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.

Efficient and identifiable noise models for symmetry-preserving quantum circuits

Yi Pan
School of Computing, University of Georgia https://orcid.org/0009-0001-2938-7646
Meng Hsiu Tsai
Department of Computer Science and Engineering, University of Tennessee at Chattanooga
https://orcid.org/0009-0001-7752-0253
Hanqi Jiang
School of Computing, University of Georgia https://orcid.org/0009-0007-2778-959X
Junhao Chen
School of Computing, University of Georgia https://orcid.org/0009-0008-5888-129X
Weihang You
School of Computing, University of Georgia https://orcid.org/0000-0002-1563-6293
Yifan Zhou
School of Computing, University of Georgia https://orcid.org/0009-0007-8985-4537
Wei Zhang
School of Computer and Cyber Sciences, Augusta University https://orcid.org/0000-0001-6508-8144
Isaac Lyngaas
National Center for Computational Sciences, Oak Ridge National Laboratory https://orcid.org/0000-0002-1682-4309
Yingfeng Wang
yingfeng-wang@utc.edu

Department of Computer Science and Engineering, University of Tennessee at Chattanooga
https://orcid.org/0000-0002-3715-5124
Tianming Liu
tliu@uga.edu

School of Computing, University of Georgia https://orcid.org/0000-0002-8132-9048

Method Article

Keywords:

Posted Date: October 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9740811/v2

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

Additional Declarations: The authors declare no competing interests.

Efficient and identifiable noise models for symmetry-preserving quantum circuits

Yi Pan[1†], Meng Hsiu Tsai[2†], Hanqi Jiang[1†], Junhao Chen[1], Weihang You[1], Yifan Zhou[1], Wei Zhang[3], Isaac Lyngaas[4], Yingfeng Wang[2*], Tianming Liu[1*]

[1*]School of Computing, University of Georgia, Athens, 30602, GA, USA.
[2*]Department of Computer Science and Engineering, University of Tennessee at Chattanooga, Chattanooga, 37403, TN, USA.
[3]School of Computer and Cyber Sciences, Augusta University, Augusta, 30912, GA, USA.
[4]National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, 37830, TN, USA.

*Corresponding author(s). E-mail(s): yingfeng-wang@utc.edu; tliu@uga.edu;
†These authors contributed equally to this work.

Abstract

Noise models support quantum-circuit simulation and need recalibration as hardware and workloads change. For circuits whose ideal outputs occupy one Hamming-weight sector, independent bit-flip (IBF) models tie leakage from the sector to the flip rate. We introduce SA-IBF, a two-parameter closed-form model whose sector-preserving weight mixes independent bit flips with flips conditioned on the sector to control this leakage separately. For nonzero weights, no mixture of product bit-flip channels reproduces its kernel. Given ideal distributions, we derive local identifiability conditions and a rank diagnostic for each fit. Across two data sets on four IBM Heron processors, SA-IBF predicts test data with lower Hellinger distance than IBF on every execution. On molecular quantum-autoencoder circuits, it matches a 21-parameter noise model fitted by Bayesian optimization to within 0.01 in Hellinger distance at three of four qubit numbers, at over 200-fold lower computational cost. This enables low-cost noise modeling of small symmetry-preserving circuits.

1 Introduction

Noise models describe how imperfect quantum operations alter the measurement statistics of a circuit. They support the characterization of hardware errors [1–3], the evaluation of variational and quantum machine-learning algorithms under realistic conditions [4–8], and the development of model-based error mitigation [9–12]. At the level of circuit outputs, a noise model maps the ideal output distribution of a circuit to the distribution observed on a device. Data-driven methods learn related corrections directly from such outputs [13–18].

Many circuits in these applications conserve a quantity. A Hamming-weight sector is the set of bitstrings with a fixed number of ones. Particle-number-conserving chemistry circuits and magnetization-conserving spin circuits that start in one sector keep their ideal output in that sector [19]. On hardware, noise redistributes probability within this sector and leaks probability to other sectors. Only the leakage violates the ideal constraint, and it provides the signal that symmetry verification uses to detect errors [19]. A noise model for such circuits therefore has to describe two features of the measured distribution: how probability spreads over bitstrings, and how much of it leaks out of the ideal sector. A noise model is updated as calibration conditions and workloads change, so it should also be compact. A compact model also has to be identifiable, and two parameters that both affect the population of the sector can be nearly indistinguishable in finite-shot data. Two questions follow: can a compact noise model describe the leakage separately from the spreading, and under what conditions can the two be distinguished from output measurements?

Existing approaches balance physical detail against computational cost. Noise models built from device calibration data and learned Pauli-channel models describe several error mechanisms [3, 20–23]. Fitting them can require a repeated numerical search over many parameters, 20 or 21 in the noise models compared here. Simple noise kernels, such as Pauli-twirled depolarization, independent bit flips, or depth-dependent white noise [24–27], reduce this cost by strongly restricting the output distribution. In the independent bit-flip (IBF) model, for example, one flip rate controls both the spreading and the leakage out of the sector, and this coupling limits a compact fit. Because the sector is known, the coupling can be relaxed without a detailed gate-level model. Conditioning the same bit flips on the known sector adds a sector-preserving weight and keeps the parameter space small enough for direct identifiability analysis.

Here, we introduce SA-IBF, the Symmetry-Aware Independent Bit-Flip model, for circuits whose ideal output is supported on a single Hamming-weight sector (Fig. 1). SA-IBF mixes independent bit flips with the same flips conditioned on remaining in the input's sector. Both components share the flip rate q, and the sector-preserving weight α sets their proportions. At a fixed flip rate, α changes only the population of the sector and leaves relative probabilities within the sector, and separately outside it, unchanged (Eq. (5)). The flip rate q sets the spreading and the IBF leakage, and α is the fraction of that leakage the model suppresses. Given the ideal distribution computed from the circuit, both parameters are fitted to hardware histograms by minimum Hellinger distance and evaluated on test data. An analytic Jacobian at the fitted parameters then assesses local identifiability and conditioning. The model applies unchanged to ideal outputs concentrated in one sector, such as the molecular circuits studied here.

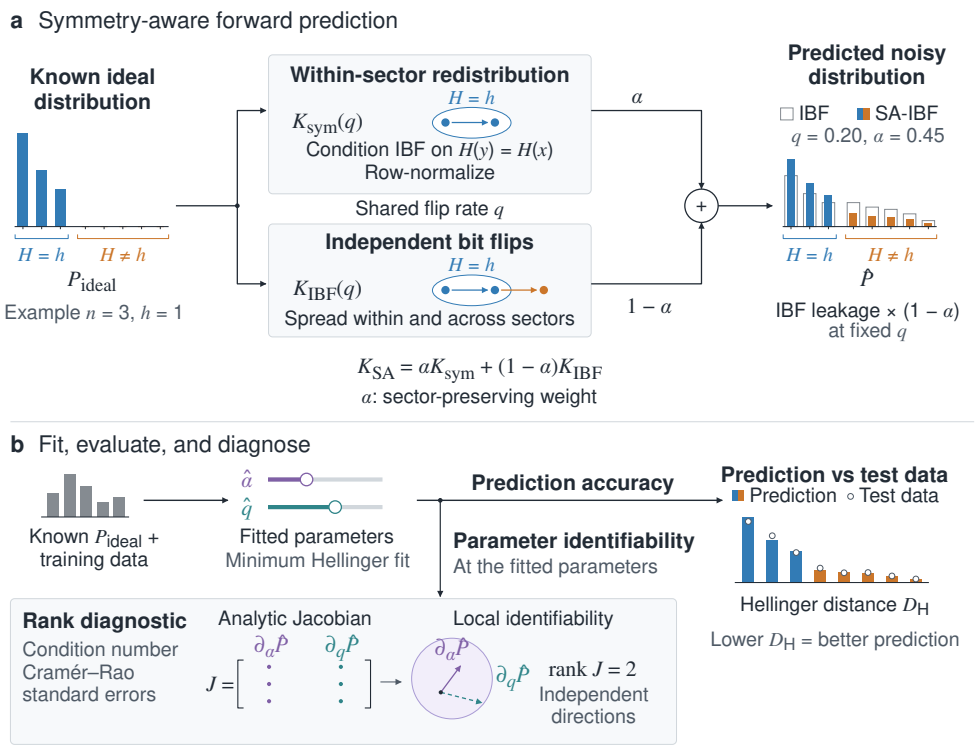


Fig. 1 SA-IBF model and fitting pipeline. **a** Forward prediction for $n = 3$ qubits with the ideal distribution in the sector of weight $h = 1$. Both components act on the known ideal distribution P_{ideal} with a shared flip rate q. The sector component K_{sym} applies independent bit flips conditioned on keeping the Hamming weight of each input, and the bit-flip component K_{IBF} spreads probability within and across sectors. Their mixture with weight α (Eq. (3)) gives the predicted distribution $\hat{P}$. Blue and orange bars mark probability inside and outside the ideal sector, and open bars show the IBF prediction at the same q. **b** Fitting, evaluation, and diagnosis. Both parameters are fitted by minimizing the Hellinger distance to the training data, the prediction (bars) is evaluated by its Hellinger distance D_{H} to the test data (open circles), and the rank diagnostic checks the analytic Jacobian at the fitted parameters.

