300 Haar samples and 300 Clifford samples. The Clifford-point formulas of Proposition 4 are checked, instance by instance, against exact GF(2) partner counts computed by code that shares no path with the sensitivity measurement. The branch tree is checked against dense simulation. The pair-orbit probabilities used in Proposition 5 and the isotropy used in Proposition 6 are checked by Monte-Carlo Clifford sampling. Each step of the proof of Proposition 7 and the degree-transition matrix of Eq. (6) are checked on dense matrices. A single driver regenerates every number quoted in this paper from the shipped data and fails on any mismatch.

6 Results

6.1 Information persists while Pauli sensitivity disperses

Figure 2 shows the one-shot maps at $n = 8$. At $f = 0.75$ the available information F_Q^A/F_Q stays between 0.94 and 1.00 across the whole magic axis. Over the same axis η_{needle} falls by a factor of 29, from 1.000 at $t = 0$ to 0.034 at $t = 32$. At $f = 0.25$, by contrast, the reduced Fisher information is zero to numerical precision for $t \leq 2$. The location of the crossover agrees with the exact Haar-average result that the reduced Fisher information is recovered only once the retained subsystem exceeds half the system [5]. The two panels therefore separate cleanly: access controls whether information is present, magic controls whether it is concentrated. Interleaved encoding restores appreciable local information at low access in the tested circuits, so the two effects also respond differently to protocol design.

6.2 The mean is exact; the transition is a concentration effect

Table 2 tests Lemma 2 and Proposition 4. The predicted mean $2g = 0.49817$ at $n = 6$, $a = 4$ is reproduced by Haar sampling (0.50012 ± 0.00249) and by the Clifford ensemble

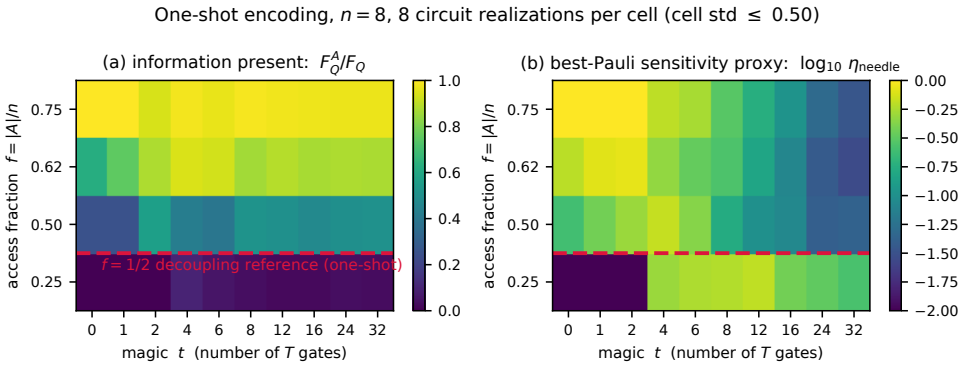

Fig. 2 Access decides presence and magic decides concentration. One-shot encoding, $n = 8$, 8 circuits per cell. (a) Available information F_Q^A/F_Q is set by the access fraction f and is flat in t; the dashed line marks the $f = 1/2$ decoupling reference. (b) Best-Pauli sensitivity fraction $\log_{10}\eta_{\text{needle}}$ collapses with t at every $f > 1/2$. The maps describe the sampled ensemble at finite size.

(0.4767 ± 0.0188). The Clifford standard deviation is 7.6 times the Haar value. The Clifford-point formula $\eta(0) = \min(1, 2^{s_A}/N)$ agrees with the independently measured needle fraction in all 120 tested instances ($n = 12$, $a \in \{6, 8, 9\}$). The measured $t = 0$ rank fractions match the exact visibility probability $P[N \geq 1]$; for example 0.03 observed against 0.031 predicted at $n = 10$, $a = 3$.

6.3 Magic fills the rank while draining the best setting

Figure 3 shows the window at $n = 10$. At $f = 0.5$ the mean rank fraction rises from 0.34 at $t = 0$ to 0.98 at $t = 6$ and reaches the full-rank threshold by $t = 12$. Over the same range η_{needle} falls from 0.42 at $t = 6$ to 0.014 at $t = 32$. The interval $t \in [6, 16]$ is the finite-size window in which identification has become possible and the best Pauli setting is still useful. At $f = 0.3$ the reduced Fisher fraction is small but nonzero, $F_Q^A/F_Q \approx 0.03$, yet the Gram matrix reaches full rank for $t \geq 16$. Structural identifiability and the amount of information available for precise estimation are thus different things, and the first can hold where the second is poor.

The Gram spectrum, measured across $n = 8$–14 (Appendix C), shows what full rank does and does not buy. Deep in the doped phase the median spectrum concentrates at the analytic mean $2g$: the condition number at $t = 32$ is 1.2–2.6 and decreases with n. Within the sampled range, therefore, no conditioning penalty accrues beyond

Table 2 The ensemble mean is magic-independent and the Clifford point is exactly solved. Verification of Eq. (3) and of Proposition 4.

	prediction	Haar (300)	Clifford (300)
$\mathbb{E}\langle R, R\rangle$ (n=6, a=4)	0.49817	0.50012(249)	0.4767(188)
std. dev.	—	0.043	0.326
$\eta(0)$ formula (n=12)	exact	—	120/120

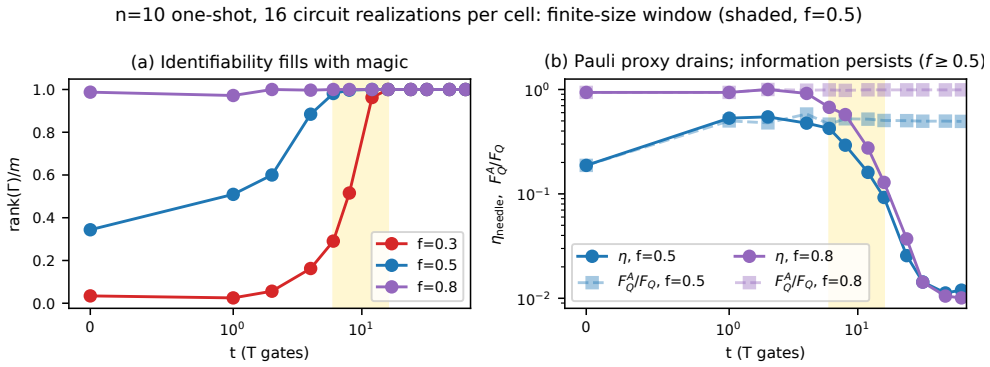

Fig. 3 A finite-size window in which identification and Pauli estimation both work. $n = 10$, one-shot, 16 circuits per cell. (a) Gram-rank fraction versus t; at $f = 0.3$ the $t = 0$ value 0.03 matches the exact prediction $P[N \geq 1] = 0.031$. Rank does not report conditioning. (b) η_{needle} (solid) and F_Q^A/F_Q (dashed) for $f \geq 0.5$. The shaded interval is the region at $f = 0.5$ where rank, information, and best-setting fraction are all $O(1)$.

the visibility scale $2g$ itself. At the window edge, $t \approx 6$ and $f = 0.5$, the condition number is about 10, so the onset of identifiability carries a finite stability cost.

6.4 The decay rate matches the branching benchmark at these sizes

Figure 4 collects the decay fits. Across twelve one-shot sweeps with $f > 1/2$ and $n = 8$–18, the fitted rates have mean and between-fit standard deviation $\alpha = 0.142 \pm 0.014$ per T gate. The largest size, $n = 18$, gives $\alpha = 0.148 \pm 0.014$ at $f = 0.556$ and 0.141 ± 0.006 at $f = 0.667$, with fitted floors at the numerical bound 10^{-6}. The circuit-ensemble bootstrap places the grand mean in $[0.130, 0.148]$. Restricting the fit window to $t \leq 16$, 24, and 32 shifts the twelve-fit mean to 0.122, 0.132, and 0.138, so the quoted rate carries a systematic window dependence of order 15% (Appendix C).

The mean agrees with the branching benchmark $\ln(4/3)/2 = 0.1438$ of Eq. (4). Three supporting observations relate the two. The measured decay rates of the A-restricted sensitivity purity are 0.282–0.290 in four data sets, close to $\ln(4/3)$, as Proposition 5 predicts. The ratio of fitted rates is $\alpha_{\text{needle}}/\alpha_{\text{purity}} = 0.500$. The branch-tree simulator gives global purity rates 0.290 at $n = 100$ and 0.279 at $n = 1000$ with zero branch collisions, so the collision-free idealization of Eq. (4) is exact at scale. The fitted floor falls from 2.4×10^{-2} at $n = 8$ to 1.5×10^{-4} at $n = 16$ ($f = 0.625$) and to the numerical bound at $n = 18$. An idealized occupancy model of the branch tree reproduces the agreement between α and the benchmark as a finite-size effect mediated by the partner multiplicity $M \approx 2^{2a-n}$ (Appendix D).

6.5 Commuting families and the per-setting distribution

Measuring the joint eigenbasis of a greedily assembled commuting family retains information longer than a single setting but still decays: $\alpha_{\text{family}} = 0.107 \pm 0.010$ and 0.072 ± 0.011 at $n = 10$ ($f = 0.6, 0.8$), and 0.090 ± 0.009 and 0.073 ± 0.010 at $n = 12$

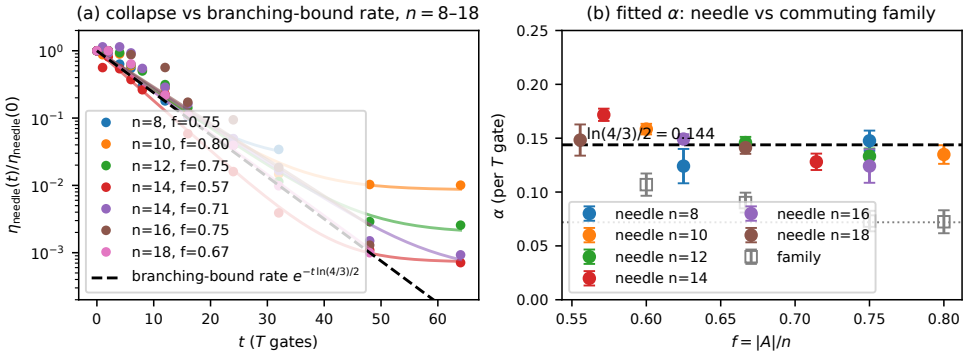

Fig. 4 The best-setting fraction decays at a rate close to the branching benchmark across $n = 8$–18. (a) Normalized $\eta_{\text{needle}}(t)$ for $f > 1/2$ with fits to Eq. (7); the dashed line is the benchmark $e^{-t\ln(4/3)/2}$. (b) Fitted α for the twelve one-shot sweeps (circles) and for greedily constructed commuting families (open squares). Error bars are fit standard errors; each point averages eight circuits (four at $n = 18$), and the rates carry a fit-window dependence of order 15% (Appendix C).

($f = 0.67, 0.75$). At $f \geq 0.75$ this is close to $\alpha/2$; at $f \leq 0.67$ it exceeds $\alpha/2$ by 30–50%. Because the construction is greedy, these values do not establish the optimum over commuting families.

The redistribution is visible directly in the per-setting distribution. At $n = 10$, $f = 0.8$, the fraction of total sensitivity captured by the union of the 64 highest-sensitivity settings falls from 1 at $t = 0$ to 0.015 at $t = 32$, while m_{eff} grows to about 2×10^4 of the $4^a - 1$ settings (Appendix C). A non-adaptive control that divides a fixed Pauli budget uniformly over all settings is flat in t, as Proposition 6 requires. As a descriptive statistic, pooled per-realization data give slopes between -0.63 and -0.75 for $\log \eta$ against $\log m_{\text{eff}}$ across eight data sets; the slope depends on the fit window and is not an independent scaling law. An extensive-generator sweep at $n = 8$ gives $\eta(0) = 0.130$ at $f = 0.75$, close to the multiplicity prediction $1/n = 0.125$, and $\alpha = 0.170 \pm 0.014$ at $f = 0.62$; the $f = 0.75$ trace lacks the dynamic range for a reliable fit.

6.6 The three regimes survive observation noise

The finite-shot protocol (Appendix E) reproduces the three regimes. At $n = 10$, $f = 0.5$, and 10^4 shots per setting, identification succeeds with probability 0.93 at $t = 6$. At the Clifford point it approaches instead the ceiling of the blind arg-max classifier, $\mathbb{E}[\text{rank}\,\Gamma]/m + (1 - \mathbb{E}[\text{rank}\,\Gamma]/m)/m \approx 0.38$. At $t = 32$ identification remains successful while the estimation error is 3.3 times that at $t = 6$ (4.1 times at 10^5 shots). A 2% readout error changes the identification probability at $t = 6$ from 0.93 to 0.89. For $t \geq 6$ the identification threshold is approximately flat, $S^*_{\text{id}} \approx 2.3$–$2.7 \times 10^3$ shots per setting, consistent with its dependence on the full response norm $\sum_j s_j^2$ rather than on the largest component.

