

Supplementary Information for “Efficient and identifiable noise models for symmetry-preserving quantum circuits”

Yi Pan^{1†}, Meng Hsiu Tsai^{2†}, Hanqi Jiang^{1†}, Junhao Chen¹,
 Weihang You¹, Yifan Zhou¹, Wei Zhang³, Isaac Lyngaas⁴,
 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.

³School of Computer and Cyber Sciences, Augusta University, Augusta, 30912, GA, USA.

⁴National 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.

Supplementary Note 1: Notation and Pauli channels

This Supplementary Note uses the notation and kernel definitions of the main-text section “The SA-IBF noise model”: $P_{\text{ideal}} \in \Delta_{2^n}$ is the ideal distribution; K_{IBF} , K_{sym} , K_{SA} are the kernels of Eqs. (1) to (3); $\hat{P}$ is the predicted distribution of Eq. (4); and $Z_x(q) := \sum_{y': H(y')=H(x)} K_{\text{IBF}}(y' | x; q)$ is the sector normalizer. We write $\text{Int}(\Delta_{2^n})$ for the open interior of the probability simplex, parameterized by $2^n - 1$ affine coordinates. The following positivity lemma is used across all proofs in this Supplementary Information.

Lemma S1 (Kernel positivity and real-analyticity) *For $q \in (0, 1/2)$, $\alpha \in [0, 1)$, $P_{\text{ideal}} \in \text{Int}(\Delta_{2^n})$, and every $x, y \in \{0, 1\}^n$:*

$$(i) \quad Z_x(q) = (1 - q)^n + \sum_{k=1}^{\min(H(x), n-H(x))} \binom{H(x)}{k} \binom{n-H(x)}{k} q^{2k} (1 - q)^{n-2k} > 0.$$

- (ii) $Z_x(q) > 0$ on $(-\infty, 1)$ (every q^{2k} factor is nonnegative, $(1-q)^{n-2k} > 0$ for $q < 1$, $k=0$ term $(1-q)^n > 0$). Hence $1/Z_x(q)$ is real-analytic on $(-\infty, 1)$; in particular on $(0, 1) \supset (0, 1/2)$. Moreover, $Z_x(q) \geq (1-q)^n \geq 9/16 > 1/2$ on $[0, 1/(2n)]$ for $n \geq 2$.
- (iii) $K_{\text{SA}}(y | x; \alpha, q) \geq (1-\alpha) K_{\text{IBF}}(y | x; q) = (1-\alpha) q^{d(x,y)} (1-q)^{n-d(x,y)} > 0$.
- (iv) $\hat{P}(y; \alpha, q, P_{\text{ideal}}) > 0$.

Proof (i) Every summand is positive on $q \in (0, 1)$, and the $k=0$ term $(1-q)^n$ is strictly positive. (ii) Z_x is a polynomial in q ; for $q < 1$, every $q^{2k} \geq 0$ (even power) and $(1-q)^{n-2k} > 0$, with the $k=0$ summand $(1-q)^n > 0$, so $Z_x(q) > 0$ on $(-\infty, 1)$. A polynomial strictly positive on an interval has a real-analytic reciprocal there, so $1/Z_x(q)$ is real-analytic on $(-\infty, 1)$. The lower bound $Z_x(q) \geq (1-1/(2n))^n \geq 9/16$ on $[0, 1/(2n)]$ follows directly. (iii) $K_{\text{sym}} \geq 0$, so $K_{\text{SA}} \geq (1-\alpha) K_{\text{IBF}}$. (iv) Each summand in $\hat{P}(y) = \sum_x P_{\text{ideal}}(x) K_{\text{SA}}(y | x)$ is positive when $P_{\text{ideal}}(x) > 0$ (interior) and $K_{\text{SA}}(y | x) > 0$. $\square$

1.1 Product Pauli correspondence

The kernel K_{IBF} equals the computational-basis output distribution of a product Pauli channel with uniform per-qubit bit-flip marginal, as follows.