We first establish what the model can represent and what the data can determine. For nonzero sector-preserving weights, no mixture of product bit-flip channels reproduces the SA-IBF kernel (Proposition 1). With the ideal distribution known, we give conditions for local identifiability of (α, q) [28–30], including a rank condition that can be checked for each fit (Theorems 1 to 3). We also bound the information on α at weak noise. For up to twelve qubits, the Fisher information on α when q is estimated jointly is at most proportional to Nq^2 for N shots (Theorem 1(b)). We then test the model on IBM Heron hardware. On the 230 circuits of the main data set, every fit passes the rank diagnostic (Fig. 2b,c). SA-IBF has lower Hellinger distance to test data than IBF on all of these circuits and on all 400 executions of a four-processor data set (Table 1 and Figs. 3 and 4a). On molecular quantum-autoencoder (MolQAE) circuits [31], we compare it with Ji-BO [20], a noise model built from device calibration data whose

parameters are fitted by Bayesian optimization. At three of four qubit numbers, the 95% confidence interval of the difference in Hellinger distance between SA-IBF and the 21-parameter sector-aware Ji-BO lies within ±0.01, and SA-IBF is fitted over 200 times faster (Table 2). On magnetization-conserving circuits with four qubits, SA-IBF has a lower mean Hellinger distance than every Ji-BO variant tested.

2 Results

2.1 The SA-IBF noise model

We consider circuits whose noiseless output distribution P_{ideal} is computable from the circuit specification and whose noisy output is observed as a finite-shot hardware histogram P_{obs}. A noise model maps P_{ideal} to a predicted distribution through a stochastic kernel fitted to training data and evaluated on test data.

Fix $n \geq 2$ qubits, written n_q in the hardware sections. Write $\{0,1\}^n$ for bitstrings, $H(x) := \sum_i x_i$ for the Hamming weight of x, and $d(x,y) := \|x \oplus y\|_1$ for Hamming distance. Let Δ_{2^n} be the probability simplex on $\{0,1\}^n$, with strictly positive interior $\text{Int}(\Delta_{2^n})$. The *sector* $[x]_S$ of x is the set of bitstrings with the same Hamming weight as x. The fitted parameters are a pair (α, q) with $\alpha \in [0,1]$ and $q \in (0, 1/2)$, defined below.

IBF noise describes incoherent bit-flip errors, and its kernel does not depend on the sector of the input. Particle-number- and magnetization-conserving circuits keep P_{ideal} in the sector of their initial state [19]. Neither IBF nor the other baselines compared below spread probability while keeping it in the input's sector. We therefore mix IBF with a sector-preserving component, and we take K_{sym} to be IBF conditioned on the sector of each input.

Kernel definition. For $\alpha \in [0,1]$ and $q \in (0, 1/2)$, define

$$K_{\text{IBF}}(y \mid x; q) := q^{d(x,y)}(1-q)^{n-d(x,y)}, \tag{1}$$

$$K_{\text{sym}}(y \mid x; q) := \mathbf{1}[H(y) = H(x)]\, K_{\text{IBF}}(y \mid x; q)/Z_x(q), \tag{2}$$

$$K_{\text{SA}}(y \mid x; \alpha, q) := \alpha\, K_{\text{sym}}(y \mid x; q) + (1-\alpha)\, K_{\text{IBF}}(y \mid x; q), \tag{3}$$

where $Z_x(q) := \sum_{y' \in [x]_S} K_{\text{IBF}}(y' \mid x; q)$ is the sector normalizer. Throughout, $P_{\text{ideal}} \in \Delta_{2^n}$ is the noiseless output distribution computed from the circuit specification and is a known input to the fit. Only (α, q) are estimated. The SA-IBF predicted distribution, sketched in Fig. 1a, is

$$\begin{aligned} \hat{P}(\cdot; \alpha, q, P_{\text{ideal}}) : y \mapsto \sum_x P_{\text{ideal}}(x)\, K_{\text{SA}}(y \mid x; \alpha, q), \\ \hat{p}_y(\alpha, q, P_{\text{ideal}}) := \hat{P}(y; \alpha, q, P_{\text{ideal}}). \end{aligned} \tag{4}$$

Sector population. Let P_{ideal} be supported on the sector of weight h, and write $B_q := K_{\text{IBF}}^{\top} P_{\text{ideal}}$ for the IBF prediction. The unitary coupled-cluster (UCCSD) state-preparation circuits and the magnetization-conserving spin circuits (Magnetization) used below satisfy this exactly. The MolQAE circuits of the main and four-processor data sets place at least 94% of their ideal mass on the all-zero bitstring, which

forms the weight-0 sector, and the remaining mass is treated below. The normalizer $Z_x(q)$ depends on x only through $H(x)$, so on the support it takes the single value $Z_h(q) = \sum_{j=0}^{\min(h,n-h)} \binom{h}{j}\binom{n-h}{j} q^{2j}(1-q)^{n-2j}$, the probability that bit-flip noise leaves the Hamming weight unchanged. Eq. (4) then reduces to

$$\hat{P}(y) = \begin{cases} \big(1-\alpha+\alpha/Z_h(q)\big)\, B_q(y), & H(y)=h, \\ (1-\alpha)\, B_q(y), & H(y)\neq h, \end{cases} \qquad \hat{P}(H{=}h) = \alpha + (1-\alpha)\, Z_h(q). \tag{5}$$

SA-IBF therefore rescales the IBF output uniformly inside the sector and uniformly outside it, so it changes only the population of the sector. Fig. 2a shows this population as a function of q for several values of α. The flip rate q sets the spreading and the IBF leakage $1 - Z_h(q)$, and α is the fraction of that leakage the model suppresses, since SA-IBF leaks $(1-\alpha)\big[1-Z_h(q)\big]$. Suppose a fraction η of the ideal mass lies outside the sector, as for these MolQAE circuits with $\eta \le 0.06$. The predicted population of the sector is then $(1-\eta)\big[\alpha + (1-\alpha)Z_h(q)\big] + \eta s$ for some $s \in [0,1]$, so it differs from $\alpha + (1-\alpha)Z_h(q)$ by at most η.

Every stochastic sector-preserving component in place of K_{sym} gives the same sector population $\hat{P}(H{=}h) = \alpha + (1-\alpha)Z_h(q)$. Among such components, K_{sym} is the unique one whose output equals the IBF output conditioned on the input sector for every single-sector P_{ideal}, as follows by taking point-mass inputs. Because both components share q, SA-IBF adds a single degree of freedom to IBF. Eq. (5) also gives the sensitivity of $\hat{P}$ to α: $\hat{P}$ is affine in α with $\|\partial_\alpha \hat{P}\|_1 = 2\big[1 - Z_h(q)\big] \le 2nq$, twice the IBF leakage, which tends to $2\big[1 - \binom{n}{h}2^{-n}\big]$ as $q \to 1/2$.

Physical interpretation. The kernel K_{IBF} is the computational-basis output distribution of a product Pauli channel with effective per-qubit flip rates $r_i = p_i^X + p_i^Y$, and uniform $r_i \equiv q$ gives Eq. (1) exactly [24, 32]. The $(1-\alpha)K_{\mathrm{IBF}}$ component therefore describes Pauli bit-flip noise, and K_{SA} as a whole is a model of the output distribution (Supplementary Note 1). The fitted $\hat{\alpha}$ is an effective parameter of this model and is specific to a circuit family on one processor. A simulation with readout error only, at the median readout error rates of the `ibm_boston` processor (Boston), gives a vanishing median $\hat{\alpha}$ in each of the seven configurations, so readout error alone does not account for the fitted weight (Supplementary Table 4).

Separation from heterogeneous bit-flip noise. The sector component is also distinct from heterogeneous bit-flip noise, in which each qubit has its own flip rate. Let $\mathcal{F} := \{K_r : r \in [0,1]^n\}$ with $K_r(y \mid x) := \prod_{i:x_i\neq y_i} r_i \prod_{i:x_i=y_i}(1-r_i)$ be the family of bit-flip kernels with per-qubit rates, which contains K_{IBF} at equal rates.

Proposition 1 (Separation from heterogeneous bit-flip kernels) *For $n \ge 2$, $q \in (0, 1/2)$, and $\alpha \in (0,1]$, the kernel $K_{\mathrm{SA}}(\cdot \mid \cdot; \alpha, q)$ lies outside the closed convex hull of $\mathcal{F}$.*

Every kernel in the closed convex hull of $\mathcal{F}$ depends on (x,y) only through the flip pattern $x \oplus y$, whereas $K_{\mathrm{SA}}(e_2 \mid e_1) - K_{\mathrm{SA}}(e_1 \oplus e_2 \mid 0^n) = \alpha q^2(1-q)^{n-2}/Z_{e_1}(q) > 0$, where e_i has a single one at position i (proof in Supplementary Note 2).