Charging all calibration shots changes the picture for identification. When both tasks are restricted to K random Pauli settings, at $S = 4096$ and $t = 16$ the identification probability rises from 0.05 at $K = 16$ through 0.29(4) at $K = 128$ and 0.75(5) at $K = 512$ to 0.98(2) at $K = 1023$, the full basis (480 trials per point). Reliable identification therefore requires calibrating essentially the full Pauli basis, whose size is exponential in a. In this per-Pauli measurement model the budget is calibration-dominated (95% of B at every K), classical post-processing is negligible ($\leq 0.2\,\mathrm{ms}$ per identification), and each Pauli setting costs one layer of single-qubit rotations. The identification/estimation asymmetry is thus a statement about how sensitivity distributes across settings at a fixed calibrated pool, not a total-resource complexity separation.

6.7 Interleaved encoding protects learnability

Re-imprinting the signal between L scrambling blocks changes the $t = 0$ starting point and the decay. At the highest tested access, $f = 0.8$, the expected value $\eta(0) = 1/L$ holds in 22 of 24 seeds across $L = 2, 4, 8$. At lower access, degenerate realizations with merged or invisible copies dominate and $\eta(0)$ deviates from $1/L$ in up to 8 of 8 seeds (Appendix C). Interleaving slows the collapse, with α decreasing from 0.227 ± 0.028 to 0.070 ± 0.012 as f grows at $n = 10$, and leaves saturation floors 7.7 and 17 times the one-shot values at $f = 0.6$ and 0.8 (and 441 times at $f = 0.5$, where the one-shot floor is nearly zero). A persistent signal is qualitatively harder to hide.

7 Other scrambler families

The doped-Clifford results raise a question of scope: which features are specific to stabilizer magic, and which belong to scrambling more generally? Figure 5 and Table 3 answer it for three further families. Values are quoted for $n = 10$, with $n = 12$ in parentheses where it differs.

7.1 Finite depth: the light cone sets the information ceiling

In a one-dimensional brickwork circuit the far block receives no information until the light cone of the signal qubit reaches it. F_Q^A vanishes to numerical precision for every depth $D < n - a$ and is nonzero at $D = n - a$, for every a and at both sizes [Fig. 5(a)]. It then rises to the global-Haar value within $D \approx 12$–24 layers.

With Haar-random gates the best-setting fraction of the near block decays at 0.27–0.30 per layer (0.26–0.29 at $n = 12$) and reaches the global-Haar value, $\eta_{\mathrm{needle}} \approx 0.009$–$0.012$, by $D \approx 20$; the Gram matrix reaches full rank at $D = 4$–8. With Clifford gates none of this happens. In every realization and at every depth η_{needle} is either 0 or 1, and the rank fraction saturates near 0.41, 0.78, and 0.96 for $f = 0.5$, 0.6, and 0.8 [Fig. 5(b)]. Depth supplies the information ceiling. It does not by itself lower the readout ceiling or fill the rank.

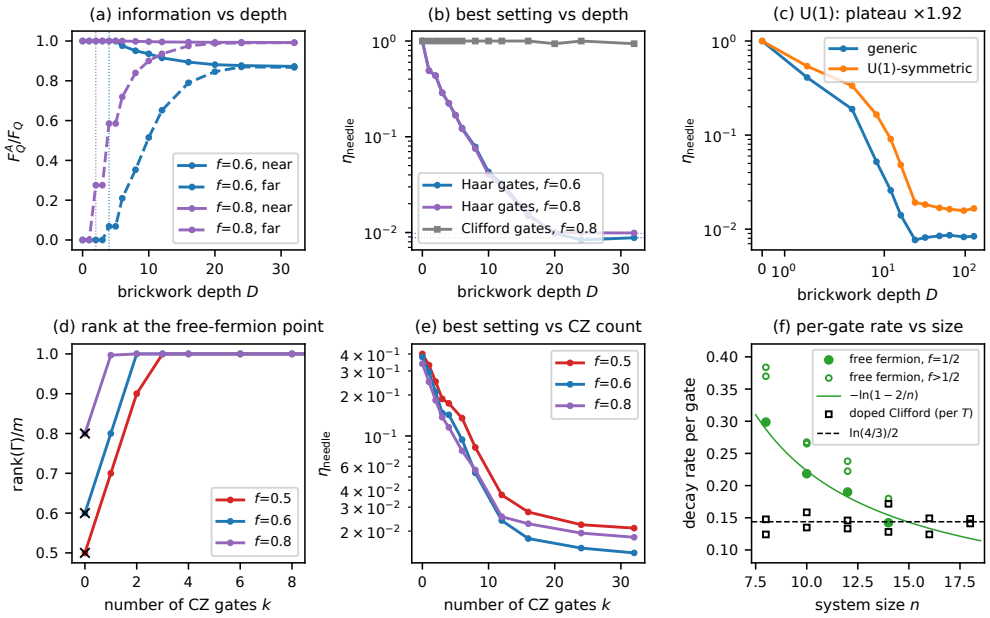

Fig. 5 Readout concentration tracks non-Clifford and non-Gaussian resources rather than depth or symmetry. All panels $n = 10$ except (f). (a) Brickwork with Haar gates: information reaches the far block (dashed) at depth $n - a$ (dotted lines) and then approaches the deep-circuit value. (b) The best-setting fraction decays with depth for Haar gates and stays at its all-or-nothing value for Clifford gates. (c) A U(1) symmetry doubles the late-depth plateau and slows the approach. (d) At the free-fermion point the Gram rank fraction equals f (crosses, Proposition 7); two or three CZ gates fill it. (e) CZ gates disperse the sensitivity. (f) Floor-subtracted decay rate per CZ gate for the free-fermion family against the benchmark of Eq. (6), and fitted rate per T gate for doped Clifford circuits against $\ln(4/3)/2$.

7.2 U(1) symmetry: a constant factor

For the signal generator X_0, which changes the conserved charge by one unit, a U(1) symmetry does not reduce the available information: the late-depth values of F_Q^A/F_Q are 0.50, 0.87, and 0.99 for $f = 0.5$, 0.6, and 0.8 with and without the symmetry. What the symmetry changes is the set of settings that can carry sensitivity. Because $\partial_\theta \rho_A$ connects neighboring charge sectors of A, only Pauli strings with an odd number of X or Y letters contribute. These are 2^{2a-1} of the $4^a - 1$ settings, and the sensitivity on all others vanishes identically. Accordingly m_{eff} is halved (ratio 0.49–0.50) and the late-depth best-setting fraction is about twice the unconstrained value: the ratio is 1.91–1.97 at $n = 10$ and 1.91–2.29 at $n = 12$, where eight circuits per cell give a larger scatter [Fig. 5(c)]. The approach to the plateau is slower, 0.19–0.21 per layer against 0.30–0.35 without the symmetry (0.19–0.22 against 0.26–0.29 at $n = 12$). A symmetry thus raises the readout ceiling by a constant factor and leaves its exponential dependence on a unchanged.

Table 3 The same inversion in two families, and its absence in two others. "Resource" is the quantity along which the best-setting fraction decays; rates in different units are not comparable.

Family	Resource	Zero-resource point	Decay of η_{needle}	Deep limit
Doped Clifford	T gates	$\eta \in \{0, 1\}$, rank deficient	0.142 per T, size independent	global Haar
Free fermion + CZ	CZ gates	$\eta \approx 0.35$, $\operatorname{rank}\Gamma/m = f$	$\approx -\ln(1 - 2/n)$ per CZ	1.6–1.9× Haar
Haar brickwork	layers	(trivial at $D = 0$)	0.27–0.30 per layer	global Haar
U(1) brickwork	layers	(trivial at $D = 0$)	0.19–0.21 per layer	$\approx 2\times$ Haar
Clifford brickwork	layers	—	none	$\eta \in \{0, 1\}$

7.3 Free fermions: the inversion recurs with non-Gaussian gates

At the free-fermion point the Gram rank equals $2a$ in every realization at $n = 8$, 10, 12, and 14, so the bound of Proposition 7 is saturated and $\operatorname{rank}\Gamma/m = f$ [Fig. 5(d)]. The best-setting fraction is large, $\eta_{\text{needle}} = 0.34$–$0.40$ (0.30–0.37 at $n = 12$), while the global-Haar value falls from about 0.010 to about 0.003 between the two sizes. The available information is $F_Q^A/F_Q = 0.73$, 0.79, and 0.94 for $f = 0.5$, 0.6, and 0.8, which differs from the Haar values. Generic matchgates are not Clifford gates, yet the sensitivity stays concentrated.

Non-Gaussian gates change both properties at once. Two or three CZ gates bring the Gram matrix to full rank, and the best-setting fraction decays with the number of CZ gates toward a plateau that is 1.6–1.9 times the global-Haar value, the factor again reflecting a selection rule, here for fermion parity [Fig. 5(e)]. This is the inversion found for doped Clifford circuits, with the non-Gaussian gate count in the role of t.

The decay rate per gate behaves differently from the doped-Clifford case. At $f = 1/2$ the floor-subtracted rate over $k \leq 8$ is 0.299, 0.218, 0.190, and 0.142 per CZ gate at $n = 8$, 10, 12, and 14, within 8% of the benchmark $-\ln(1 - 2/n) = 0.288$, 0.223, 0.182, and 0.154 of Eq. (6). At larger f the measured rate exceeds the benchmark by 17%–33%. In both cases it falls with n, whereas the rate per T gate of doped Clifford circuits shows no systematic size dependence over $n = 8$–18 [Fig. 5(f)].

8 Experimental feasibility

The protocol uses only operations that are native on the platforms named in the Introduction; this section translates the resource counts of Sec. 6 into laboratory quantities. Repetition rates below are stated assumptions used for conversion, not measured device parameters.

Mapping. The scrambler is a sequence of Clifford layers with t single-qubit T gates interleaved; scrambling circuits of this type have been run on trapped-ion processors [30], and Clifford layers plus single-qubit phase gates are standard on superconducting and Rydberg-array hardware. The calibration stage requires the ability to apply a known small rotation $e^{\mp i\delta G_\mu}$ for each dictionary generator G_μ before the scrambler; this is the probe-access assumption. Readout in every stage is one layer of single-qubit basis rotations on the accessible qubits followed by computational-basis measurement, so the entangling depth of the readout is zero. The interleaved protocol repeats the encoding between scrambling blocks and needs no additional control.

Gate budget. In the window at $n = 10$, $f = 0.5$, the number of non-Clifford gates is $t = 6$–16, and the twelve decay fits span $t \leq 64$. The T count sets the magic axis

directly; the Clifford layers are the dominant gate count and are compiled natively on all three platforms.

Shot budget and wall-clock time. The binding resource is the number of Pauli settings that must be calibrated. Table 4 converts the charged budget $B = SK(2m+2)+S$ of the same-budget experiment ($n = 10$, $a = 5$, $m = 20$, $S = 4096$) into wall-clock time at three assumed repetition rates; the conversion is reproduced by `feasibility_units.py` in the package. Reliable identification needs K near the full basis of $4^a - 1 = 1023$ settings, which at 10 kHz is a five-hour calibration. The setting count grows as 4^a, so the per-Pauli protocol is practical for accessible blocks of $a \lesssim 6$ qubits; larger blocks require structured readout families, whose mean captured sensitivity is magic-independent (Proposition 6) but whose per-realization performance is open.

Noise requirements. At 10^4 shots per setting a readout bit-flip probability of 2% per qubit lowers the identification probability in the window from 0.93 to 0.89, and the identification threshold is 2.3–2.7×10^3 shots per setting for $t \geq 6$. Estimation deep in the magic phase ($t = 32$) costs 3.3 times the shots of the window at equal accuracy. Gate errors in the scrambler itself are not modeled; because the protocol never inverts the scrambler, coherent errors act as an unknown change of U and are absorbed by calibration, whereas incoherent errors reduce F_Q^A and are not studied here.

Observables that test the predictions. Four predictions are measurable with the single-qubit readout above. (i) The best-setting sensitivity $\max_j s_j^2$, obtained from calibrated Pauli expectations at $\theta = \pm\delta$, decays with the T count at a rate close to 0.14 per gate at $f > 1/2$, while the sum $\sum_j s_j^2$ over all settings stays flat. (ii) The identification success probability rises from the Clifford-point ceiling ≈ 0.38 to above 0.9 once $t \gtrsim 6$ at $f = 0.5$, with a shot threshold that does not grow with t. (iii) At $f = 0.8$ and $t = 0$ the interleaved protocol gives a best-setting fraction of $1/L$. (iv) The Gram matrix of the calibrated responses becomes full rank at $t \approx 6$–12 for $f = 0.5$, with a condition number that falls toward unity deep in the doped phase.