Proposition S1 (Product Pauli channel $\leftrightarrow$ bit-flip kernel) (i) Independent X -channel. Let $\Lambda_p^{(X)}$ be the n -qubit product Pauli channel that applies X to each qubit independently with probability $p \in [0, 1]$ and identity otherwise. For any state ρ whose computational-basis probabilities are $P_{\text{ideal}}(x) := \langle x | \rho | x \rangle$, the computational-basis measurement distribution of $\Lambda_p^{(X)}(\rho)$ equals $y \mapsto \sum_x P_{\text{ideal}}(x) K_{\text{IBF}}(y | x; p)$.

(ii) General product Pauli channel. If the physical channel is any n -qubit product Pauli channel with per-qubit error probabilities (p_i^X, p_i^Y, p_i^Z) , then under computational-basis measurement Z_i acts as identity and the effective per-qubit bit-flip rate is $r_i := p_i^X + p_i^Y$. When all $r_i \equiv q$ uniformly, the output distribution equals $y \mapsto \sum_x P_{\text{ideal}}(x) K_{\text{IBF}}(y | x; q)$.

Proof (i) Write $\rho = \sum_{x,x'} \rho_{xx'} |x\rangle\langle x'|$; the diagonal entries are $\rho_{xx} = P_{\text{ideal}}(x)$. $\Lambda_p^{(X)}$ factorizes across qubits and acts on each single-qubit Pauli component by $\sigma \mapsto (1-p)\sigma + pX\sigma X$. On a computational-basis projector $|x\rangle\langle x| = \bigotimes_i |x_i\rangle\langle x_i|$, each factor evolves as $|x_i\rangle\langle x_i| \mapsto (1-p)|x_i\rangle\langle x_i| + p|1-x_i\rangle\langle 1-x_i|$; taking the tensor product over qubits and expanding yields $|x\rangle\langle x| \mapsto \sum_y p^{d(x,y)} (1-p)^{n-d(x,y)} |y\rangle\langle y|$. Computational-basis measurement reads out diagonal entries, so the output distribution is $\sum_x P_{\text{ideal}}(x) p^{d(x,y)} (1-p)^{n-d(x,y)} = \sum_x P_{\text{ideal}}(x) K_{\text{IBF}}(y | x; p)$.

(ii) For the general product Pauli case, $Y_i = iX_i Z_i$ acts on $|b_i\rangle$ as $Y_i |b_i\rangle = \pm i |1-b_i\rangle$, and the computational-basis projector satisfies $Y_i |b_i\rangle\langle b_i| Y_i = |1-b_i\rangle\langle 1-b_i|$ identically to $X_i |b_i\rangle\langle b_i| X_i$; the phase $\pm i$ vanishes when sandwiching. $Z_i |b_i\rangle\langle b_i| Z_i = |b_i\rangle\langle b_i|$ leaves the computational-basis projector invariant. Hence the computational-basis measurement outcome depends only on $(p_i^X + p_i^Y)$ per qubit; setting $r_i = p_i^X + p_i^Y$ and factorizing across qubits recovers case (i) with $r_i \equiv q$ (standard Pauli-channel identity; see, e.g., [1]). $\square$

The kernel K_{IBF} is therefore the computational-basis image of a product Pauli channel with bit-flip marginal q , and it enters $K_{\text{SA}} = \alpha K_{\text{sym}} + (1-\alpha) K_{\text{IBF}}$ with

weight $1 - \alpha$. The component αK_{sym} summarizes sector preservation in the output distribution, which can arise from several mechanisms, such as stabilizer constraints or physical symmetries. We therefore treat α as an effective parameter.

Supplementary Note 2: Proof of Proposition 1

Let $\mathcal{F} := \{K_r : r \in [0, 1]^n\}$ with $K_r(y | x) := \prod_{i: x_i \neq y_i} r_i \prod_{i: x_i = y_i} (1 - r_i)$. At equal rates $r = (q, \dots, q)$ the kernel K_r is $K_{\text{IBF}}(\cdot | \cdot; q)$.