Fit objective. Let $P^{\mathrm{tr}}_{\mathrm{obs}}$ and $P^{\mathrm{te}}_{\mathrm{obs}}$ denote disjoint training and test histograms, the first with N shots. The Hellinger distance is $D_{\mathrm{H}}(P,Q) := 2^{-1/2}\|\sqrt{P}-\sqrt{Q}\|_2$, a bounded minimum-distance objective that is well defined with zero-count bins [33–35]. For a fixed search region $[0,1]\times[q_{\min},q_{\max}]$ with $0<q_{\min}<q_{\max}<1/2$ chosen in Methods, the fitted parameters are

$$(\hat{\alpha}(P_{\mathrm{ideal}},P^{\mathrm{tr}}_{\mathrm{obs}}),\hat{q}(P_{\mathrm{ideal}},P^{\mathrm{tr}}_{\mathrm{obs}}))\in\arg\min_{(\alpha,q)\in[0,1]\times[q_{\min},q_{\max}]} D_{\mathrm{H}}\Big(\hat{P}(\cdot;\alpha,q,P_{\mathrm{ideal}}),\,P^{\mathrm{tr}}_{\mathrm{obs}}\Big),$$

with ties broken by preferring smaller α. The rank diagnostic of Methods is evaluated at the constrained optimum. Accuracy is measured by the Hellinger distance $D_{\mathrm{H}}(\hat{P}(\cdot;\hat{\alpha},\hat{q},P_{\mathrm{ideal}}),P^{\mathrm{te}}_{\mathrm{obs}})$ to the test histogram, and D_{H} denotes this distance to test data throughout unless stated otherwise.

2.2 Identifiability of SA-IBF

We now establish local identifiability of (α,q) from $\hat{P}$ (Eqs. (3) and (4)) and bound the information on α as $q\to 0$.

Assumption 1 (Regularity of the identifiability domain) *Fix $n\geq 2$. The parameters satisfy $\alpha\in[0,1-\epsilon]$ for some $\epsilon\in(0,1)$, $q\in(0,1/2)$, and the ideal distribution P_{ideal} lies in the strictly positive interior $\mathrm{Int}(\Delta_{2^n})$. The observation model is N-shot multinomial with probabilities $\hat{p}_y(\alpha,q,P_{\mathrm{ideal}})$ given by Eq. (4); P_{ideal} is known and (α,q) is the estimand.*

Theorem 1 (SA-IBF identifiability) *Under Assumption 1, let $J(\alpha,q,P_{\mathrm{ideal}}) := [\partial_\alpha\hat{P},\ \partial_q\hat{P}]\in\mathbb{R}^{2^n\times 2}$.*

(a) *Small-q local identifiability. There exists a threshold $q_1(n,\epsilon)\in(0,1/2)$ depending only on (n,ϵ) such that for every $(\alpha,q)\in[0,1-\epsilon]\times(0,q_1)$, the set $\{P_{\mathrm{ideal}}\in\mathrm{Int}(\Delta_{2^n}):\mathrm{rank}\,J<2\}$ is a proper real-analytic subvariety of the simplex interior in its affine coordinates. On its open dense complement, the analytic extension of $(\alpha,q)\mapsto\hat{P}$ is locally injective; for $\alpha=0$ this means local injectivity relative to the stated parameter domain.*

(b) *Computer-assisted finite-n Fisher upper bound and Cramér–Rao divergence ($2\leq n\leq 12$). For every $(\alpha,q,P_{\mathrm{ideal}})\in[0,1-\epsilon]\times(0,1/(8n)]\times\mathrm{Int}(\Delta_{2^n})$ with $n\in\{2,\dots,12\}$, the marginal multinomial Fisher information for α under N-shot observations of $\hat{P}$ satisfies*

$$I_{\alpha|q}\leq c_+(n,\alpha)\,Nq^2,\qquad c_+(n,\alpha):=\frac{10\,n^4}{(1-\alpha)^5},\tag{6}$$

with the projection convention $I_{\alpha|q}:=I_{\alpha\alpha}-I_{\alpha q}^2/I_{qq}$ when $I_{qq}>0$ (the Schur complement; when I is rank-deficient but $I_{qq}>0$ the Schur complement equals 0), and $I_{\alpha|q}:=I_{\alpha\alpha}$ when $I_{qq}=0$. When $I_{\alpha|q}>0$, any regular unbiased estimator $\hat{\alpha}$ at a two-sided interior point, e.g. $\alpha\in(0,1-\epsilon)$ and $q\in(0,1/(8n))$, with unbiasedness and regularity in a neighborhood, satisfies $\mathrm{Var}(\hat{\alpha})\geq 1/(c_+(n,\alpha)\,Nq^2)$; when $I_{\alpha|q}=0$, α carries no first-order efficient information and the Cramér–Rao lower bound is $+\infty$. For fixed (n,N,α) with $\alpha<1$, this gives a uniform lower envelope $\Omega(q^{-2})$ for the population Cramér–Rao lower bound as $q\to 0$ whenever $I_{\alpha|q}>0$; if $I_{\alpha|q}=0$, the local bound is $+\infty$.

Theorem 2 (Generic support-face identifiability) *Fix $n \geq 2$, $\epsilon \in (0,1)$, and a nonempty known support $T \subseteq \{0,1\}^n$, and let* relint Δ_T *denote distributions that are strictly positive on T and zero off T. Under the SA-IBF observation map over all output bitstrings, there exists $q_{\text{face}}(n,\epsilon) > 0$ such that, for every $(\alpha, q) \in [0, 1-\epsilon] \times (0, q_{\text{face}})$, the set of $P_{\text{ideal}} \in$* relint Δ_T *for which* rank $J(\alpha, q, P_{\text{ideal}}) < 2$ *is either empty or a proper real-analytic subvariety of the support face. On its relative open dense complement, the analytic extension of $(\alpha, q) \mapsto \hat{P}$ is locally injective; for $\alpha = 0$ this means local injectivity relative to the stated parameter domain. Moreover, the joint rank-deficient set in $(0, 1-\epsilon) \times (0, 1/2) \times$* relint Δ_T *is relative-Lebesgue-null.*

Theorem 3 (Rank-conditioned local identifiability on a support face) *Fix $n \geq 2$, a known support $T \subseteq \{0,1\}^n$, $P_0 \in$* relint Δ_T*, $\alpha_0 \in (0,1)$, and $q_0 \in (0, 1/2)$. Let $\Phi_{P_0}(\alpha, q) := \hat{P}(\cdot; \alpha, q, P_0)$ and $J(\alpha_0, q_0, P_0) := D_{(\alpha,q)}\Phi_{P_0}(\alpha_0, q_0)$. If $J(\alpha_0, q_0, P_0)$ has column rank 2, equivalently if some 2×2 row minor of J is nonzero, then Φ_{P_0} is locally injective at (α_0, q_0).*

Theorem 2 holds for generic P_{ideal} on a fixed support face, and Theorem 3 drops the small-q requirement by conditioning on the rank of J at the given point. The bound in Theorem 1(b) is computer-assisted. The Fisher aggregation is analytic after reduction to four finite-n pointwise residual envelopes, which are certified by interval arithmetic for $2 \leq n \leq 12$ (Remark S1), and the remaining contributions for each Hamming distance are exact in closed form. Proofs of Theorems 1 to 3 are in Supplementary Note 4.

Rank diagnostic. Theorem 3 holds for any flip rate $q \in (0, 1/2)$ and on any support face. It therefore covers the UCCSD and Magnetization ideal outputs, which lie on symmetry-restricted support faces, and the MolQAE fits at high $\hat{q}$. We check its rank condition for every circuit of the main data set with the diagnostic of Definition 1 (Fig. 1b), which requires strictly positive predictions, a nonsingular Fisher information, $\kappa(J) < 10^3$, and Cramér–Rao standard errors below 0.1 (Methods). Every fit passes, both with the 15,000-shot training histograms (Fig. 2b,c) and with the full 30,000 shots, and every fit lies in the parameter interior that Theorem 3 requires (Supplementary Note 5). Refitting resampled histograms gives standard deviations of $\hat{\alpha}$ close to these Cramér–Rao standard errors (Supplementary Table 3).

The conditioning of the fits follows Eq. (5). As $q \to 0$ the population curves for different α merge (Fig. 2a), so the two Jacobian columns approach collinearity. Across all 230 fits, the condition number is largest at small fitted flip rates and smallest near $\hat{q} = 1/2$ (Fig. 2b). We refer to $\hat{q} < 0.15$, $0.15 \leq \hat{q} < 0.40$, and $\hat{q} \geq 0.40$ as the signal, intermediate, and saturation regimes. The molecular fits lie in the saturation regime, where IBF leakage is large and the Jacobian is well conditioned. The Magnetization fits lie in the signal regime and have the largest condition numbers and standard errors, which remain inside both diagnostic thresholds (Fig. 2b,c).