Other platforms. Nothing in the protocol depends on the physical encoding beyond the availability of Clifford layers, single-qubit phase gates, and single-qubit readout rotations, so the same counts apply to superconducting, trapped-ion, and neutral-atom processors with the repetition rates of Table 4 as the only platform-specific input.

Table 4 Charged shot budget of the same-budget identification experiment and its wall-clock time at assumed repetition rates ($n = 10$, $a = 5$, $m = 20$, $S = 4096$ shots per setting). Calibration accounts for 95% of B at every K.

K	B (shots)	1 kHz	10 kHz	100 kHz
16	2.76×10^6	46 min	4.6 min	28 s
128	2.20×10^7	6.1 h	37 min	3.7 min
512	8.81×10^7	1.0 d	2.4 h	15 min
1023	1.76×10^8	2.0 d	4.9 h	29 min

9 Scope and limitations

The statements above hold under the following conditions, which we collect here rather than repeat at every claim.

- *Ensemble.* All results refer to the sampled t-doped Clifford ensemble with single-qubit T gates. Grouping T gates into parallel layers of four leaves the per-T rate similar at $n = 8$, while a CCZ gate gives a larger per-gate rate; these checks establish convention dependence, not universality. Continuous chaotic dynamics and hardware noise beyond readout bit flips are not studied.
- *Other families.* The brickwork, U(1), and free-fermion results are for one-dimensional circuits at $n = 10$ and 12 (free fermion: $n = 8$–14) with 8 or 16 circuits per cell. Equation (6) is an operator-level benchmark derived for Haar-random Gaussian blocks; the fitted rates agree with it within 8% at $f = 1/2$ and exceed it by 17%–33% at larger f. Generators that commute with the conserved charge are not probed. The reading that concentration is governed by the departure from free operator dynamics is a hypothesis supported by two families, not a theorem.
- *Finite size.* The statevector results span $n = 8$–18; the window $t \in [6, 16]$, the floors, and the agreement between α and $\ln(4/3)/2$ are finite-size statements. The occupancy model attributes the agreement to the partner multiplicity realized in these sweeps, and the asymptotic needle rate is open. The branch-tree results at $n \leq 1000$ verify only the global operator benchmark.
- *Proxy.* η_{needle} is a Pauli-readout concentration proxy, not an achievable Fisher fraction over arbitrary measurements. It differs from the exact best-Pauli Fisher fraction by at most 3% in the $n = 10$ sweeps. Its operational meaning is checked only by the finite-shot protocol at $n = 10$.
- *Benchmark versus theorem.* Equation (4) is a theorem about global operator purity. Of the two steps needed to transfer it to the accessible spectrum, Proposition 5 settles the visibility filter; the state-level partner factor of Proposition 4 is open. Collision corrections are bounded only empirically.
- *Rank versus stability.* Full rank is a structural statement. Measured conditioning tracks the analytic scale $2g$ at $n \leq 14$ deep in the doped phase; an asymptotic conditioning theorem is not provided.
- *Decoupling, not a wall.* The $f = 1/2$ crossover is a decoupling reference for the one-shot ensemble. No-cloning motivates this scale for recovering arbitrary quantum information but imposes no exact threshold for estimating a scalar; the $f = 0.3$ data show reduced but nonzero information.
- *Measurement class.* No lower bound over arbitrary POVMs, adaptive collective measurements, kernels, or variational decoders is claimed, and none is possible while F_Q^A persists (Proposition 6). The commuting-family result concerns a greedy construction, not the optimum.
- *Resource accounting.* The flat identification threshold is conditional on calibrating the full Pauli basis. Under a common budget neither task enjoys a subexponential guarantee from our data.

- *Uncertainties.* Fit standard errors are conditional on the sampled circuits; the bootstrap interval is the ensemble statement; the window dependence of order 15% is a systematic.

10 Discussion

The central lesson is an inversion within the specified calibration-and-readout model. Magic makes dictionary responses more typically full rank while dispersing sensitivity across Pauli settings. At the Clifford point the ensemble mean of the identifiability matrix is already Haar-like, but individual responses are sparse; rank filling is therefore a fluctuation effect, not a change of the mean. Injecting non-Clifford resources concentrates the response ensemble and spreads the Pauli sensitivity, producing the empirical $e^{-0.14t}$ decay over the simulated range. Calibration-assisted meta-metrology consequently has an intermediate-magic window. Its lower edge is set by response concentration and its budget-dependent upper edge by Pauli-readout sensitivity.

The other families show what this inversion depends on. It does not depend on depth: Clifford brickwork of any depth keeps every realization at the all-or-nothing point. It is not removed by a symmetry: a conserved charge halves the set of active settings and changes nothing else. And it is not tied to stabilizer magic as such: free-fermion circuits built from non-Clifford matchgates remain rank deficient and concentrated until non-Gaussian gates are added, at which point the rank fills and the sensitivity disperses. What the two zero-resource points share is that Heisenberg-evolved operators stay inside a small, basis-aligned set—a single Pauli string for Clifford circuits, linear combinations of Majorana operators for Gaussian circuits. We therefore conjecture that the readout ceiling is governed by the departure from such free operator dynamics, of which stabilizer magic and fermionic non-Gaussianity are two instances. The size dependence of the per-gate rate, constant for T gates and of order $1/n$ for CZ gates on free fermions, shows that the two instances differ quantitatively.

This picture is a measurement-side analogue of barren-plateau behavior [11–13]. In a barren plateau the gradient of a training cost concentrates exponentially; here the sensitivity available to any one fixed readout setting concentrates exponentially even though F_Q^A stays large. The analogy is structural, not mechanistic: the barren-plateau results concern optimization landscapes, whereas Proposition 6 shows that our effect lives in the selection of a setting from an exponentially large family whose mean is flat. The observed functional form resembles the decoder-fidelity decay found by Leone *et al.* [8], and Pauli splitting offers a possible common mechanism, consistent with approximate-design convergence [56] and magic-spreading studies [52, 53].

Several directions follow. The state-level partner factor in the operator-to-accessible transfer is the sharpest open theoretical step; its resolution would turn the benchmark into a theorem for the needle rate. Concentration below $f = 1/2$, finite-shot scaling beyond $n = 10$, and the scaling of the window edges with n are numerical questions within reach of the present pipelines. Optimization over structured measurement families, and the implementation cost of the optimal measurement, are the correctly posed versions of the measurement-complexity question. Interleaved sweeps

at $L = 2, 4, 8$ reject the naive $\alpha_{\text{int}} \propto 1/L$ picture and instead show elevated finite-size floors whose mechanism is open. Hardware realization remains to be tested.

11 Conclusion

We introduced a calibration-assisted meta-metrology problem: classify an unknown encoding generator within a known probe dictionary from subsystem responses alone, then estimate the encoded value. To our knowledge, the separation of this identification task from the concentration of sensitivity across readout settings has not been formulated before. In the sampled doped-Clifford ensemble the two tasks are governed by different resources. Access fraction controls a one-shot decoupling crossover near $f = 1/2$, while magic fills the response rank and disperses sensitivity, so that the best-Pauli fraction decays at $\alpha = 0.142 \pm 0.014$ per T gate over $n = 8$–18 (bootstrap interval $[0.130, 0.148]$, window-dependent at the 15% level). Compared with learned-decoder studies of the same scramblers [8, 9], which report a non-stabilizerness-dependent crossover for an inversion task, the present results isolate the part of that difficulty that is due to readout concentration alone, with the ensemble-mean identifiability matrix proved to be magic-independent.

The same inversion appears in free-fermion scramblers doped with non-Gaussian gates, where the response rank is bounded by twice the subsystem size (Proposition 7), and it is absent in Clifford circuits of any depth; a U(1) symmetry changes the readout ceiling only by a factor of two. These results place calibration-assisted sensing in many-body devices between two established pictures. Chaos-assisted metrology is right that scrambling preserves information under partial access, and barren-plateau results are right that unknown scramblers resist variational inversion; what decides whether a hardware-native readout can use the preserved information is how far the dynamics depart from a free, classically tractable structure—measured by magic for Clifford-based scramblers and by non-Gaussianity for free-fermion ones. A sensing protocol designed without this distinction will either over-provision access or under-provision calibration, since the two costs scale with different parameters. The exact Clifford endpoint, the magic-independent mean, the visibility-filter theorem, and the flatness of every fixed readout family (Lemma 2, Propositions 4–7) provide a baseline against which hardware implementations and alternative scrambler ensembles can be compared. Closing the state-level partner factor in the operator-to-accessible transfer, and extending the finite-shot protocol beyond $n = 10$, are the next steps toward a complete theory of readout under scrambling.

Statements and Declarations

Funding. This work received no external funding.

Competing interests. Both authors are employees of QIRI (Quantum Integrated Research Institute Inc.); the authors declare no other competing interests.

Author contributions. H.W. conceived the study, developed the theory, performed the numerical calculations, and wrote the manuscript. T.T. secured the

internal funding and supervised the project. Both authors reviewed and approved the manuscript.

Generative-AI disclosure. The authors disclose that a large language model, Claude (Anthropic), was used extensively under the authors' direction in preparing this work: drafting and revising the prose of all sections and appendices; writing and refactoring the simulation, verification, and figure-generation code in the reproduction package; assisting with the derivation and presentation of the proofs in Appendices A, B, D, and F; and performing internal adversarial-review passes and cross-checks of the cited literature. The research questions, the model, and the claims were set by the authors; every numerical result is regenerated from fixed seeds by the gated driver `reproduce.py`; and the authors reviewed, verified, and edited all AI-assisted content and take full responsibility for it.

Data availability. The data that support the findings of this article were generated by numerical simulations. The simulation code (dense statevector for $n \leq 8$, matrix-free statevector for $n \leq 18$, and the symplectic branch-tree simulator for $n \leq 1000$), the pipeline for the other scrambler families, the verification suites for Lemma 2 and Propositions 4–7, the figure-generation scripts, the 39 data sets underlying every number quoted in the text, and a one-command gated reproduction driver (`reproduce.py`, over 140 checks that fail on any mismatch) are publicly available on Zenodo [57]; the development repository is at https://github.com/deeptell-inc/qnoise-pack. All random seeds are fixed in the code, so reruns are deterministic.

Appendix A Proof of Lemma 2 (Weingarten constant)

For U Haar (or any exact unitary 2-design) and traceless X, Y on $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$,

$$\mathbb{E}_U \operatorname{Tr}_A\big[\operatorname{Tr}_{\bar{A}}(UXU^\dagger)\operatorname{Tr}_{\bar{A}}(UYU^\dagger)\big] = \tilde{g}\operatorname{Tr}(XY), \qquad \tilde{g} = \frac{d\,d_A - d_{\bar{A}}}{d^2-1}. \tag{A1}$$

Inputs. (I1) The second-moment (Schur–Weyl) twirl identity for the Haar measure—equivalently any unitary 2-design—which is the two-copy special case of the Weingarten calculus [58]: for any operator M on $\mathcal{H}^{\otimes 2}$,

$$\mathbb{E}_U\big[(U\otimes U)\,M\,(U^\dagger\otimes U^\dagger)\big] = c_1\mathbb{1} + c_2 S, \tag{A2}$$

where S is the swap and c_1, c_2 are fixed by taking $\operatorname{Tr}(\cdot)$ and $\operatorname{Tr}(S\,\cdot)$ of both sides, using $\operatorname{Tr}\mathbb{1} = d^2$, $\operatorname{Tr} S = d$, $\operatorname{Tr}(S(A\otimes B)) = \operatorname{Tr}(AB)$:

$$c_1 d^2 + c_2 d = \operatorname{Tr} M, \qquad c_1 d + c_2 d^2 = \operatorname{Tr}(SM). \tag{A3}$$

(I2) The elementary identities $\operatorname{Tr}(S(A\otimes B)) = \operatorname{Tr}(AB)$ (write $S = \sum_{jk}|jk\rangle\langle kj|$ and contract) and $S = S_A \otimes S_{\bar{A}}$ (the swap factorizes over the tensor factors of each copy).

Step 1 (rewrite the LHS as a two-copy trace). In index form, $\operatorname{Tr}_A[\operatorname{Tr}_{\bar{A}}(M)\operatorname{Tr}_{\bar{A}}(N)] = \sum_{aa'ee'} M_{(ae),(a'e)} N_{(a'e'),(ae')}$, which is exactly

$\mathrm{Tr}[(M \otimes N)\, W]$ with $W = S_A \otimes \mathbb{1}_{\bar{A}\bar{A}}$ (the A indices are swapped between the copies; the $\bar{A}$ indices close within each copy). Apply this with $M = UXU^\dagger$, $N = UYU^\dagger$.

Step 2 (twirl). Solving Eq. (A3) for $M = X \otimes Y$ with $\mathrm{Tr}\, X = \mathrm{Tr}\, Y = 0$ gives $\mathrm{Tr}\, M = 0$, $\mathrm{Tr}(SM) = \mathrm{Tr}(XY)$, hence

$$c_1 = -\frac{\mathrm{Tr}(XY)}{d(d^2-1)}, \qquad c_2 = \frac{\mathrm{Tr}(XY)}{d^2-1}. \tag{A4}$$

Therefore $\mathbb{E}[\cdot] = c_1\, \mathrm{Tr}\, W + c_2\, \mathrm{Tr}(SW)$.