Proof of Proposition 1 Each $K_r(y | x)$ depends on (x, y) only through the flip pattern $x \oplus y$, so every $K \in \mathcal{F}$ satisfies the linear constraints $K(y | x) = K(x \oplus y | 0^n)$ for all $x, y \in \{0, 1\}^n$. The map $r \mapsto K_r$ is continuous on the compact cube, so $\mathcal{F}$ is compact and its convex hull is closed in the finite-dimensional space of kernels; the linear constraints hold on this hull. For $n \geq 2$ let e_i denote the bitstring with a single one at position i . The strings e_1 and e_2 lie in the same sector at distance two, $e_1 \oplus e_2$ has weight two, and 0^n is alone in its sector. Hence $K_{\text{sym}}(e_2 | e_1) = q^2(1 - q)^{n-2}/Z_{e_1}(q)$ and $K_{\text{sym}}(e_1 \oplus e_2 | 0^n) = 0$, while $K_{\text{IBF}}(e_2 | e_1) = K_{\text{IBF}}(e_1 \oplus e_2 | 0^n) = q^2(1 - q)^{n-2}$. Therefore

$$K_{\text{SA}}(e_2 | e_1) - K_{\text{SA}}(e_1 \oplus e_2 | 0^n) = \frac{\alpha q^2(1 - q)^{n-2}}{Z_{e_1}(q)} > 0$$

for every $\alpha \in (0, 1]$ and $q \in (0, 1)$, where $Z_{e_1}(q) = (1 - q)^n + (n - 1)q^2(1 - q)^{n-2} > 0$. The kernel K_{SA} thus violates a constraint that holds on the closed convex hull of $\mathcal{F}$. $\square$

The violation is proportional to α , and at $\alpha = 0$ the kernel is K_{IBF} , a member of $\mathcal{F}$. The case $\alpha = 1$ gives the same conclusion for K_{sym} alone. At $n = 4$ and $q = 0.42$ the witness transition carries $K_{\text{sym}}(e_2 | e_1) = 0.204$.

Supplementary Note 3: Certified residual bounds

Define the differential operator $\mathcal{L}_\alpha := \partial_\alpha + \frac{q}{1-\alpha} \partial_q$ (for $\alpha < 1$).

Lemma S2 (Residual identity) *For every $x, y \in \{0, 1\}^n$, $\alpha \in [0, 1)$, and $q \in (0, 1/2)$:*

$$R_x(y; \alpha, q) := (\mathcal{L}_\alpha K_{\text{SA}})(y | x; \alpha, q) = \mathcal{O}(q^2) \quad \text{as } q \rightarrow 0.$$

Explicit leading coefficients (write $c_x := H(x)(n - H(x))$, $C := \binom{n}{2}$):

- (i) $y = x$: $R_x(x) = [C - c_x \frac{1+\alpha}{1-\alpha}]q^2 + \mathcal{O}(q^3)$.
- (ii) $d(x, y) = 1$, $H(y) \neq H(x)$: $R_x(y) = -(n - 1)q^2 + \mathcal{O}(q^3)$.
- (iii) $d(x, y) = 2$, $H(y) = H(x)$: $R_x(y) = \frac{2}{1-\alpha}q^2 + \mathcal{O}(q^3)$.
- (iv) $d(x, y) = 2$, $H(y) \neq H(x)$: $R_x(y) = q^2 + \mathcal{O}(q^3)$.
- (v) $d(x, y) \geq 3$: $R_x(y) = \mathcal{O}(q^{d(x, y)})$.