2.3 Comparison with baseline noise models

To test whether the sector-preserving weight improves prediction of test data, we fit SA-IBF, IBF, and three other baseline noise models (Methods) to the 230 circuits of the main data set, run on Boston (Table 1). SA-IBF has lower D_{H} than IBF on

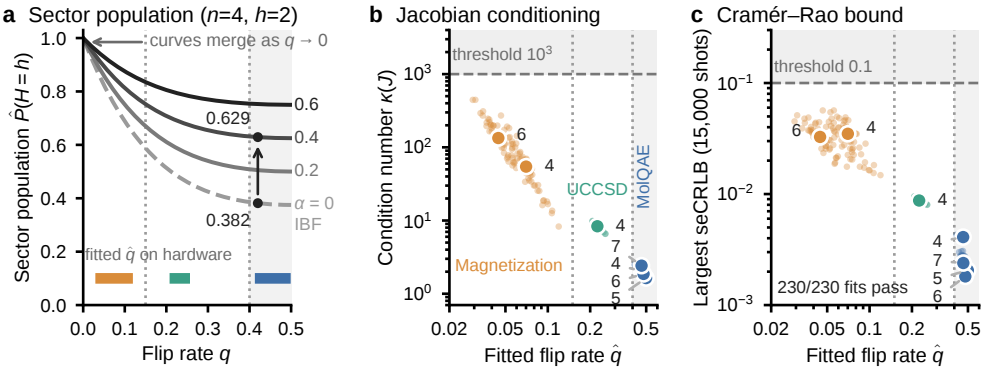


Fig. 2 Sector population and conditioning of the hardware fits. **a** Population of the ideal sector, $\hat{P}(H{=}h)$, against the flip rate for four values of α; the dashed curve is IBF, and the arrow shows the population gained by one α curve at fixed q. Colored bars give the range of fitted flip rates on hardware in each family (blue, MolQAE; orange, Magnetization; green, UCCSD). **b** Condition number of the analytic Jacobian and **c** larger of the two Cramér–Rao standard errors (seCRLB) of $(\hat{\alpha}, \hat{q})$ for the 15,000-shot training histograms, for the fit of each circuit in the main data set (small circles), against the fitted flip rate. Large circles are configuration medians labeled by n_q. Dotted lines separate the signal, intermediate, and saturation regimes, with saturation shaded grey, and dashed lines with grey horizontal bands mark the diagnostic thresholds of Definition 1. Per-configuration values are in Supplementary Tables 1 and 2.

every circuit of every configuration (Table 1 and Fig. 3). Hamming-weight shrinkage, the only two-parameter baseline, is less accurate than SA-IBF on every UCCSD and Magnetization configuration and matches it on MolQAE.

The improvement over IBF extends to other processors and physical registers (Fig. 4a). When the same 100 MolQAE circuits are run on four processors, SA-IBF has the lower D_{H} on every execution. Grouping 85 MolQAE $n_q{=}4$ executions by physical register gives a positive median improvement in each of the six groups on Boston and Kingston. Across these ten groups, the median improvement rises with the group's $\hat{\alpha}$, as expected from Eq. (5). To test whether fitted parameters carry over between registers, we apply the pooled $(\hat{\alpha}, \hat{q})$ of each group unchanged to the other groups on the same processor. In 12 of the 14 transfers, this gives a lower mean D_{H} than IBF fitted to the target group. The other two transfer from J1 and J2 into A2, the group with the smallest $\hat{\alpha}$ (Fig. 4b and Supplementary Note 7).

Replacing the within-sector shape of K_{sym} by a q-independent uniform component or an identity component changes D_{H} on MolQAE $n_q{=}4$ by a small fraction of the tolerance δ defined below (Supplementary Table 5). All these components give the same sector population at fixed (α, q), which therefore carries the improvement over IBF.

The gains are largest on MolQAE, intermediate on Magnetization, and smallest on UCCSD (Fig. 3 and Table 1). Fig. 5 tests the result in settings closer to downstream use. The improvement persists when both models are fitted to only 1k, 3k, or 10k training shots (Fig. 5a). On MolQAE it extends to the population of the ideal sector and to the populations of all sectors (Fig. 5b and Supplementary Table 10). The improvement also holds under zero-noise extrapolation (ZNE) of the fitted predictions with gate folding. For 15 circuits run at the original depth and with folded gates,

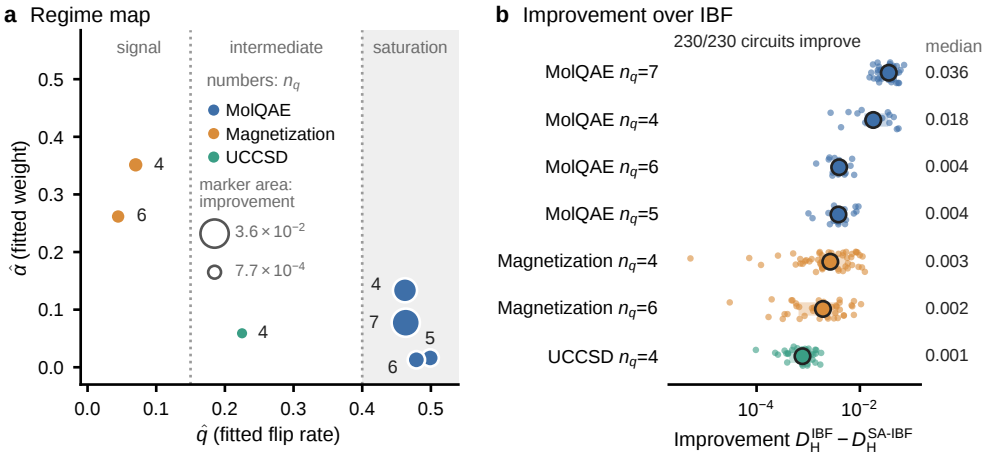


Fig. 3 Regime map and improvement over IBF. **a** Median $(\hat{q}, \hat{\alpha})$ of each configuration, colored by family (blue, MolQAE; orange, Magnetization; green, UCCSD) and sized by the median improvement in D_{H}; dotted lines separate the signal, intermediate, and saturation regimes, with the saturation regime shaded grey. **b** Improvement $D_{\mathrm{H}}^{\mathrm{IBF}} - D_{\mathrm{H}}^{\mathrm{SA\text{-}IBF}}$ of every circuit (small dots) on a logarithmic axis, with the median of each configuration (large circles, values at right) and its interquartile range (shaded bars), colored as in **a**.

Table 1 Median Hellinger distance to test data for SA-IBF and four baseline noise models in each configuration. Lower D_{H} is better. Δ is the median paired difference, IBF minus SA-IBF, so positive values favor SA-IBF. Wins count circuits where SA-IBF has lower D_{H} than IBF. N_{circ} is the number of circuits. Baselines are defined in Methods: Depol. is depolarization toward uniform, Conc. is concentration shrinkage, and HW-shr. is Hamming-weight shrinkage. Shading marks SA-IBF, and bold marks the smallest median in each row, ranked before rounding.

Family	n_q	N_{circ}	SA-IBF	IBF	Depol.	Conc.	HW-shr.	Δ	Wins
UCCSD	4	30	**0.2250**	0.2261	0.2887	0.6141	0.2887	$+7.7\times10^{-4}$	30/30
Magnetization	4	60	**0.0754**	0.0780	0.1437	0.2895	0.1437	$+2.7\times10^{-3}$	60/60
Magnetization	6	40	**0.0930**	0.0953	0.2236	0.3087	0.2232	$+1.9\times10^{-3}$	40/40
MolQAE	4	20	**0.2218**	0.2399	0.2300	0.6821	0.2219	$+1.8\times10^{-2}$	20/20
MolQAE	5	20	**0.0809**	0.0857	0.0809	0.8221	0.0809	$+3.8\times10^{-3}$	20/20
MolQAE	6	20	0.0863	0.0900	0.0925	0.8477	**0.0862**	$+4.0\times10^{-3}$	20/20
MolQAE	7	40	**0.1946**	0.2355	0.2072	0.7727	0.1947	$+3.6\times10^{-2}$	40/40

the extrapolated SA-IBF prediction lies closer to P_{ideal} in every metric and for every circuit (Fig. 5c). We also extrapolate the probability Π that qubits 0 and 1 both read 0. Across 50 circuits, the SA-IBF value is closer to its ideal value than the IBF value on all but four circuits (Fig. 5d and Supplementary Table 11).

2.4 Accuracy and computational cost against Ji-BO

We compare SA-IBF with Ji-BO [20], a noise model built from device calibration data whose parameters are fitted by Bayesian optimization with the tree-structured

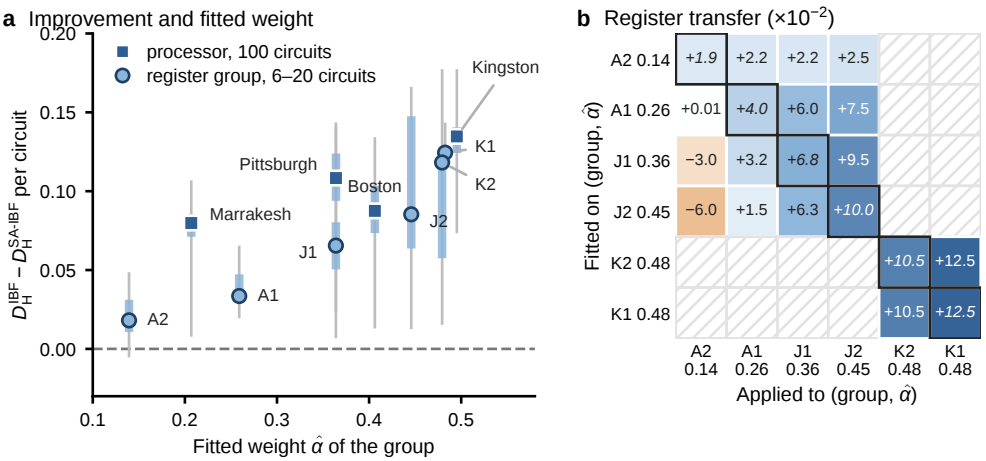


Fig. 4 Improvement over IBF across processors and physical registers. **a** Improvement $D_{\mathrm{H}}^{\mathrm{IBF}} - D_{\mathrm{H}}^{\mathrm{SA\text{-}IBF}}$ of each circuit against the pooled $\hat{\alpha}$ of each group: the same MolQAE circuits on four processors (squares) and MolQAE circuits grouped by physical register (circles). Markers give the median, thick bars the interquartile range, and grey bars the full range. **b** Mean D_{H} of IBF fitted to the target group minus that of the source group's $(\hat{\alpha}, \hat{q})$ applied unchanged, in units of 10^{-2}, with groups ordered by $\hat{\alpha}$. Italic diagonal entries give each group's own improvement. Transfers are tested between groups on the same processor, and hatched cells pair groups on different processors. Positive values favor SA-IBF. In **b**, positive values are blue and negative values orange. Register groups J1–K2 are defined in Supplementary Table 8.