Step 3 (the two traces of W). $\mathrm{Tr}\, W = \mathrm{Tr}_{AA}[S_A] \cdot \mathrm{Tr}_{\bar{A}\bar{A}}[\mathbb{1}] = d_A \cdot d_{\bar{A}}^2$. Using (I2), $SW = (S_A S_A) \otimes S_{\bar{A}} = \mathbb{1}_{AA} \otimes S_{\bar{A}}$, so $\mathrm{Tr}(SW) = d_A^2 \cdot d_{\bar{A}}$.

Step 4 (collect). With $d = d_A d_{\bar{A}}$,

$$\mathbb{E} = \frac{\mathrm{Tr}(XY)}{d^2-1}\left(d_A^2 d_{\bar{A}} - \frac{d_A d_{\bar{A}}^2}{d}\right) = \frac{\mathrm{Tr}(XY)}{d^2-1}\left(d\, d_A - d_{\bar{A}}\right), \tag{A5}$$

which is $\tilde{g}\, \mathrm{Tr}(XY)$; the equivalent form $g = d(d_A^2-1)/(d_A(d^2-1))$ follows from $d_{\bar{A}}(d_A^2-1) = d\, d_A - d_{\bar{A}}$ upon substituting $d_{\bar{A}} = d/d_A$. □

Step 5 (application to the Gram matrix). For Hermitian responses $R_\mu = -i\, \mathrm{Tr}_{\bar{A}}(UC_\mu U^\dagger)$ with $C_\mu = [G_\mu, \rho_0]$ anti-Hermitian, $\langle R_\mu, R_\nu \rangle = -\mathbb{E}\, \mathrm{Tr}_A[\mathrm{Tr}_{\bar{A}}(UC_\mu U^\dagger)\mathrm{Tr}_{\bar{A}}(UC_\nu U^\dagger)] = -\tilde{g}\, \mathrm{Tr}(C_\mu C_\nu)$. For pure ρ_0 ($\rho_0^2 = \rho_0$), expanding the commutators and using cyclicity,

$$-\mathrm{Tr}(C_\mu C_\nu) = \langle \{G_\mu, G_\nu\} \rangle_{\rho_0} - 2\langle G_\mu \rangle_{\rho_0} \langle G_\nu \rangle_{\rho_0}. \tag{A6}$$

On $|0 \ldots 0\rangle$ with the dictionary $\{X_i, Y_i\}$: all single-generator expectations vanish, $\{X_i, Y_i\} = 0$ on the same site, different sites give products of vanishing single-site expectations, and $\{X_i, X_i\} = 2 \cdot \mathbb{1}$; hence $-\mathrm{Tr}(C_\mu C_\nu) = 2\delta_{\mu\nu}$, giving Eq. (3). Numerically, the predicted $\lambda_{\min}$ plateaus $2g = 0.0624,\ 0.125,\ 0.500$ for $a = 5, 6, 8$ at $n = 10$ match the observed $0.047,\ 0.110,\ 0.485$ ($\lambda_{\min}$ biased slightly below the mean, as it must be).

Appendix B Exact Clifford-point solution (Proposition 4)

Inputs. (I1) The stabilizer expansion of a stabilizer state, $\hat{\rho} = 2^{-n} \sum_{S \in \mathcal{S}} S$, and the symplectic (GF(2)) representation of Pauli conjugation by Clifford unitaries; both are standard [59, 60]. (I2) The SLD form of the quantum Fisher information, $F_Q = 2 \sum_{jk} |\langle j|\partial_\theta \rho|k\rangle|^2/(\lambda_j + \lambda_k)$ [50]. (I3) For two Pauli operators, $[\hat{P}, S] = 0$ if they commute and $[\hat{P}, S] = 2\hat{P}S$ if they anticommute (from $\hat{P}S = \pm S\hat{P}$).

Let U be Clifford, $\hat{P} = UG_\mu U^\dagger$, and $\mathcal{S} = U\langle Z_1, \ldots, Z_n\rangle U^\dagger$ the stabilizer group of $|\varphi\rangle = U|0 \ldots 0\rangle$. Expanding $\hat{\rho}$ by (I1) and applying (I3) term by term,

$$R = 2^{1-a} \sum_{S \in \mathcal{P}} (\pm)\, (\hat{P}S)_A, \qquad \mathcal{P} = \{S \in \mathcal{S} : \{S, \hat{P}\} = 0,\ \mathrm{supp}(\hat{P}S) \subseteq A\}. \tag{B7}$$

Distinct partners give distinct Pauli strings on A (a stabilizer group contains no $\pm$-pairs), hence orthogonal contributions and no cancellation. Writing each candidate combination of stabilizer generators as $c \in \mathbb{F}_2^n$, membership in $\mathcal{P}$ is an affine linear system over GF(2): $2(n-a)$ support constraints (the $\bar{A}$ symplectic components of the combination must equal those of $\hat{P}$) plus one anticommutation constraint. Thus $N = |\mathcal{P}| \in \{0\} \cup \{2^k\}$ is exactly computable per instance by Gaussian elimination. With $s_A = \log_2 |\{S \in \mathcal{S} : \mathrm{supp}(S) \subseteq A\}|$ (the homogeneous version of the same system), the reduced state has flat spectrum of rank 2^{a-s_A} and the N partner Paulis carry equal sensitivity weight, giving

$$\eta_{\mathrm{needle}}(t{=}0) = \min(1,\ 2^{s_A}/N)\,. \tag{B8}$$

Explicitly: the reduced state of a stabilizer state is a normalized projector, here of rank 2^{a-s_A} with flat eigenvalues $\lambda = 2^{s_A-a}$. *Scope of the derivation:* the following computation assumes a proper subsystem, $a < n$, and that the response is supported inside the state's support projector; support–kernel cross terms of the SLD sum have denominator $\lambda+0 = \lambda$ rather than 2λ and are not covered ($N = 0$ gives $\eta = 0$ trivially). At the boundary $a = n$ the containment assumption can fail — e.g. $n = a = 1$, $U = \mathbb{1}$, $G = X$ connects support to kernel and the formula below would give $F_Q = 2$ against the true $F_Q = 4$ (the min-clip still returns the correct $\eta = 1$ there); a fully general block-decomposed derivation is left as a technical step. Within this scope — which covers all 120 tested instances ($a < n$, containment verified numerically) — every nonvanishing term of the SLD sum (I2) has denominator $\lambda_j + \lambda_k = 2 \cdot 2^{s_A-a}$, so

$$F_Q^A = \frac{2}{2 \cdot 2^{s_A-a}} \sum_{jk} |\langle j|\partial_\theta \rho_A|k\rangle|^2 = 2^{a-s_A}\,\mathrm{Tr}[(\partial_\theta \rho_A)^2] = 2^{a-s_A} \cdot 2^{-a} \sum_k s_k^2 = 4N\, 2^{-s_A}, \tag{B9}$$

using Parseval ($\partial_\theta \rho_A = 2^{-a} \sum_k s_k P_k$, so $\mathrm{Tr}[(\partial_\theta \rho_A)^2] = 2^{-a} \sum_k s_k^2$) and $\sum_k s_k^2 = 4N$ (N partners, each with $s = \pm 2$). Since $\max_j s_j^2 = 4$, the ratio yields the formula. Verified in all 120 tested instances ($n = 12$, $a \in \{6, 8, 9\}$, 40 random Cliffords each; the formula and the measurement share no code path). Consistency with Lemma 2: $\|R\|^2 = 2^{2-a}N$ and $\mathbb{E}[N] \approx 2^{2a-n-1}$ give $\mathbb{E}\|R\|^2 \approx 2^{a-n+1} \approx 2g$ for $d \gg 1$.

Appendix C Extended data

Finite-size window at $n = 10$ (one-shot, 16 realizations per cell, dictionary $m = 2n = 20$):

a (f)	t	rank(Γ)/m	F_Q^A/F_Q	η_{needle}
3 (0.3)	0	0.03	0.06	0.062
3 (0.3)	16	1.00	0.03	0.354
5 (0.5)	0	0.34	0.19	0.187
5 (0.5)	6	0.98	0.47	0.424
5 (0.5)	32	1.00	0.50	0.014
8 (0.8)	0	0.99	0.94	0.937
8 (0.8)	32	1.00	0.99	0.014

The Gram rank uses an absolute eigenvalue tolerance ($2g \times 10^{-6}$) anchored to the analytic mean, because a $\lambda_{\max}$-relative tolerance counts floating-point noise as rank when the Gram matrix is numerically zero (at $t = 0$, $a = 3$ it would report 0.54). The value 0.03 agrees with the exact visibility prediction $P[N \geq 1] = 0.031$. Below the wall ($f = 0.3$) the mean F_Q^A and η reflect all-or-nothing per-realization visibility (Proposition 4) and fluctuate accordingly; their identifiability content is carried by the rank column.

Interleaved fits and the $\eta(0) = 1/L$ check: the identity $\eta(0) = 1/L$ presumes a nondegenerate realization in which all L signal copies are visible from A and contribute to distinct settings. Counting seeds with $|\eta(0) - 1/L| > 0.02$ per (L, a) cell at $n = 10$ (8 seeds each): $L = 2$: 5/8 (a=6), 1/8 (a=8); $L = 4$: 8/8 (a=3, all copies invisible), 7/8 (a=5), 5/8 (a=6), 1/8 (a=8); $L = 8$: 6/8 (a=6), 0/8 (a=8). Deviating seeds show merged copies ($\eta(0)$ at a multiple of $1/L$) or invisible copies; the identity holds cleanly only in the high-access cells ($a = 8$: 22 of 24 seeds). Fit results ($L = 4$): $\alpha = 0.227 \pm 0.028$, 0.147 ± 0.023, 0.070 ± 0.012 at $n = 10$ for $f = 0.5, 0.6, 0.8$; floor ratios to one-shot 441, 7.7, 17 at $f = 0.5, 0.6, 0.8$ (the $f = 0.5$ ratio is large because the one-shot floor is 1.4×10^{-4}).

Extensive generator ($G = \sum_i X_i$, $n = 8$, one-shot): $\eta(0) = 0.130$ at $f = 0.75$ versus the needle-multiplicity prediction $1/n = 0.125$; $\alpha = 0.170 \pm 0.014$ at $f = 0.62$; at $f = 0.75$ the initial spread leaves insufficient dynamic range for a reliable fit ($\alpha = 0.061 \pm 0.022$ with a large floor, excluded).

Gram conditioning across sizes. The finite-shot stability of the identification inverse problem is controlled by the smallest Gram eigenvalue and the condition number, not by rank alone. We measured the full Gram spectrum of the $m = 2n$ dictionary at $n = 8, 10, 12, 14$ (8 realizations per cell; `gram_conditioning_n8_14.json`). Medians of $\lambda_{\min}/2g$ (where $2g$ is the analytic ensemble mean of Lemma 2) and of the condition number $\kappa = \lambda_{\max}/\lambda_{\min}$:

		$f = 0.5$		$f \approx 0.75$	
n		t=6	t=32	t=6	t=32
8	$\lambda_{\min}/2g$	0.18	0.59	0.72	0.89
	κ	11.5	2.6	1.8	1.24
10	$\lambda_{\min}/2g$	0.14	0.76	0.88	0.97
	κ	10.5	1.7	1.29	1.07
12	$\lambda_{\min}/2g$	0.22	0.86	0.88	0.98
	κ	8.2	1.33	1.29	1.04
14	$\lambda_{\min}/2g$	0.22	0.92	0.88	0.99
	κ	9.6	1.18	1.29	1.02

(The $f \approx 0.75$ column uses $a = 3n/4$ rounded, i.e., $f = 0.75, 0.8, 0.75, 0.71$; at $t = 0$ and high access the visible responses are exactly orthogonal, $\kappa = 1.00$, as Proposition 4 predicts.) Two finite-size trends emerge. Deep in the doped phase the spectrum *concentrates at the analytic mean*: $\lambda_{\min}/2g$ rises with n (from 0.59 to 0.92 at $f = 0.5$, $t = 32$) and κ falls toward 1—within the sampled range, conditioning imposes no penalty beyond the visibility scale $2g \approx 2^{1-(n-a)}$ that already sets the mean response. At the window edge ($t = 6$, $f = 0.5$), by contrast, $\kappa \approx 8$–11 roughly independent of n: the onset of identifiability carries a finite conditioning cost. These are medians over 8 circuits at $n \leq 14$; no asymptotic claim is made.