Parzen estimator (TPE) in Optuna [36, 37]. Ji-BO has 20 parameters, and a sector-aware variant adds a sector weight, for 21 parameters (Methods). The comparison uses five-fold cross-validation over circuits on MolQAE at four qubit numbers and on Magnetization at n_q=4 (Table 2), with 100 TPE trials per fold for each Ji-BO variant unless stated otherwise. Two models agree when the 95% confidence interval of their difference in D_{H} lies within $\pm\delta$, with tolerance δ=0.01 (Methods).

Accuracy. On MolQAE, SA-IBF and sector-aware Ji-BO agree within the tolerance at $n_q \in \{4, 5, 7\}$ (Fig. 6a and Table 2). At n_q=6, sector-aware Ji-BO is more accurate, with its confidence interval entirely below $-\delta$, and both Ji-BO variants reach the same fit. The 20-parameter Ji-BO is less accurate than SA-IBF at n_q=4 by more than five times the tolerance. On Magnetization at n_q=4, SA-IBF has a lower mean D_{H} than every Ji-BO variant tested, including sector-aware Ji-BO with up to 500 warm-started trials (Fig. 6a and Table 2).

Computational cost. Each SA-IBF fold is fitted in seconds, whereas each sector-aware Ji-BO fold takes minutes to over half an hour under the same CPU implementation (Fig. 6b). SA-IBF is therefore fitted over 200 times faster at every qubit number, and a GPU-accelerated Ji-BO run at n_q=4 gives a similar speedup (Table 2 and Supplementary Table 12).

Zero-noise extrapolation. On the 35 Boston circuits of the three-scale gate-folding experiment, for which Ji-BO uses the calibration from the day of execution, the extrapolated Π from the SA-IBF fit is closer to its ideal value than that from the

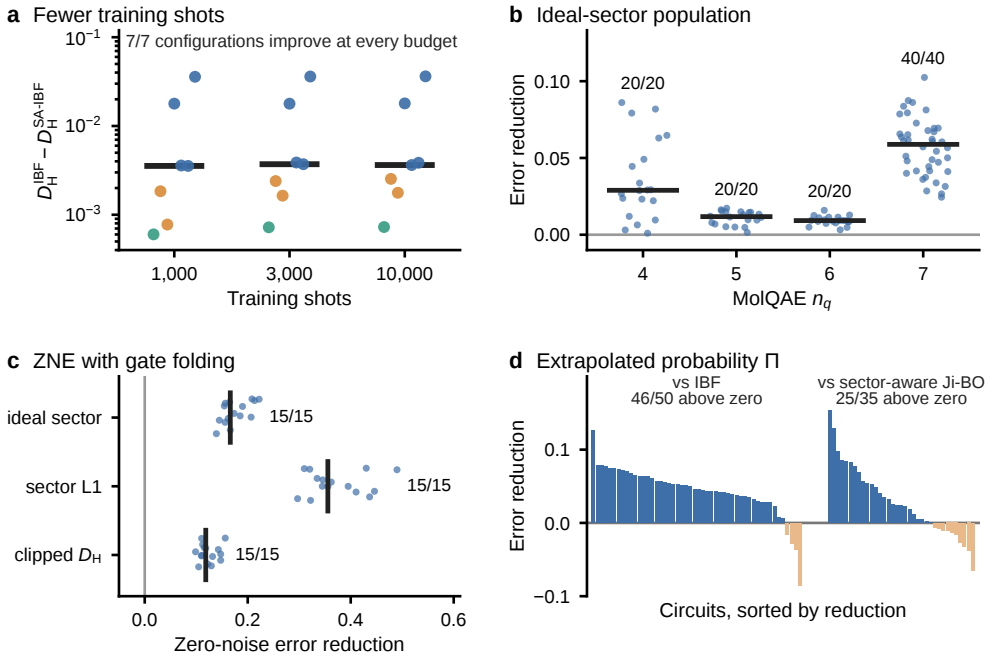


Fig. 5 Fewer training shots, sector observables, and zero-noise extrapolation (ZNE). **a** Median improvement $D_{\mathrm{H}}^{\mathrm{IBF}} - D_{\mathrm{H}}^{\mathrm{SA\text{-}IBF}}$ after fitting to subsampled training data; each colored circle is one configuration, colored by family as in Fig. 2, and black lines are medians across configurations. **b** Reduction of the error in the population of the ideal sector for each MolQAE circuit (dots) and its median (black line) at each qubit number. **c** Reduction of the zero-noise error with respect to P_{ideal} for each pair of original and folded circuits (dots) and its median (black line) in three metrics: the error in the population of the ideal sector, the L1 error of the sector populations, and the clipped D_{H}. **d** Reduction of the absolute error of the zero-noise-extrapolated probability Π that qubits 0 and 1 both read 0, for each circuit, relative to IBF (left) and to sector-aware Ji-BO on the circuits with a calibration from the day of execution (right), sorted within each comparison; blue bars favor SA-IBF and orange bars favor the comparator. Labels in **b**–**d** give wins over circuits or pairs.

sector-aware Ji-BO fit on 25 circuits. Its median error is also lower (Fig. 5d and Supplementary Table 11).

3 Discussion

The identifiability results yield a check that can be run for every fit. Theorem 1(a) establishes generic local identifiability on the open simplex interior for small q. Theorem 3 needs only a full-rank Jacobian at the fitted parameters, and the diagnostic of Definition 1 checks this condition from the analytic Jacobian at the same order of cost as one prediction (Methods). All 230 fits of the main data set pass it. The construction extends to any partition of the outcomes into symmetry sectors. Theorem 3 and the diagnostic apply unchanged, because they use only the block structure and a nonzero 2×2 Jacobian minor. Theorem 2 and the Fisher bound are proved for the Hamming partition, through a distance-two witness and a decomposition by Hamming distance.

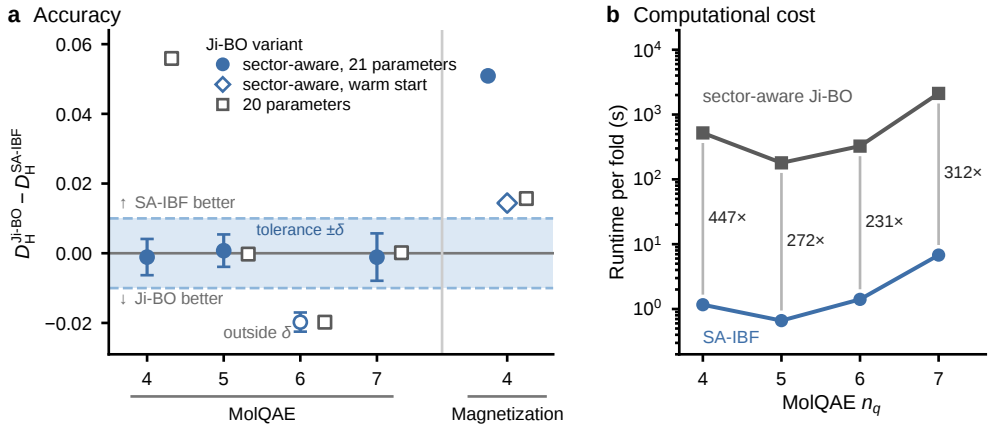


Fig. 6 SA-IBF against Ji-BO. **a** Ji-BO minus SA-IBF D_{H} for every Ji-BO variant: sector-aware with 21 parameters (blue circles; error bars are 95% confidence intervals over folds, computed as in Methods) and 20 parameters (open grey squares) on MolQAE, and both variants plus warm-started sector-aware Ji-BO (open blue diamond) on Magnetization. The shaded band is the tolerance $\pm\delta$. On MolQAE, filled circles mark intervals inside the band and the open circle an interval outside it. The other points are fold means without intervals. Positive values favor SA-IBF. **b** CPU runtime per fold for both models, labeled with the speedup of SA-IBF.

Table 2 Accuracy and computational cost of SA-IBF and Ji-BO under five-fold cross-validation. Lower D_{H} is better. The Ji-BO columns give the sector-aware variant with 21 parameters and the variant with 20 parameters. Δ is sector-aware Ji-BO minus SA-IBF fold-mean D_{H}, with its 95% confidence interval (CI) over folds (Methods), so positive values favor SA-IBF. Rows whose interval lies inside $\pm\delta$ agree within the tolerance. Speedup is the summed five-fold CPU time of sector-aware Ji-BO divided by that of SA-IBF under the same implementation. [a]Best of four warm-started sector-aware search strategies. Supplementary Table 12 details the n_q=4 rows. Shading marks SA-IBF.

		D_{H}						
			Ji-BO					
n_q	N_{circ}	SA-IBF	21 par.	20 par.	Δ	95% CI	Speedup	Outcome
MolQAE								
4	100	0.15146	0.15035	0.20736	−0.00111	[−0.0063, +0.0041]	447×	Within δ
5	20	0.08106	0.08179	0.08082	+0.00072	[−0.0039, +0.0054]	272×	Within δ
6	20	0.08492	0.06514	0.06514	−0.01978	[−0.0225, −0.0170]	231×	Ji-BO better
7	40	0.20115	0.20002	0.20130	−0.00114	[−0.0079, +0.0057]	312×	Within δ
Magnetization								
4	60	0.0776	0.0920[a]	0.0933	+0.0144	—	—	SA-IBF lowest

The fitted parameters have a direct reading through Eq. (5), and $\hat{q}$ places each circuit family in the signal, intermediate, or saturation regime. Refitting both parameters when the processor or circuit family changes takes seconds per fold (Fig. 6b). At MolQAE n_q=6, sector-aware Ji-BO reaches a lower D_{H} at higher computational cost (Table 2). Identifying the physical terms behind this gap is a direction for future work. All empirical evaluations use 4–7-qubit circuits and a classically available P_{ideal}.

The gate-folding experiments (Fig. 5c,d) suggest using SA-IBF inside zero-noise extrapolation and other error-mitigation pipelines [9, 12]. The rank diagnostic can flag ill-conditioned fits before a model enters such a pipeline.

4 Methods

4.1 Circuits and hardware execution

MolQAE circuits load a molecule from a 512-molecule subset of QM9 [38] into n_q qubits with one U3 rotation per qubit followed by a ring of CNOT gates. They then apply a trained quantum autoencoder with $n_q - 2$ latent qubits [31] and invert the state preparation, so that the probability of the all-zero bitstring is the reconstruction fidelity. The molecules come from the part of this subset not used for autoencoder training, and the trained checkpoints reach mean noiseless reconstruction fidelities between 0.9744 and 0.9761 on the autoencoder training set. UCCSD circuits prepare a four-qubit, two-particle state from the Hartree–Fock bitstring with two layers of a Jordan–Wigner double excitation and nearest-neighbor single excitations, with angles drawn uniformly from $[-0.2, 0.2]$. Magnetization circuits start from a bitstring of Hamming weight $n_q/2$ and apply four layers, each made of random Z rotations followed by a brick-wall of random $XX{+}YY$ gates. Three diagnostic families complete the set. GHZ circuits apply a Hadamard gate, a random Z rotation, and a CNOT chain. Clifford circuits apply five layers of random two-qubit Clifford gates on random qubit pairs. Near-boundary circuits follow a depth-four even-parity circuit with a rotation $R_Y(0.7)$ between two R_{XX} gates, which moves 12% of the ideal mass out of the even-parity sector. Ideal distributions are computed by exact statevector simulation in Qiskit [39].

Circuits were transpiled at optimization level 1 without a fixed initial layout and executed with the Qiskit Runtime sampler. The 100 circuits of the four-processor data set were transpiled once and run unchanged on all four processors. The four-processor data set used the default sampler options, which disable dynamical decoupling, and all other executions used XY4 dynamical decoupling. No readout-error mitigation or twirling was requested, and all fits use the measured counts directly. After transpilation, the MolQAE circuits contain 101, 203, 325, and 497 CZ gates at $n_q = 4$, 5, 6, and 7, and the UCCSD and Magnetization circuits contain 100 and 12–20.

4.2 Hardware data and splits

The analyses use the IBM Heron 156-qubit processors `ibm_boston`, `ibm_kingston`, `ibm_pittsburgh`, and `ibm_marrakesh`. All 1,040 executions use 30,000 shots (Table 3).

The data span six families: MolQAE, UCCSD, Magnetization, GHZ, Clifford, and near-boundary, with $n_q \in \{4, 5, 6, 7\}$. Comparisons with Ji-BO use MolQAE at n_q=4 (N_{circ}=100) and at $n_q \in \{5, 6, 7\}$ (N_{circ}=20, 20, 40), and Magnetization at n_q=4 (N_{circ}=60). The main data set of 230 circuits covers MolQAE, UCCSD, and Magnetization in seven configurations of family and n_q. GHZ, Clifford, and near-boundary circuits enter only the Boston diagnostics (Supplementary Note 7).

Table 3 lists all subsets. The four-processor data set runs the same 100 MolQAE n_q=4 circuits once on each of the four processors. About ten days after the Kingston

run of this data set, a second Kingston run uses 30 MolQAE n_q=4 circuits with a different molecule set on the same physical qubits and layout (Supplementary Note 7).

Table 3 Hardware data, counted as circuit executions, with the role of each subset in the analyses.

Processor	Subset	Executions	Role
All four	Four-processor data set	4×100	MolQAE n_q=4 comparison across processors, and with Ji-BO on Boston
Boston	Main data set I	190	UCCSD n_q=4 (30), Magnetization n_q=4 (60), and MolQAE n_q=4, ..., 7 (20,20,20,40)
Boston	Main data set II	40	Magnetization n_q=6 (40)
Boston	Diagnostic circuits	120	Diagnostic families and two-scale gate folding (Supplementary Table 9)
Pittsburgh	Comparison with Boston	30	Magnetization n_q=4 (30)
Kingston	Comparison with Boston	80	Magnetization n_q=4 and MolQAE n_q=6 (40 each)
Kingston	Second Kingston run	30	MolQAE n_q=4 on the register of the Kingston four-processor run
Boston	Three-scale gate folding	105	MolQAE n_q=4, 35 circuits at noise scales 1, 2, 3
Kingston	Three-scale gate folding	45	MolQAE n_q=4, 15 circuits at noise scales 1, 2, 3

For the main data set, every 30,000-shot histogram is split at the shot level by binomial thinning of each bitstring count. For each observed count c_y, the training count is Binomial$(c_y, 0.5)$ and the test count is c_y minus the training count. Splits use a fixed random seed, and test histograms are used only for evaluation. The resulting training and test histograms contain approximately 15,000 shots each, and each circuit is fitted to its own training histogram and evaluated on its test histogram.

Comparisons across circuits use five-fold cross-validation with shuffled folds and a fixed seed. Each fold uses a fifth of the circuits as test circuits: 20 of 100 for MolQAE at n_q=4, 4 of 20 at $n_q \in \{5, 6\}$, 8 of 40 at n_q=7, and 12 of 60 for Magnetization. In these comparisons and in the four-processor data set, each model fits one parameter set per training fold by minimizing the mean Hellinger distance over the full 30,000-shot histograms of the training circuits. The fit is evaluated on the full test histograms.

For the fits to fewer training shots in Fig. 5a, 1,000, 3,000, or 10,000 training counts are drawn without replacement from each 30,000-shot histogram, and the remaining counts form the test histogram. Each circuit and number of shots uses five draws, and each plotted value is the median paired difference over circuits and draws.

4.3 Model fitting

SA-IBF. SA-IBF uses the box $\alpha \in [0, 1]$, $q \in [0.001, 0.499]$. The optimizer first evaluates a 31-point uniform α grid on $[0, 1]$. For each grid value it solves the one-dimensional q subproblem by bounded scalar minimization in SciPy [40] with tolerance 10^{-4}. The best grid point initializes a two-dimensional Nelder–Mead refinement,

constrained to the same box, with x-tolerance 10^{-5}, function tolerance 10^{-7}, and at most 300 iterations. The refined solution is accepted if it remains feasible and lowers the Hellinger objective, and the grid solution is kept otherwise. Raising $q_{\max}$ to 0.99 leaves six of the seven configurations of the main data set unchanged and changes the median D_{H} of MolQAE n_q=5, where most fits lie at the bound, by $7{\times}10^{-5}$ (Supplementary Note 6). To test the within-sector shape, K_{sym} is replaced by its column-normalized or row/column-averaged form, by a q-independent uniform distribution within each sector, or by the identity within the sector. Each variant is fitted under the same five-fold cross-validation, and a fine-grid check fits SA-IBF on a 21×17 grid of (α, q) followed by Nelder–Mead refinement (Supplementary Table 5). The three-parameter variants of SA-IBF in Supplementary Table 13 give the two components separate flip rates, or assign two flip rates to the logical qubits in two contiguous blocks or in even and odd positions.

Baseline noise models. Each baseline replaces K_{SA} by a Markov kernel applied to P_{ideal} and is fitted by the same Hellinger objective on the same split. Depolarization toward uniform uses $(1-\theta)\mathbf{1}[y{=}x] + \theta\, 2^{-n}$, and concentration shrinkage uses $(1-\theta)\mathbf{1}[y{=}x] + \theta\, \mathbf{1}[y{=}x^\star]$ with $x^\star = \arg\max_x P_{\mathrm{ideal}}(x)$. Hamming-weight shrinkage uses $(1-F)\mathbf{1}[y{=}x] + F\, g_\lambda(y)$ with $g_\lambda(y) \propto e^{-\lambda H(y)}$, and IBF uses $K_{\mathrm{IBF}}(y \mid x; q)$ with the same q bounds and bounded scalar minimization as SA-IBF. Depolarization and concentration shrinkage optimize their scalar interpolation parameter on $[0, 1]$. Hamming-weight shrinkage uses an 11×9 coarse grid over $F \in [0, 1]$ and $\lambda \in [0, 4]$, followed by Nelder–Mead refinement under $F \in [0, 1]$, $\lambda \in [0, 5]$.