Bootstrap confidence intervals and fit-window sensitivity for the twelve α fits. The quoted $\alpha = 0.142 \pm 0.014$ is the mean and standard deviation of twelve descriptive fits. Two further uncertainties are quantified (`alpha_robustness.py`; 2000 bootstrap resamples, fixed seed). *Statistical:* within each (n, a) cell the per-t circuit sets are resampled with replacement, the mean trace $\eta(t)/\eta(0)$ is rebuilt and refit, and percentile intervals $[2.5\%, 97.5\%]$ are formed; combining independent cell resamples gives the grand-mean interval [0.130, 0.148]. *Systematic:* each cell is refit with the maximum included depth restricted to $t \leq 16, 24, 32$; the twelve-fit mean shifts monotonically through 0.122 ± 0.023, 0.132 ± 0.020, 0.138 ± 0.017, versus 0.142 ± 0.014 for the full window, because short windows weight the transient while the full window constrains the floor. The quoted rate is therefore window-dependent at the $\sim$15% level—consistent with a finite-size empirical rate, not an asymptotic constant. Four-circuit cells ($n = 18$) give correspondingly wide intervals; the $(18, 10)$ interval reaches the fit bound at zero.

n	a	f	circuits	α	bootstrap CI	$\alpha_{t\leq 16}$	$\alpha_{t\leq 24}$	$\alpha_{t\leq 32}$
8	5	0.62	8	0.124	$[0.062, 0.172]$	0.115	0.116	0.124
8	6	0.75	8	0.148	$[0.137, 0.162]$	0.137	0.141	0.148
10	6	0.60	8	0.158	$[0.117, 0.185]$	0.147	0.146	0.150
10	8	0.80	8	0.135	$[0.131, 0.140]$	0.111	0.123	0.127
12	8	0.67	8	0.146	$[0.129, 0.157]$	0.132	0.139	0.141
12	9	0.75	8	0.133	$[0.130, 0.137]$	0.103	0.120	0.129
14	8	0.57	8	0.172	$[0.167, 0.178]$	0.160	0.166	0.170
14	10	0.71	8	0.128	$[0.113, 0.137]$	0.094	0.110	0.119
16	10	0.62	8	0.149	$[0.143, 0.161]$	0.147	0.155	0.157
16	12	0.75	8	0.124	$[0.120, 0.129]$	0.081	0.091	0.108
18	10	0.56	4	0.148	$[0.000, 0.176]$	0.114	0.144	0.142
18	12	0.67	4	0.141	$[0.138, 0.145]$	0.128	0.132	0.138
mean				0.142	$[0.130, 0.148]$	0.122	0.132	0.138

Achievable per-setting Fisher information versus the needle proxy ($n = 10$, one-shot, eight circuits per cell). The proxy $\eta_{\text{needle}} = s^2_{\max}/F_Q^A$ omits the binomial variance factor $1 - \langle P_j\rangle^2$ of the exact per-setting Fisher information $F_j = s_j^2/(1 - \langle P_j\rangle^2)$. Regenerating the $n = 10$ circuits (identical seeds; `setting_distribution.py`, cross-checked record-by-record against the archived sweep) and evaluating F_j for all $4^a - 1$ settings gives $\eta_{\text{ach}} = \max_j F_j/F_Q^A$: the ratio $\eta_{\text{ach}}/\eta_{\text{needle}}$ stays within $[1.00, 1.03]$ at every sampled (a, t), because the best settings sit near $\langle P_j\rangle \approx 0$ where the variance factor is $\approx$1. The proxy therefore tracks the achievable per-setting fraction closely at these sizes. The same data quantify the dispersal: the cumulative share of total sensitivity $\sum_j s_j^2$ captured by the top-k settings collapses with t,

a	t	η_{needle}	η_{ach}	ratio	top-1	top-16	top-64	m_{eff}
6	0	0.75	0.75	1.000	0.321	0.819	0.874	44
6	6	0.354	0.354	1.000	0.167	0.916	1.000	13
6	16	0.0645	0.0645	1.001	0.028	0.190	0.416	212
6	32	0.0101	0.0101	1.001	0.004	0.045	0.129	1280
6	64	0.0085	0.0086	1.002	0.004	0.039	0.118	1364
8	0	1.00	1.00	1.000	0.031	0.500	1.000	32
8	6	0.599	0.599	1.000	0.019	0.175	0.431	194
8	16	0.121	0.121	1.000	0.004	0.034	0.086	3110
8	32	0.0147	0.0148	1.001	0.000	0.005	0.015	20501
8	64	0.0101	0.0101	1.001	0.000	0.004	0.013	21862

(top-k = share of $\sum_j s_j^2$ in the k largest settings, averaged over circuits.) At $t \geq 32$, even the union of the 64 best settings retains $\sim$1% of the total sensitivity at $f = 0.8$: the collapse is not an artifact of scoring a single setting. These are finite-size statements at $n = 10$; they complement, but do not prove, optimality claims over measurement families.

Appendix D Global operator-branching benchmark and finite-size comparison

Inputs. (I1) The SLD quantum Fisher information and its bound $F_Q^A \geq \mathrm{Tr}[(\partial_\theta \rho_A)^2]$, which follows term by term from $\lambda_j + \lambda_k \leq 2$ in the SLD sum [50]. (I2) The T-gate conjugation rule, verified by direct 2×2 computation with $T = \mathrm{diag}(1, e^{i\pi/4})$: $TZT^\dagger = Z$, $TXT^\dagger = (X+Y)/\sqrt{2}$, $TYT^\dagger = (Y-X)/\sqrt{2}$. (I3) The fact that Clifford conjugation maps the Pauli group to itself as a signed permutation and that the uniform Clifford ensemble maps any fixed non-identity Pauli to a uniformly random non-identity Pauli (2-design property of the Clifford group [51]). (I4) The elementary inequalities $\ell_\infty \leq \ell_2$ ($\max_j p_j \leq (\sum_j p_j^2)^{1/2}$) and Jensen for the concave square root ($\mathbb{E}\sqrt{\Pi} \leq \sqrt{\mathbb{E}\,\Pi}$).

Step (i), deterministic. With $p_j = s_j^2 / \sum_k s_k^2$ and $\Pi = \sum_j p_j^2$,

$$\eta_{\mathrm{needle}} = \frac{\sum_k s_k^2}{F_Q^A} \max_j p_j \;\leq\; C_{n,a}\,\sqrt{\Pi}, \qquad C_{n,a} = \frac{\sum_k s_k^2}{F_Q^A}, \tag{D10}$$

since $\max_j p_j \leq (\sum_j p_j^2)^{1/2}$ ($\ell_\infty \leq \ell_2$).

Step (ii), branching law. Consider the Pauli-weight distribution of the conjugated generator across the doped circuit. A T gate on qubit q leaves a Pauli component invariant if its letter on q is I or Z, and splits it into two equal-weight orthogonal components ($TXT^\dagger = (X+Y)/\sqrt{2}$) if the letter is X or Y. A split halves the component's contribution to Π: $\pi^2 \to 2(\pi/2)^2 = \pi^2/2$. Because a uniformly random Clifford conjugation precedes each T gate, each component's letter on q is uniform over $\{I, X, Y, Z\}$, so the split occurs with probability $\frac{1}{2}(1 + \frac{1}{4^n - 1})$ (uniformity over non-identity Paulis; exactly $1/2$ up to $O(4^{-n})$), giving per gate

$$\mathbb{E}[\Pi_{t+1}] = \left(\tfrac{1}{2} \cdot 1 + \tfrac{1}{2} \cdot \tfrac{1}{2}\right) \mathbb{E}[\Pi_t] = \tfrac{3}{4}\,\mathbb{E}[\Pi_t], \tag{D11}$$

hence $\mathbb{E}[\Pi_t] = (3/4)^t\,\Pi_0$. Clifford conjugation permutes Pauli strings, so distinct components cannot merge; only split branches can collide, a correction suppressed by the ratio of component number to 4^n (explicit bound open). *Scope of the branching result.* The accessible inequality $\eta_{\mathrm{needle}} \leq 2^a \sqrt{\Pi_A}$ is exact, but it does not determine the decay of Π_A. Step (ii) gives the collision-free global operator-purity benchmark $\sqrt{\mathbb{E}[\Pi_t^{\mathrm{op}}]} = (3/4)^{t/2} \sqrt{\Pi_0^{\mathrm{op}}}$, with $\alpha_{\mathrm{branch}} = \ln(4/3)/2 = 0.1438\ldots$. An analytic transfer from Π_t^{op} to Π_A is not supplied here, so this is not a bound on η_{needle} itself.

Finite-size comparison. Measured purity decay rates of the A-restricted sensitivity spectrum over $t \in [4, 16]$ are $0.285, 0.282$ ($n = 10$; $f = 0.6, 0.8$) and $0.285, 0.290$ ($n = 12$; $f = 0.67, 0.75$), close to $\ln(4/3) = 0.2877$. The fitted-rate ratio $\alpha_{\mathrm{needle}}/\alpha_{\mathrm{purity}} = 0.143/0.286 = 0.500$ and the mean of twelve needle fits, $\alpha = 0.142 \pm 0.014$, are empirical comparisons, not analytic consequences of the global branching law.

Branch combinatorics: what sets the observed rate. Under the twirl idealization the branch tree is exactly solvable: the expected number of leaves with k splits after t gates obeys $\mathbb{E}[N_k(t+1)] = \frac{1}{2}\mathbb{E}[N_k(t)] + \mathbb{E}[N_{k-1}(t)]$ (survive with probability $\frac{1}{2}$, or

split into two children of one higher split count), giving

$$\mathbb{E}[N_k(t)] = \binom{t}{k} \left(\tfrac{1}{2}\right)^{t-k}, \tag{D12}$$

which reproduces the observed total branch growth $\sum_k \mathbb{E}N_k = (3/2)^t$, weight conservation $\sum_k \mathbb{E}N_k\, 2^{-k} = 1$, and the purity law $\sum_k \mathbb{E}N_k\, 4^{-k} = (3/4)^t$ combinatorially. A first-moment extreme-value estimate places the smallest occupied split number at $k^* = \beta^* t$ with $H_2(\beta^*) + \beta^* = 1$, $\beta^* = 0.2271$ (first moment only; concentration would require a second-moment argument and is not claimed). Direct simulation of this occupancy process (2×10^5 realizations; `maxw_model.py`) shows that $\mathbb{E}[\max_b w_b]$ decays at a slowly drifting rate: 0.182 over $t \in [1, 32]$, 0.163 over $t \in [50, 100]$, approaching $\beta^* \ln 2 = 0.157$ only logarithmically—while $\mathbb{E}[\sqrt{\Pi}]$ decays at $0.145 \approx \ln(4/3)/2$, confirming that Eq. (4) measures the purity route, not the needle statistic. The two rates are reconciled by the partner multiplicity: at $f > 1/2$ each branch carries $M \approx 2^{2a-n}$ partner slots, and the maximum over M independent trees decays at rate 0.182, 0.174, 0.164, 0.152, 0.137 for $M = 1, 4, 16, 64, 256$—dragged monotonically downward, crossing $\ln(4/3)/2 = 0.144$ between $M = 64$ (0.152) and $M = 256$ (0.137). The M values realized in our twelve (n, a) sweeps span 4–256 (including $M = 256$ at $(n, a) = (16, 12)$ and $M = 4$ at $(18, 10)$), so the sweeps sit in the regime where the M-dragged rate straddles the benchmark; this heuristic independent-tree model does not account for partner correlations through the shared final Clifford layer. The observed $\alpha \approx 0.144$-coincidence is therefore an M-mediated finite-size effect, not evidence of asymptotic tightness; the asymptotic rate (and the M-corrected extreme-value theory) is the sharpest remaining open step.

Transfer to the A-restricted spectrum: the visibility filter (Proposition 5). The final Clifford layer acts *after* all splits, so each branch's Pauli identity—hence its visibility from A—is randomized independently of its weight (split history). The following proposition makes the filter argument quantitative at the second-moment level.

Proposition 5 (the visibility filter preserves the branching rate). Condition on the branch data $\{(w_b, P_b)\}_{b=1}^{B}$ produced by the doped layers: distinct non-identity Pauli strings P_b (the collision-free regime, verified up to $n = 1000$ below), weights $w_b \geq 0$ with $\sum_b w_b = 1$, purity $\Pi_{\mathrm{op}} = \sum_b w_b^2$, and maximal weight $w_{\max}$. Let C be the final layer, uniform in the n-qubit Clifford group and independent of the branch data, let $v_b = \mathbb{1}[\mathrm{supp}(CP_bC^\dagger) \subseteq A]$ and $q = (4^a - 1)/(4^n - 1)$, and define the filtered weight $W = \sum_b v_b w_b$, the filtered numerator $N = \sum_b v_b w_b^2$, and the filtered purity $\Pi_A^{\mathrm{br}} = N/W^2$ (on the event $W > 0$). Then: (i) $\mathbb{E}[v_b] = q$ for every b, hence $\mathbb{E}[W] = q$ and $\mathbb{E}[N] = q\,\Pi_{\mathrm{op}}$ exactly; (ii) for $b \neq b'$, $|\mathrm{Cov}(v_b, v_{b'})| \leq 3q^2/(4^a - 1)$; (iii) consequently, for any $\delta \in (0, \frac{1}{2}]$, if $\Pi_{\mathrm{op}} \leq \varepsilon^2 q$, $w_{\max}^2 \leq \varepsilon^2 q\,\Pi_{\mathrm{op}}$, and $3/(4^a - 1) \leq \varepsilon^2$, then

$$\Pr\left[\left|\Pi_A^{\mathrm{br}} \frac{q}{\Pi_{\mathrm{op}}} - 1\right| > 5\delta\right] \;\leq\; \frac{4\varepsilon^2}{\delta^2}. \tag{D13}$$

Thus, once the branch tree is deep enough that its purity and maximal weight are small on the scale of the visibility probability $q \approx 4^{-(n-a)}$, the filtered purity concentrates

at $\Pi_{\rm op}/q$: the filter changes the prefactor $1/q$ and *preserves the decay rate* $\ln(4/3)$ of $\mathbb{E}[\Pi_t^{\rm op}]$.