Simulation with readout error only. To test whether readout error alone produces a nonzero $\hat{\alpha}$, each P_{ideal} is passed through the n-qubit tensor-product readout confusion matrix built from the median per-qubit readout error rates of Boston over all 156 qubits, $p_{10}{\approx}0.0037$ and $p_{01}{\approx}0.0051$. One 30,000-shot multinomial sample is drawn from the result, and SA-IBF is fitted to it. The difference $\hat{\alpha}_{\mathrm{obs}} - \hat{\alpha}_{\mathrm{RO}}$ between the hardware fit and the readout-only fit of each circuit is summarized by its median and interquartile range (Supplementary Table 4).

4.4 Rank diagnostic

Theorem 3 gives local identifiability wherever the analytic Jacobian has rank 2. We check this condition numerically at seven representative operating points and for the fit of every circuit in the main data set. The representative points pair the median $(\hat{\alpha}, \hat{q})$ of each configuration, fitted to the full 30,000-shot histograms, with the ideal distribution of its first circuit. The check for each circuit uses its own fit $(\hat{\alpha}_i, \hat{q}_i, P^{(i)}_{\mathrm{ideal}})$.

Definition 1 (Rank diagnostic) For each fitted $(\hat{\alpha}, \hat{q}, P_{\mathrm{ideal}})$:

(i) Require the strict-positivity precondition $\min_y \hat{P}(y; \hat{\alpha}, \hat{q}, P_{\mathrm{ideal}}) > 0$ (which can fail when $\hat{\alpha} = 1$ and P_{ideal} has sparse sector support, since K_{sym} vanishes off-support). The remaining steps are computed only when it holds.

(ii) Compute $J(\hat{\alpha}, \hat{q}, P_{\mathrm{ideal}}) \in \mathbb{R}^{2^n \times 2}$ analytically.

(iii) Compute singular values $\sigma_{\min}(J) \le \sigma_{\max}(J)$ and condition number $\kappa(J) := \sigma_{\max}/\sigma_{\min}$.

(iv) Compute multinomial Fisher information $I = J^\top \mathrm{diag}(N/\hat{P})J \in \mathbb{R}^{2\times 2}$ for N shots (well defined by step (i)).

(v) Compute the Cramér–Rao standard errors $\mathrm{seCRLB}_\alpha := \sqrt{[I^{-1}]_{11}}$, $\mathrm{seCRLB}_q := \sqrt{[I^{-1}]_{22}}$ when I is nonsingular.

The diagnostic *passes* at $(\hat{\alpha}, \hat{q}, P_{\mathrm{ideal}})$ if the positivity precondition (i) holds, I is nonsingular, $\kappa(J) < 10^3$, and $\max(\mathrm{seCRLB}_\alpha, \mathrm{seCRLB}_q) < 0.1$; otherwise it fails.

The condition-number bound limits how collinear the Jacobian columns can be. The Cramér–Rao threshold is a necessary condition for any regular unbiased estimator to reach standard errors below 0.1 in both coordinates for the given number of shots. The thresholds add numerical margins to the exact rank condition of Theorem 3.

Remark 1 The smallest singular value at the representative points is $\sigma_{\min}(J) = 1.86\times10^{-2}$. The Fisher information is linear in N, so moving from 30,000 shots to the 15,000-shot training histograms of the main data set multiplies every seCRLB by $\sqrt{2}$ and leaves $\kappa(J)$ unchanged. All reported passes hold for both numbers of shots.

Parameter recovery (Supplementary Table 3) uses the first ten circuits of each configuration of the main data set. For each circuit, 100 multinomial samples of 15,000 shots are drawn from the fitted SA-IBF prediction and 100 from the measured 30,000-shot histogram, and SA-IBF is refitted to every sample with the same optimizer.

The four pointwise residual bounds behind Theorem 1(b) were certified by outward-rounded interval arithmetic at 50-digit precision on 8,000 subintervals per case, using the mpmath library (https://mpmath.org) and SymPy [41] (Remark S1).

4.5 Ji-BO settings and agreement tolerance

Ji-BO has 20 parameters [20]. For one- and two-qubit gates, these are a scale factor on the calibrated error rate ($[0, 5]$) with an offset ($[-0.1, 0.1]$), an amplitude-damping offset, three coherent rotation angles (X, Y, Z and IX, ZX, ZZ), and a dephasing exponent ($[0.5, 2]$). A two-qubit dephasing offset and a ZZ crosstalk rate ($[0, 0.5]$) complete the gate parameters, and the remaining offsets and angles lie in $[-0.05, 0.05]$. Four readout parameters set the scale factors ($[-1, 1]$) and offsets ($[-0.1, 0.1]$) of correlated $|00\rangle \leftrightarrow |11\rangle$ and $|01\rangle \leftrightarrow |10\rangle$ flips on coupled qubit pairs. The sector-aware variant adds a weight $\alpha_{\mathrm{sector}} \in [0, 1]$ that mixes the simulator output with P_{ideal} restricted to its sectors of ideal mass above 0.01 and renormalized. Each TPE trial [36, 37] simulates the training circuits with the Qiskit Aer simulator at 30,000 shots, and warm-start points enqueued before the search count toward the trial budget. At $n_q \in \{5, 6, 7\}$, the sector-aware search also enqueues every warm start of the 20-parameter search at $\alpha_{\mathrm{sector}} = 0$. The main Ji-BO comparisons use 100 TPE trials per fold, and the Magnetization comparison also reports the best of four warm-started sector-aware search strategies with 200–500 trials (Supplementary Table 12). All Ji-BO variants are evaluated on the same folds as the corresponding SA-IBF rows.

Agreement with Ji-BO is judged from the five fold-level differences in D_{H}: their 95% confidence interval must lie within $\pm\delta$.

The tolerance δ=0.01 is about 38 times the half-width of 2.6×10^{-4} that multinomial shot noise gives for a fold mean. This half-width comes from 20,000 multinomial resamples of the test histograms with the five-fold predictions fixed (Supplementary Table 14). The confidence intervals in Table 2 exceed this floor, so their width reflects how the model difference varies across folds and circuits.

4.6 Sector observables and zero-noise extrapolation

The MolQAE observable checks use the sector populations $P(H{=}w)$. The population of the ideal sector is $P(H{=}w^\star)$, where $w^\star$ is the sector carrying most of the ideal mass ($w^\star$=0 for every MolQAE circuit), and the L1 error of the sector populations is $\sum_{w=0}^{n}|\hat{P}(H{=}w) - P_{\mathrm{obs}}^{\mathrm{te}}(H{=}w)|$. Both errors are computed against the test half of the binomial-thinning split, and wins count circuits where SA-IBF has the smaller error.

The two-scale gate-folding experiment uses 15 MolQAE n_q=4 circuits run on Boston at the original depth and with every two-qubit gate G replaced by $GG^\dagger G$, which triples the two-qubit gate count. SA-IBF and IBF are fitted separately to the full histogram at each depth, the two predictions are extrapolated to zero noise as $\hat{P}_0 = (3\hat{P}_1 - \hat{P}_3)/2$, and $\hat{P}_0$ is compared with P_{ideal}. Clipped D_{H} sets the negative entries of $\hat{P}_0$ to zero and renormalizes before computing the Hellinger distance.

The three-scale gate-folding experiment runs 35 MolQAE n_q=4 circuits on Boston at noise scales 1, 2, and 3, where scale 2 triples the first half of the two-qubit gates, and repeats 15 circuits on Kingston at the same scales. Its observable is the probability that qubits 0 and 1 both read 0, $\Pi(P) = \sum_{y:\, y_0=y_1=0} P(y)$, evaluated on each fitted prediction. The value is extrapolated to zero noise from scales 1 and 3 as $(3\Pi_1 - \Pi_3)/2$, or from all three scales as $3\Pi_1 - 3\Pi_2 + \Pi_3$, and its error is $|\Pi_0 - \Pi(P_{\mathrm{ideal}})|$. Sector-aware Ji-BO is fitted to each circuit at scales 1 and 3 with 100 TPE trials, using the calibration from the day of execution, available for the 35 Boston circuits.

The register analysis groups the MolQAE n_q=4 circuits by the physical register recorded for each execution. It uses the 35 Boston and 15 Kingston circuits of the three-scale experiment at scale 1, the 15 circuits of the two-scale experiment at the original depth, and the 20 MolQAE n_q=4 circuits of the main data set. This gives six groups on five registers with 85 circuits, and five-fold cross-validation is applied within each group. In the transfer test, the pooled $(\hat{\alpha}, \hat{q})$ of one group is applied unchanged to another group on the same processor and compared with IBF fitted to that group.