Proof. (i) The conjugation action of the Clifford group on non-identity Pauli strings modulo phase is the action of $\mathrm{Sp}(2n, \mathrm{GF}(2))$ on $\mathrm{GF}(2)^{2n} \setminus \{0\}$ [59]; it is transitive (shown equivalent to the 2-design property in Lemma 2 of Ref. [51]), so $CP_bC^\dagger$ is uniform over the $4^n - 1$ non-identity Paulis and $\mathbb{E}[v_b]$ equals the fraction q supported in A. (ii) By Witt's theorem for symplectic spaces over GF(2) [61], the action is also transitive on ordered pairs of distinct nonzero vectors with fixed symplectic product, i.e., on distinct Pauli pairs with fixed commutation relation. Counting pairs supported in A against all pairs (for fixed non-identity P there are $2^{2n-1}-2$ distinct non-identity commuting partners and 2^{2n-1} anticommuting ones),

$$\Pr[v_b = v_{b'} = 1] = \begin{cases} \dfrac{(4^a-1)(2^{2a-1}-2)}{(4^n-1)(2^{2n-1}-2)} & [P_b, P_{b'}] = 0, \\ \dfrac{(4^a-1)\,2^{2a-1}}{(4^n-1)\,2^{2n-1}} & \{P_b, P_{b'}\} = 0, \end{cases} \tag{D14}$$

giving $\mathrm{Cov}(v_b, v_{b'})/q^2 = -3(4^n - 4^a)/[(4^n - 4)(4^a - 1)]$ (commuting) or $+(4^n - 4^a)/[4^n(4^a - 1)]$ (anticommuting), both bounded in magnitude by $3/(4^a - 1)$; both formulas are verified by Monte-Carlo Clifford sampling at $n = 5$, $a = 2$ (`verify_proofs.py`, check 3). (iii) From (i)–(ii), $\mathrm{Var}(W) \le q\,\Pi_{\rm op} + 3q^2/(4^a - 1)$ and $\mathrm{Var}(N) \le q\,w_{\max}^2\,\Pi_{\rm op} + 3q^2\Pi_{\rm op}^2/(4^a - 1)$ (using $\mathrm{Var}(v_b) \le q$, $\sum_{b\neq b'} w_b w_{b'} \le 1$, $\sum_b w_b^4 \le w_{\max}^2 \Pi_{\rm op}$, $\sum_{b\neq b'} w_b^2 w_{b'}^2 \le \Pi_{\rm op}^2$). Under the stated conditions both relative variances are $\le 2\varepsilon^2$, so by Chebyshev each of W/q and $N/(q\Pi_{\rm op})$ deviates from 1 by more than δ with probability at most $2\varepsilon^2/\delta^2$; on the complement, $\Pi_A^{\rm br} q/\Pi_{\rm op} = [N/(q\Pi_{\rm op})]/[W/q]^2 \in [(1-\delta)/(1+\delta)^2, (1+\delta)/(1-\delta)^2] \subseteq [1-5\delta, 1+5\delta]$ for $\delta \le \frac{1}{2}$. □

What remains open in the transfer. Proposition 5 controls the *filtered branch* purity $\Pi_A^{\rm br}$. Identifying $\Pi_A^{\rm br}$ with the state-based A-restricted sensitivity purity Π_A additionally requires the per-branch state factors—the partner multiplicities of Proposition 4, treated by the M-multiplicity model above—to be weight-independent as well; that step, and the collision bound at fixed n, remain the open elements. The measured Π_A rates 0.282–0.290 agree with the rate preservation that Proposition 5 predicts.

Large-n verification of the branching law. The symplectic branch-tree simulator tracks the $\sim 1.5^t$ Pauli branches of $UGU^\dagger$ through the actual circuit (real Clifford layers, not the idealized twirl), with no statevector, and is validated against dense matrices at $n = 6$ (exact agreement per realization). Fitted operator-purity decay rates: 0.2903 at $n = 100$ (24 realizations, $t \le 32$) and 0.2794 at $n = 1000$ (8 realizations, $t \le 24$), versus $\ln(4/3) = 0.2877$; the prefactor $\mathbb{E}[\Pi_t]/(3/4)^t$ stays $O(1)$ throughout, and *no branch collisions* occur at any checkpoint at either size — the collision-neglect idealization of Step (ii) is exact in the large-n regime. The operator-level branching law is thus verified two orders of magnitude beyond the statevector simulations.

Commuting families. The classical Fisher information of the joint eigenbasis of the greedily maximized commuting family gives $\alpha_{\rm family} = 0.107(10), 0.072(11)$ (n=10; $f = 0.6, 0.8$), $0.090(9), 0.073(10)$ (n=12; $f = 0.67, 0.75$) — consistent with $\ln(4/3)/4 =$

0.072 at $f \geq 0.75$. A methodological caution recorded for reproducibility: the naive family score $\sum_{j\in\mathrm{fam}} s_j^2$ exceeds F_Q^A (by up to 32× in our data) because commuting Paulis measured in one basis are correlated estimators; the classical Fisher information of the joint eigenbasis is the correct, rigorously bounded quantity.

Fixed measurement families: the mean is flat (Proposition 6). The two-copy twirl that proves Lemma 2 yields more than the Gram mean: writing $U = CV$ with the final layer C uniform Clifford and V the remaining (arbitrary, independent) circuit, the response is $R = -i\,\mathrm{Tr}_{\bar{A}}(C\,M_V\,C^\dagger)$ with $M_V = V[G,\rho_0]V^\dagger$ fixed by conditioning, so the second moment of R over C alone is a 2-design twirl of a fixed traceless operator and depends only on $\mathrm{Tr}[M_V^2] = \mathrm{Tr}[[G,\rho_0]^2]$, which is t-independent. Expanding in A-Paulis, the twirl forces

$$\mathbb{E}[s_P\, s_Q] = c\,\delta_{PQ} \quad \text{for non-identity } P, Q \text{ on } A, \tag{D15}$$

with c independent of both P and t (isotropy; the off-diagonal vanishes because the identity and swap components of the twirl contribute $\mathrm{Tr}[P]\mathrm{Tr}[Q] = 0$ and $\mathrm{Tr}[PQ] = 0$ for $P \neq Q$). Consequently, for *every fixed* set F of non-identity A-Paulis chosen before the circuit is drawn,

$$\frac{\mathbb{E}\big[\sum_{P\in F} s_P^2\big]}{\mathbb{E}\big[\sum_j s_j^2\big]} = \frac{|F|}{4^a - 1} \qquad \text{at every } t. \tag{D16}$$

For a maximal commuting (stabilizer) family, $|F| = 2^a - 1$ and the mean captured share is $1/(2^a+1)$, independent of magic. Moreover, for the joint eigenbasis B_F of such a family, $p_x = 2^{-a}(1 + \sum_{P\in F}\chi_x(P)\langle P\rangle)$ gives $\partial_\theta p_x = 2^{-a}\sum_{P\in F}\chi_x(P)\,s_P$ and, by Parseval over the character group, $\sum_x(\partial_\theta p_x)^2 = 2^{-a}\sum_{P\in F} s_P^2$—so the ensemble mean of this squared-derivative norm is flat in t for every fixed stabilizer readout. The observed decay of α_{family} is therefore, like the needle collapse, a *concentration* phenomenon exposed by adaptive (greedy, per-realization) selection, not a decay of the mean; isotropy and the fixed-family flatness are verified numerically at $n = 6$, $a = 4$ (`verify_proofs.py`, check 4).

Scope: why no shot-count no-go over arbitrary POVMs is possible. Since F_Q^A/F_Q remains near unity at $f > 1/2$ for all sampled t (Fig. 2), the SLD-optimal POVM keeps the achievable classical Fisher information non-decaying; no lower bound on shot counts quantified over *all* POVMs can therefore exhibit an $e^{\alpha t}$ cost. The operational barrier documented in this paper lives in the hardware-native class of fixed Pauli or stabilizer settings—where Proposition 6 shows the mean is flat and the cost is concentration—and in the fact that constructing the SLD eigenbasis requires knowledge of the realized scrambler at a complexity comparable to U itself. Bounding the measurement *implementation* complexity (e.g., over bounded-depth readout circuits) rather than the shot count is the correctly posed open problem.

Appendix E End-to-end protocol under observation noise

We simulate the complete meta-metrology protocol with two forms of observation noise: finite sampling (S shots per Pauli setting, Gaussian approximation to the binomial) and readout bit-flip error ($\langle P\rangle \to (1-2q)^{w(P)}\langle P\rangle$ for a weight-w Pauli, error probability q per measured qubit). Protocol: (1) *calibrate*: for each dictionary element $G_\mu \in \{X_i, Y_i\}$, inject $\theta = \pm\delta$ ($\delta = 0.1$) and estimate all Pauli expectations on A, yielding noisy reference responses; (2) *identify*: nature encodes a uniformly random unknown G_*; its noisy response is decomposed over the references by least squares and the largest coefficient is declared; (3) *estimate*: the identified generator's calibrated best setting is used to invert $\theta_{\mathrm{sig}} = 0.05$ linearly. Eight circuit realizations, twenty noise trials each.

The two figures of merit at $n = 10$, $f = 0.5$, $q = 0$, as functions of the shots per setting S (a 2% readout error changes these values only marginally, e.g. $P_{\mathrm{id}} : 0.93 \to 0.89$ at $t = 6$, $S = 10^4$):

	identification success P_{id}			median rel. error of $\hat{\theta}$			
t	S=10^3	10^4	10^5	S=10^3	10^4	10^5	regime
0	0.25	0.31	0.35	0.55	0.32	0.31	identification principle-limited
6	0.19	**0.93**	0.99	0.53	**0.17**	0.06	**window: both succeed**
32	0.23	1.00	1.00	0.55	0.55	0.23	identify yes, estimate expensive

Read row-wise: at the Clifford point ($t = 0$) more shots barely help either task; in the window ($t = 6$) both columns improve together; deep in the magic phase ($t = 32$) the identification block reaches unity while the error block stays 3–4 times behind the window values at equal budget (0.55 vs 0.17 at 10^4 shots, 0.23 vs 0.06 at 10^5).

At the Clifford point the success probability climbs only slowly toward the ceiling of the *implemented* classifier—a blind arg-max over all m dictionary coefficients, whose asymptote is $\mathbb{E}[\mathrm{rank}\,\Gamma]/m + (1 - \mathbb{E}[\mathrm{rank}\,\Gamma]/m)/m = 0.377$ here (0.35 observed at 10^5 shots; invisible needles cannot be identified at any shot count). A Bayes-optimal classifier that excluded calibrated visible candidates upon a null response would reach the higher value $\mathbb{E}[\mathrm{rank}\,\Gamma]/m + \Pr(\mathrm{rank}\,\Gamma < m)/m \approx 0.394$; the quoted ceiling is classifier-specific, not information-theoretic. Inside the window both tasks succeed at 10^4 shots per setting, and deep in the magic phase identification is perfect while estimation inherits the $e^{-\alpha t}$ shot-cost inflation. At $f = 0.8$ the $t = 0$ identification failure disappears ($P_{\mathrm{id}} = 0.99$ at $S = 10^3$), confirming it is an $f \approx 1/2$ concentration phenomenon rather than a noise effect.

The identification threshold is flat in magic. Interpolating the shot count at which $P_{\mathrm{id}} = 0.5$:

t	2	6	12	16	32
S^*_{id}	6.9×10^3	2.7×10^3	2.6×10^3	2.6×10^3	2.3×10^3

For $t \geq 6$—the regime in which Γ is full rank, so all dictionary responses are well separated—the threshold is constant (mildly *decreasing*): the least-squares identification uses the full response vector, whose signal-to-noise is set by the basis-invariant $\sum_j s_j^2$ and is therefore immune to the extraction collapse; below $t \approx 6$ the failure mode is visibility (rank deficit), not shot noise.

Same-budget comparison with charged calibration. The flat threshold above presupposes that all $4^a - 1$ Pauli settings are calibrated; that cost is charged here. Both tasks are restricted to K settings drawn uniformly at random, and the total budget $B = SK(2m+2)+S$ counts every calibration shot (m generators $\times$ two θ-points $\times$ K settings), the signal acquisition, and the estimation shots ($n = 10$, $f = 0.5$, $m = 20$, eight circuit realizations, twenty noise trials; `fair_budget.py`). P_{id} | median relative error of $\hat{\theta}$ at $S = 4096$:

t	K=128	K=512	K=1023 (full)	B at K=1023
0	0.09 | 0.45	0.22 | 0.52	0.34 | 0.39	1.8×10^8
6	0.24 | 0.49	0.51 | 0.41	0.88 | 0.26	1.8×10^8
16	0.41 | 0.52	0.75 | 0.48	0.98 | 0.48	1.8×10^8
32	0.29 | 0.51	0.79 | 0.49	0.95 | 0.57	1.8×10^8

Resource frontier over the calibrated pool size. At the window operating point $t = 16$ the K sweep is refined (`resource_frontier.py`; eight circuit realizations $\times$ 60 noise trials per point, i.e., 480 trials versus 160 in the table above—in the steep region P_{id} fluctuates visibly with the drawn subsets, e.g., 0.41 [160 trials] versus 0.29(4) [480 trials] at $K = 128$; parenthetical uncertainties are between-circuit standard deviations). The budget decomposes as $B = B_{\mathrm{cal}} + B_{\mathrm{sig}} + B_{\mathrm{est}}$ with $B_{\mathrm{cal}} = 2mKS$: calibration carries 95% of B at every K, so the frontier cost is set almost entirely by the calibration stage. Classical post-processing (least-squares identification against $m = 20$ references) takes ≤ 0.21 ms per trial on a single CPU core, and each Pauli setting is implemented by one layer of single-qubit basis rotations (entangling readout depth zero): shots, not classical compute or readout depth, are the binding resource in this per-Pauli protocol. A joint-eigenbasis measurement of a stabilizer family would instead require an entangling Clifford readout circuit—in canonical form, $O(a^2/\log a)$ gates at parallel depth $O(\log a)$ for an a-qubit stabilizer circuit [59]—rather than a single layer of local basis rotations, and every such fixed family has magic-independent mean captured sensitivity (Proposition 6); whether structured or adaptive readout shifts this frontier is open.

K	B (total shots)	cal. share	P_{id}	median rel. err.	classical time (ms)
16	2.8×10^6	95%	0.05(4)	0.53	0.03
32	5.5×10^6	95%	0.09(5)	0.54	0.05
64	1.1×10^7	95%	0.14(4)	0.52	0.06
128	2.2×10^7	95%	0.29(4)	0.54	0.07
256	4.4×10^7	95%	0.49(5)	0.55	0.09
512	8.8×10^7	95%	0.75(5)	0.53	0.14
1023	1.8×10^8	95%	0.98(2)	0.49	0.20

Identification approaches reliability only as K approaches the full basis; with $K = 128$ of 1023 settings it fails at every depth. Under this per-Pauli measurement model, therefore, *neither* task enjoys a subexponential total-resource guarantee from our data: the identification/estimation contrast established elsewhere in this paper is a statement about how sensitivity distributes across settings at fixed calibrated pool, not a complexity separation under a common budget. Whether structured measurement families (e.g., simultaneous stabilizer-basis readout) restore a genuine separation is open. Estimation alone pays the $e^{-\alpha t}$ price. This is the operational, finite-shot form of the identification/estimation asymmetry.

Appendix F Free-fermion scramblers: proof of Proposition 7 and the degree chain

Inputs. (I1) The Majorana operators γ_μ defined in Sec. 4 are Hermitian and satisfy $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$. The parity operator $P = Z_0 \cdots Z_{n-1}$ anticommutes with every γ_μ. (I2) A product of unitaries generated by quadratic Majorana Hamiltonians acts as $U\gamma_\mu U^\dagger = \sum_\nu O_{\nu\mu}\gamma_\nu$ with $O \in \mathrm{SO}(2n)$. This is Eq. (31) of Ref. [54] and Theorem 3 of Ref. [55], whose index conventions differ from ours at most by a transposition of O. Nearest-neighbor matchgates are unitaries of this kind (Sec. 5 of Ref. [55]). (I3) We call a state full-rank Gaussian when $\rho = e^{-H}/\operatorname{Tr} e^{-H}$ with $H = \frac{i}{4}\sum_{\mu\nu} h_{\mu\nu}\gamma_\mu\gamma_\nu$ and h real and antisymmetric. Three properties follow from (I1) and (I2). First, P commutes with every product $\gamma_\mu\gamma_\nu$, so $P\rho P = \rho$. Second, $UHU^\dagger = \frac{i}{4}\sum_{\alpha\beta}(OhO^T)_{\alpha\beta}\gamma_\alpha\gamma_\beta$, so a Gaussian unitary maps a full-rank Gaussian state to a full-rank Gaussian state. Third, the vacuum is a limit of such states. Since $Z_j = -i\gamma_{2j+1}\gamma_{2j+2}$, the choice $h_{2j+1,2j+2} = -h_{2j+2,2j+1} = 2\beta$ with all other entries zero gives $\rho_\beta = \prod_j \frac{1}{2}(\mathbb{1} + \tanh\beta\, Z_j)$, which converges to ρ_0 as $\beta \to \infty$. Hence $U\rho_0 U^\dagger = \lim_{\beta\to\infty} U\rho_\beta U^\dagger$ is a limit of full-rank Gaussian states. The general theory of fermionic Gaussian states, which includes states of reduced rank, is given in Ref. [62].

Step 1: the dictionary is linear in the Majorana operators. $X_i = \gamma_{2i+1}S_i$ and $Y_i = \gamma_{2i+2}S_i$ with $S_i = Z_0 \cdots Z_{i-1}$, and S_i commutes with γ_{2i+1} and γ_{2i+2}. On the vacuum $S_i\rho_0 = \rho_0 S_i = \rho_0$, so $[X_i, \rho_0] = [\gamma_{2i+1}, \rho_0]$ and $[Y_i, \rho_0] = [\gamma_{2i+2}, \rho_0]$.

Step 2: reduction to $2n$ fixed operators. With $\rho = U\rho_0 U^\dagger$ and (I2), Eq. (1) gives

$$R(\gamma_\mu) = \sum_\nu O_{\nu\mu}\, T_\nu, \qquad T_\nu = -i \operatorname{Tr}_{\bar{A}}\big([\gamma_\nu, \rho]\big). \tag{F17}$$

Since O is invertible, the responses and the operators T_ν span the same space. Write $x_\nu = \operatorname{Tr}_{\bar{A}}(\rho\gamma_\nu)$ and $y_\nu = \operatorname{Tr}_{\bar{A}}(\gamma_\nu\rho)$, so that $T_\nu = -i(y_\nu - x_\nu)$.

Step 3: a Gaussian state acts linearly on Majorana operators. Let ρ be full rank. From (I1), $[\gamma_\mu\gamma_\nu, \gamma_\kappa] = 2\delta_{\nu\kappa}\gamma_\mu - 2\delta_{\mu\kappa}\gamma_\nu$, hence $[H, \gamma_\kappa] = i\sum_\mu h_{\mu\kappa}\gamma_\mu$ and $e^H\gamma_\kappa e^{-H} = \sum_\mu (e^{ih})_{\mu\kappa}\gamma_\mu$. Therefore

$$\gamma_\kappa\, \rho = \sum_\mu L_{\kappa\mu}\, \rho\, \gamma_\mu, \qquad L = (e^{ih})^T. \tag{F18}$$

Because ih is Hermitian, L is Hermitian and positive definite. Taking the partial trace of Eq. (F18) gives $y_\kappa = \sum_\mu L_{\kappa\mu} x_\mu$ for all κ.

Step 4: parity relation outside A. For $\nu > 2a$ the operator γ_ν has the form $P_A \otimes g_\nu$, where $P_A = Z_0 \cdots Z_{a-1}$ and g_ν acts on $\bar{A}$. Cyclicity of the partial trace for operators on $\bar{A}$ gives $y_\nu = P_A \operatorname{Tr}_{\bar{A}}(\rho\,(\mathbb{1} \otimes g_\nu)) = P_A\, x_\nu\, P_A$. Moreover $P_A x_\nu P_A = \operatorname{Tr}_{\bar{A}}(P\rho\gamma_\nu P) = -x_\nu$, because inserting $P_{\bar{A}}$ on both sides does not change a partial trace over $\bar{A}$, $P\rho P = \rho$, and $P\gamma_\nu P = -\gamma_\nu$. Hence

$$y_\nu = -x_\nu \qquad (\nu > 2a). \tag{F19}$$

Step 5: all responses lie in a $2a$-dimensional space. Split the index set into $A = \{1, \ldots, 2a\}$ and its complement e. Combining Step 3 with Eq. (F19) for $\nu \in e$ gives $(\mathbb{1}+L_{ee})\, x_e = -L_{eA}\, x_A$. The principal block L_{ee} of a positive-definite matrix is positive definite, so $\mathbb{1} + L_{ee}$ is invertible and

$$x_e = -(\mathbb{1} + L_{ee})^{-1} L_{eA}\, x_A. \tag{F20}$$

Thus every x_ν, and by Step 3 every y_ν, lies in $V = \operatorname{span}\{x_\mu : \mu \in A\} = \operatorname{span}\{\rho_A \gamma_\mu : \mu \le 2a\}$, whose dimension is at most $2a$. So do all T_ν, which proves $\operatorname{rank}\Gamma \le 2a$ for full-rank Gaussian states.

Step 6: pure states. The operators T_ν depend linearly and continuously on ρ, and the set of Gram matrices of rank at most $2a$ is closed. The bound therefore extends to limits of full-rank Gaussian states, in particular to $U\rho_0 U^\dagger$.

Step 7: equality. For $\nu \le 2a$, $T_\nu = -i[\gamma_\nu, \rho_A]$. If these $2a$ operators are linearly independent, the span has dimension exactly $2a$. Independence fails only if some combination $c \cdot \gamma$ of Majorana operators on A commutes with ρ_A, which does not occur for generic Gaussian states. □

Verification. Steps 1, 3, 4, and 5 are checked on dense matrices for a random full-rank Gaussian state at $n = 5$, and the rank is checked on pure Gaussian states prepared by matchgate circuits at $n = 6$ (`free_fermion_theory.py`). In the production data the rank equals $2a$ in all 136 free-fermion realizations at $n = 8$–14.

Degree chain. (i) A Gaussian unitary maps a Majorana monomial of degree d to a combination of monomials of the same degree. To see this, write $\gamma'_\mu = U\gamma_\mu U^\dagger$. These operators satisfy $\{\gamma'_\mu, \gamma'_\nu\} = 2\delta_{\mu\nu}$, so a product of d distinct γ'_μ equals its antisymmetrization over the d factors. Expanding each factor with (I2) gives antisymmetrized products of the γ_ν. Such a product vanishes when two indices coincide and is a monomial of degree d otherwise. In the Grassmann representation the same statement is the orthogonal change of variables in Eq. (16) of Ref. [62]. The total weight p_d of an operator in each degree sector is therefore conserved. (ii) The Haar measure on $\mathrm{SO}(2n)$ is invariant under signed permutations of the Majorana operators with unit determinant, which act transitively on the monomials of a given degree; after a Haar-random Gaussian unitary the expected weight is therefore uniform within each sector. (iii) A CZ gate on modes $j, j+1$ maps each of the four local Majorana operators to itself times the parity Z of the other mode, for example $\gamma_{2j+1} \to \gamma_{2j+1} Z_{j+1}$, and so permutes monomials. Let $\Delta(\mathcal{L})$ be the resulting change of degree for a local pattern $\mathcal{L} \subseteq \{1, 2, 3, 4\}$ of occupied local Majorana operators. A uniformly random monomial

of degree d has pattern $\mathcal{L}$ with probability $\binom{2n-4}{d-|\mathcal{L}|}/\binom{2n}{d}$, so

$$P_{d'd} = \sum_{\mathcal{L}:\, d+\Delta(\mathcal{L})=d'} \binom{2n-4}{d-|\mathcal{L}|} \Big/ \binom{2n}{d}, \tag{F21}$$

and conditioning on the history gives $\mathbb{E}[p(k+1)] = P\,\mathbb{E}[p(k)]$ exactly. (iv) A degree-one monomial is one of the four local operators with probability $4/2n$, and each of these is mapped to degree three, so $P_{31} = 2/n$ and $P_{11} = 1 - 2/n$; the return flow is $P_{13} = 4/\binom{2n}{3}$. This gives Eq. (6). The matrix P is checked against a direct simulation with Haar-random Gaussian unitaries at $n = 5$ (300 samples, agreement within 0.03 in every degree sector). The brickwork blocks used in the simulations only approximate the Haar average: at $n = 8$ the degree-one weight after $k = 1, 2, 3$ CZ gates is 0.69, 0.53, 0.37 against 0.75, 0.56, 0.43 from the chain.