4.7 Computational cost and scaling

For one circuit with $m = 2^n$ output states, a direct SA-IBF prediction evaluates $K_{\mathrm{IBF}}^\top P_{\mathrm{ideal}}$ and $K_{\mathrm{sym}}^\top P_{\mathrm{ideal}}$ without any quantum-simulator call. The dense bit-flip component costs $O(nm^2) = O(n4^n)$ time if Hamming distances are formed on the fly and $O(m^2)$ memory if the kernel is materialized. It can also be implemented as a hypercube convolution in $O(n2^n)$ time. The sector component is block diagonal by Hamming weight, with $\sum_{w=0}^{n}\binom{n}{w}^2 = \binom{2n}{n} = O(4^n/\sqrt{n})$ source–target pairs, and the row normalizer $Z_x(q)$ depends only on $H(x)$, so only $n{+}1$ normalizers are precomputed. The analytic Jacobian and 2×2 Fisher matrix have the same order as one prediction plus $O(m)$ assembly after the q-derivative kernels are available. With N_{eval} objective

calls, a direct SA-IBF fit costs $O(N_{\text{eval}} n 4^n)$ time and can be streamed in $O(2^n)$ working memory, and N_{eval} is set by the optimizer budget and is independent of the number of shots. In the reported range $n_q \leq 7$, the largest prediction vector has 128 bins, and the per-fold runtimes of SA-IBF and sector-aware Ji-BO under the same five-fold cross-validation are shown in Fig. 6b. All timings are wall-clock times on one host with an Intel Xeon Gold 6330 CPU, and Qiskit Aer runs on a GPU use an NVIDIA RTX A5000. The Ji-BO time covers the TPE loop of noise-model construction, simulation, and scoring. The speedups in Table 2 divide the summed five-fold CPU time of sector-aware Ji-BO by the five-fold SA-IBF fit time, and the GPU speedup divides the per-fold times. The speedups exclude the cost of constructing P_{ideal}.

4.8 Statistics and reproducibility

The unit of analysis is one circuit execution. Comparisons with the baselines use the D_{H} of each circuit and report, for each configuration, the median paired difference, its interquartile range (Fig. 3b), and the number of circuits on which SA-IBF has the lower D_{H}, with $n = N_{\text{circ}}$ circuits per configuration (Table 1). Comparisons with Ji-BO use the five fold-level differences in fold-mean D_{H}, with $n = 5$ per qubit number. Their 95% confidence interval is the two-sided Student-t interval with four degrees of freedom, with the variance of the fold mean multiplied by $1 + 5N_{\text{test}}/N_{\text{train}}$ for the training circuits shared between folds [42], where N_{train} and N_{test} are the numbers of training and test circuits per fold. Uncertainties of pooled fitted parameters are standard errors or 95% intervals from at least 1,000 resamples of the circuits with replacement, and values after $\pm$ are standard errors. Each circuit was executed once on each processor it was run on. Replication across processors, physical registers, and calibration dates is reported in Fig. 4 and Supplementary Note 7.

Acknowledgments. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. This research was supported by the UTC Applied AI & Industry Partnerships Initiative, funded through the Center of Excellence in Applied Computational Science and Engineering (CEACSE). This study was also funded by the College of Engineering and Computer Science of UTC. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

References

[1] Sarovar, M. *et al.* Detecting crosstalk errors in quantum information processors. *Quantum* **4**, 321 (2020).

[2] Rudinger, K. *et al.* Experimental characterization of crosstalk errors with simultaneous gate set tomography. *PRX Quantum* **2**, 040338 (2021).

[3] Georgopoulos, K., Emary, C. & Zuliani, P. Modeling and simulating the noisy behavior of near-term quantum computers. *Phys. Rev. A* **104**, 062432 (2021).

[4] Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).

[5] Cerezo, M. *et al.* Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644 (2021).

[6] Abbas, A. *et al.* The power of quantum neural networks. *Nat. Comput. Sci.* **1**, 403–409 (2021).

[7] Mari, A., Bromley, T. R., Izaac, J., Schuld, M. & Killoran, N. Transfer learning in hybrid classical-quantum neural networks. *Quantum* **4**, 340 (2020).

[8] Thanasilp, S., Wang, S., Cerezo, M. & Holmes, Z. Exponential concentration in quantum kernel methods. *Nat. Commun.* **15**, 5200 (2024).

[9] Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).

[10] Nation, P. D., Kang, H., Sundaresan, N. & Gambetta, J. M. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum* **2**, 040326 (2021).

[11] Bravyi, S., Sheldon, S., Kandala, A., McKay, D. C. & Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A* **103**, 042605 (2021).

[12] Cai, Z. *et al.* Quantum error mitigation. *Rev. Mod. Phys.* **95**, 045005 (2023).

[13] Czarnik, P., Arrasmith, A., Coles, P. J. & Cincio, L. Error mitigation with Clifford quantum-circuit data. *Quantum* **5**, 592 (2021).

[14] Strikis, A., Qin, D., Chen, Y., Benjamin, S. C. & Li, Y. Learning-based quantum error mitigation. *PRX Quantum* **2**, 040330 (2021).

[15] Patil, H. P., Baron, D. & Zhou, H. *Q-Cluster: Quantum error mitigation through noise-aware unsupervised learning. Proc. IEEE International Conference on Quantum Computing and Engineering (QCE)*, 849–860 (2025).

[16] Liao, W.-Y. *et al.* Sample-efficient quantum error mitigation via classical learning surrogates. *Commun. Phys.* (2026). https://doi.org/10.1038/s42005-026-02827-w.

[17] Liao, W.-Y. *et al.* Demonstration of efficient predictive surrogates for large-scale quantum processors. *Nat. Commun.* **17**, 4731 (2026).

[18] Cheng, L., Chen, Y.-Q., Zhang, S.-X. & Zhang, S. Quantum approximate optimization via learning-based adaptive optimization. *Commun. Phys.* **7**, 83 (2024).

[19] Bonet-Monroig, X., Sagastizabal, R., Singh, M. & O'Brien, T. E. Low-cost error mitigation by symmetry verification. *Phys. Rev. A* **98**, 062339 (2018).

[20] Ji, Y., Roth, M., Kreplin, D. A., Polian, I. & Wilhelm, F. K. Data-efficient quantum noise modeling via machine learning. *Phys. Rev. Appl.* **25**, 034051 (2026).

[21] Flammia, S. T. & Wallman, J. J. Efficient estimation of Pauli channels. *ACM Trans. Quantum Comput.* **1**, 3 (2020).

[22] Harper, R., Flammia, S. T. & Wallman, J. J. Efficient learning of quantum noise. *Nat. Phys.* **16**, 1184–1188 (2020).

[23] van den Berg, E., Minev, Z. K., Kandala, A. & Temme, K. Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors. *Nat. Phys.* **19**, 1116–1121 (2023).

[24] Wallman, J. J. & Emerson, J. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A* **94**, 052325 (2016).

[25] Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* 10th anniversary edn (Cambridge University Press, 2010).

[26] Dalzell, A. M., Hunter-Jones, N. & Brandão, F. G. Random quantum circuits transform local noise into global white noise. *Commun. Math. Phys.* **405**, 78 (2024).

[27] Aharonov, D., Gao, X., Landau, Z., Liu, Y. & Vazirani, U. *A polynomial-time classical algorithm for noisy random circuit sampling. Proc. 55th Annual ACM Symposium on Theory of Computing (STOC)*, 945–957 (2023).

[28] Kruskal, J. B. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **18**, 95–138 (1977).

[29] Allman, E. S., Matias, C. & Rhodes, J. A. Identifiability of parameters in latent structure models with many observed variables. *Ann. Stat.* **37**, 3099–3132 (2009).

[30] Krantz, S. G. & Parks, H. R. *A Primer of Real Analytic Functions* 2nd edn (Birkhäuser, 2002).

[31] Pan, Y. *et al. MolQAE: Quantum autoencoder for molecular representation learning. Proc. IEEE International Conference on Quantum Artificial Intelligence (QAI)*, 98–105 (2025).

[32] Yadavalli, S. A. & Marvian, I. Noisy quantum trees: Infinite protection without correction. *npj Quantum Inf.* **11**, 151 (2025).

[33] Beran, R. Minimum Hellinger distance estimates for parametric models. *Ann. Stat.* **5**, 445–463 (1977).

[34] Lindsay, B. G. Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Stat.* **22**, 1081–1114 (1994).

[35] Vidyashankar, A. N. & Collamore, J. F. Rare event analysis for minimum Hellinger distance estimators via large deviation theory. *Entropy* **23**, 386 (2021).

[36] Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. *Optuna: A next-generation hyperparameter optimization framework. Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, 2623–2631 (2019).

[37] Bergstra, J., Bardenet, R., Bengio, Y. & Kégl, B. *Algorithms for hyper-parameter optimization. Advances in Neural Information Processing Systems*, Vol. 24, 2546–2554 (2011).

[38] Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).

[39] Javadi-Abhari, A. *et al.* Quantum computing with Qiskit. *Preprint at https://arxiv.org/abs/2405.08810* (2024).

[40] Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

[41] Meurer, A. *et al.* SymPy: symbolic computing in Python. *PeerJ Comput. Sci.* **3**, e103 (2017).

[42] Nadeau, C. & Bengio, Y. Inference for the generalization error. *Mach. Learn.* **52**, 239–281 (2003).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- saibfrssi.pdf