References

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photon. **5**, 222 (2011), https://doi.org/10.1038/nphoton.2011.35.

[2] P. Hayden and J. Preskill, Black holes as mirrors: quantum information in random subsystems, JHEP **09**, 120 (2007), https://doi.org/10.1088/1126-6708/2007/09/120.

[3] H. Sharma, S. Choudhury, and J. N. Bandyopadhyay, Quantum metrology with partially accessible chaotic sensors, arXiv:2602.12914

[4] Y. Zhang, J. Zuniga Castro, and R. J. Lewis-Swan, Harnessing quantum chaos in spin-boson models for all-purpose quantum-enhanced sensing, arXiv:2410.03965

[5] P. Wysocki, J. Chwedeńczuk, and M. Płodzień, Volume-law protection of metrological advantage, arXiv:2602.09086 (2026).

[6] T. Hernández-Yanes, P. Sierant, J. Zakrzewski, and M. Płodzień, Non-stabilizerness in quantum-enhanced metrological protocols, Phys. Rev. A **113**, 012416 (2026), https://doi.org/10.1103/tmf9-fyc2.

[7] D. Bondarenko and P. Feldmann, Quantum autoencoders to denoise quantum data, Phys. Rev. Lett. **124**, 130502 (2020), https://doi.org/10.1103/PhysRevLett.124.130502.

[8] L. Leone, S. F. E. Oliviero, S. Piemontese, S. True, and A. Hamma, Retrieving information from a black hole using quantum machine learning, Phys. Rev. A **106**, 062434 (2022), https://doi.org/10.1103/PhysRevA.106.062434.

[9] L. Leone, S. F. E. Oliviero, and A. Hamma, Learning efficient decoders for quasi-chaotic quantum scramblers, Phys. Rev. A **109**, 022429 (2024), https://doi.org/

10.1103/PhysRevA.109.022429.

[10] M. Vetrano, G. Lo Monaco, L. Innocenti, S. Lorenzo, and G. M. Palma, State estimation with quantum extreme learning machines beyond the scrambling time, npj Quantum Inf. **11**, 20 (2025), https://doi.org/10.1038/s41534-024-00927-5.

[11] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, Barren plateaus preclude learning scramblers, Phys. Rev. Lett. **126**, 190501 (2021), https://doi.org/10.1103/PhysRevLett.126.190501.

[12] R. J. Garcia, K. Bu, and A. Jaffe, Quantifying scrambling in quantum neural networks, JHEP **03**, 027 (2022), https://doi.org/10.1007/JHEP03(2022)027.

[13] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018), https://doi.org/10.1038/s41467-018-07090-4.

[14] G. Aguilar, S. Jerbi, J. Eisert, and L. Bittel, Efficient learning of Clifford disentanglers and typical t-doped unitaries with exponentially more T gates, arXiv:2609.27565 (2026).

[15] T. Haug, Efficient learning of Clifford-scrambled product states, arXiv:2609.27128 (2026).

[16] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, One-shot decoupling, Commun. Math. Phys. **328**, 251 (2014), https://doi.org/10.1007/s00220-014-1990-4.

[17] W. Brown and O. Fawzi, Decoupling with random quantum circuits, Commun. Math. Phys. **340**, 867 (2015), https://doi.org/10.1007/s00220-015-2470-1.

[18] Y. Nakata, E. Wakakuwa, and M. Koashi, Black holes as clouded mirrors: the Hayden-Preskill protocol with symmetry, Quantum **7**, 928 (2023), https://doi.org/10.22331/q-2023-02-21-928.

[19] H. Tajima and K. Saito, Universal limitation of quantum information recovery: symmetry versus coherence, arXiv:2103.01876 (2021).

[20] B. Yoshida, Soft mode and interior operator in Hayden-Preskill thought experiment, Phys. Rev. D **100**, 086001 (2019), https://doi.org/10.1103/PhysRevD.100.086001.

[21] J. Liu, Scrambling and decoding the charged quantum information, Phys. Rev. Research **2**, 043164 (2020), https://doi.org/10.1103/PhysRevResearch.2.043164.

[22] N. Bao and Y. Kikuchi, Hayden-Preskill decoding from noisy Hawking radiation, JHEP **02**, 017 (2021), https://doi.org/10.1007/JHEP02(2021)017.

[23] B. Yoshida and N. Y. Yao, Disentangling Scrambling and Decoherence via Quantum Teleportation, Phys. Rev. X **9**, 011006 (2019), https://doi.org/10.1103/PhysRevX.9.011006.

[24] Y. Nakata and M. Tezuka, Hayden-Preskill Recovery in Hamiltonian Systems, Phys. Rev. Research **6**, L022021 (2024), https://doi.org/10.1103/PhysRevResearch.6.L022021.

[25] L. J. Fiderer and D. Braun, Quantum metrology with quantum-chaotic sensors, Nat. Commun. **9**, 1351 (2018), https://doi.org/10.1038/s41467-018-03623-z.

[26] B. Yoshida and A. Kitaev, Efficient decoding for the Hayden-Preskill protocol, arXiv:1710.03363

[27] A. R. Brown, H. Gharibyan, S. Leichenauer, H. W. Lin, S. Nezami, G. Salton, L. Susskind, B. Swingle, and M. Walter, Quantum Gravity in the Lab: Teleportation by Size and Traversable Wormholes, PRX Quantum **4**, 010320 (2023), https://doi.org/10.1103/PRXQuantum.4.010320.

[28] T. Schuster, B. Kobrin, P. Gao, I. Cong, E. T. Khabiboulline, N. M. Linke, M. D. Lukin, C. Monroe, B. Yoshida, and N. Y. Yao, Many-body quantum teleportation via operator spreading in the traversable wormhole protocol, Phys. Rev. X **12**, 031013 (2022), https://doi.org/10.1103/PhysRevX.12.031013.

[29] Y. Nakayama, A. Miyata, and T. Ugajin, The Petz (lite) recovery map for scrambling channel, arXiv:2310.18991 (2023).

[30] K. A. Landsman, C. Figgatt, T. Schuster, N. M. Linke, B. Yoshida, N. Y. Yao, and C. Monroe, Verified quantum information scrambling, Nature **567**, 61 (2019), https://doi.org/10.1038/s41586-019-0952-6.

[31] D. Harlow and P. Hayden, Quantum Computation vs. Firewalls, JHEP **06**, 085 (2013), https://doi.org/10.1007/JHEP06(2013)085.

[32] S. Aaronson, The Complexity of Quantum States and Transformations: From Quantum Money to Black Holes, arXiv:1607.05256 (2016).

[33] I. H. Kim, E. Tang, and J. Preskill, The ghost in the radiation: Robust encodings of the black hole interior, JHEP **06**, 031 (2020), https://doi.org/10.1007/JHEP06(2020)031.

[34] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, The Python's Lunch: geometric obstructions to decoding Hawking radiation, arXiv:1912.00228 (2019).

[35] Z. Ji, Y. Liu, and F. Song, Pseudorandom States, Non-Cloning Theorems and Quantum Money, in *Advances in Cryptology – CRYPTO 2018*, Lecture Notes in Computer Science Vol. 10993 (Springer, Cham, 2018), p. 126, https://doi.org/10

.1007/978-3-319-96878-0_5.

[36] F. Ma and H. Huang, How to Construct Random Unitaries, arXiv:2410.10116 (2024).

[37] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021), https://doi.org/10.1103/PRXQuantum.2.040316.

[38] R. A. Low, Learning and Testing Algorithms for the Clifford Group, Phys. Rev. A **80**, 052314 (2009), https://doi.org/10.1103/PhysRevA.80.052314.

[39] S. F. E. Oliviero, L. Leone, S. Lloyd, and A. Hamma, Unscrambling Quantum Information with Clifford decoders, Phys. Rev. Lett. **132**, 080402 (2024), https://doi.org/10.1103/PhysRevLett.132.080402.

[40] L. Leone, S. F. E. Oliviero, and A. Hamma, Learning t-doped stabilizer states, Quantum **8**, 1361 (2024), https://doi.org/10.22331/q-2024-05-27-1361.

[41] S. Grewal, V. Iyer, W. Kretschmer, and D. Liang, Efficient Learning of Quantum States Prepared With Few Non-Clifford Gates, Quantum **9**, 1907 (2025), https://doi.org/10.22331/q-2025-11-06-1907.

[42] H. Huang, Y. Liu, M. Broughton, I. Kim, A. Anshu, Z. Landau, and J. R. McClean, Learning shallow quantum circuits, in *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (ACM, New York, 2024), https://doi.org/10.1145/3618260.3649722.

[43] A. A. Mele and Y. Herasymenko, Efficient learning of quantum states prepared with few fermionic non-Gaussian gates, PRX Quantum **6**, 010319 (2025), https://doi.org/10.1103/PRXQuantum.6.010319.

[44] A. Gu, S. F. E. Oliviero, and L. Leone, Magic-induced computational separation in entanglement theory, PRX Quantum **6**, 020324 (2025), https://doi.org/10.1103/PRXQuantum.6.020324.

[45] A. Gu, L. Leone, S. Ghosh, J. Eisert, S. Yelin, and Y. Quek, Pseudomagic Quantum States, Phys. Rev. Lett. **132**, 210602 (2024), https://doi.org/10.1103/PhysRevLett.132.210602.

[46] H. Tang, S. Vardhan, and J. Wang, Estimating time in quantum chaotic systems and black holes, SciPost Phys. **19**, 095 (2025), https://doi.org/10.21468/SciPostPhys.19.4.095.

[47] M. Karjula, T. Ojanen, T. Ala-Nissila, and M. N. Ivaki, The ebbs and flows of quantum learning and sensing, arXiv:2608.20155 (2026).

[48] E. Bairey, I. Arad, and N. H. Lindner, Learning a local Hamiltonian from local measurements, Phys. Rev. Lett. **122**, 020504 (2019), https://doi.org/10.1103/PhysRevLett.122.020504.

[49] L. Leone, S. F. E. Oliviero, and A. Hamma, Stabilizer Rényi entropy, Phys. Rev. Lett. **128**, 050402 (2022), https://doi.org/10.1103/PhysRevLett.128.050402.

[50] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Phys. Rev. Lett. **72**, 3439 (1994), https://doi.org/10.1103/PhysRevLett.72.3439.

[51] H. Zhu, Multiqubit Clifford groups are unitary 3-designs, Phys. Rev. A **96**, 062336 (2017), https://doi.org/10.1103/PhysRevA.96.062336.

[52] X. Turkeshi, E. Tirrito, and P. Sierant, Magic spreading in random quantum circuits, Nat. Commun. **16**, 2575 (2025), https://doi.org/10.1038/s41467-025-57704-x.

[53] D. Szombathy, A. Valli, C. P. Moca, J. Asbóth, L. Farkas, T. Rakovszky, and G. Zaránd, Spectral properties versus magic generation in T-doped random Clifford circuits, Phys. Rev. Research **7**, 043080 (2025), https://doi.org/10.1103/557f-6tpb.

[54] B. M. Terhal and D. P. DiVincenzo, Classical simulation of noninteracting-fermion quantum circuits, Phys. Rev. A **65**, 032325 (2002), https://doi.org/10.1103/PhysRevA.65.032325.

[55] R. Jozsa and A. Miyake, Matchgates and classical simulation of quantum circuits, Proc. R. Soc. A **464**, 3089 (2008), https://doi.org/10.1098/rspa.2008.0189.

[56] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth, Efficient unitary designs with a system-size independent number of non-Clifford gates, Commun. Math. Phys. **397**, 995 (2023), https://doi.org/10.1007/s00220-022-04507-6.

[57] H. Wakaura and T. Tanimae, qnoise-pack: code, data, and reproduction driver for "Identifiability and Pauli-setting concentration of metrological signals across scrambler families," Version 1.3.0, [Software and dataset], Zenodo, 2026, https://doi.org/10.5281/zenodo.24156665 (concept DOI https://doi.org/10.5281/zenodo.21532188 resolves to the latest version).

[58] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys. **264**, 773 (2006), https://doi.org/10.1007/s00220-006-1554-3.

[59] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004), https://doi.org/10.1103/PhysRevA.70.052328.

[60] J. Dehaene and B. De Moor, The Clifford group, stabilizer states, and linear and quadratic operations over GF(2), Phys. Rev. A **68**, 042318 (2003), https://doi.org/10.1103/PhysRevA.68.042318.

[61] E. Artin, *Geometric Algebra* (Interscience, New York, 1957), Chap. III (Witt's theorem).

[62] S. Bravyi, Lagrangian representation for fermionic linear optics, Quantum Inf. Comput. **5**, 216 (2005), arXiv:quant-ph/0404180.