Proof Direct Taylor expansion of $K_{\text{SA}} = \alpha K_{\text{sym}} + (1 - \alpha)K_{\text{IBF}}$ at $q = 0$ through the displayed leading orders. For $d = 1, 2$, the lower-order terms cancel by construction of $\mathcal{L}_\alpha$; for $d = 0$, the order- q terms cancel, leaving the stated q^2 coefficient. For $d \geq 3$, direct evaluation of $\mathcal{L}_\alpha K_{\text{SA}}(y | x)$ gives (off-sector case): writing $\beta := 1 - \alpha$,

$$\partial_\alpha K_{\text{SA}}(y | x) = K_{\text{sym}} - K_{\text{IBF}} = -K_{\text{IBF}}(y | x) = -q^d(1 - q)^{n-d},$$

$$\begin{aligned}
\partial_q K_{\text{SA}}(y \mid x) &= \beta \partial_q K_{\text{IBF}}(y \mid x) = \beta q^{d-1} (1-q)^{n-d-1} [d(1-q) - (n-d)q], \\
R_x(y) &= -q^d (1-q)^{n-d} + \frac{q}{\beta} \beta q^{d-1} (1-q)^{n-d-1} [d(1-q) - (n-d)q] \\
&= q^d (1-q)^{n-d-1} [-(1-q) + d(1-q) - (n-d)q] \\
&= q^d (1-q)^{n-d-1} [(d-1)(1-q) - (n-d)q].
\end{aligned}$$

For $q \in (0, 1/(8n)]$, $|(d-1)(1-q) - (n-d)q| \leq (d-1) + nq \leq n + 1/8$. For $1 \leq d \leq n-1$, $(1-q)^{n-d-1} \leq 1$, so $|R_x(y)| \leq (n+1/8)q^d \leq 2nq^d$. For $d = n$ (off-sector iff $H(x) \neq n/2$, since then $y = \bar{x}$ has $H(y) \neq H(x)$), the $(n-d)q$ term vanishes and the $(1-q)^{n-d-1} = (1-q)^{-1}$ factor cancels the bracket's $(1-q)$: $R_x(y) = q^n (1-q)^{-1} \cdot (n-1)(1-q) = (n-1)q^n$, so $|R_x(y)| \leq nq^n \leq 2nq^n$. In all subcases $|R_x(y)| \leq 2nq^d$. The in-sector $d \geq 4$ case has the same form $|R_x(y)| = \mathcal{O}(q^d)$. Its finite-constant envelope $|R_x(y)| \leq 2nq^{\bar{d}}/(1-\alpha)$ for $n \in \{2, \dots, N_{\text{verified}}\}$ is the certified bound of Remark S1 family (iii), which the proof of Theorem 1(b) uses. $\square$

Lemma S3 (Finite- q remainder bound for $d = 2$ in-sector; computer-assisted) *Fix $n \in \{2, \dots, N_{\text{verified}}\}$ where N_{verified} is the range certified in Remark S1. For every $\alpha \in [0, 1-\epsilon]$ with $\epsilon \in (0, 1)$, every $x, y \in \{0, 1\}^n$ with $d(x, y) = 2$ and $H(x) = H(y)$, and every $q \in (0, 1/(8n)]$, the pointwise bounds*

$$(a) \quad |R_0(q, n, w, 2)|, |R_1(q, n, w, 2)| \leq 3n^2 q^2 \text{ on } q \in [0, 1/(2n)] \text{ and}$$

$$(b) \quad |R_i(q, n, w, 2) - 2q^2| \leq 4n q^3 \text{ on } q \in [0, 1/(8n)], i \in \{0, 1\},$$

both certified over the continuum q -interval by the interval-arithmetic verification in Remark S1, give

$$\left| R_x(y) - \frac{2q^2}{1-\alpha} \right| \leq \frac{4n q^3}{1-\alpha}.$$

$$\text{In particular, } |R_x(y)| \leq \left(\frac{2}{1-\alpha} + \frac{4n}{8n(1-\alpha)} \right) q^2 = \left(\frac{2}{1-\alpha} + \frac{1}{2(1-\alpha)} \right) q^2 \leq \frac{3}{1-\alpha} q^2.$$

Proof At $d = 2$ in-sector with $w := H(x)$, $c_x := w(n-w)$, $C := \binom{n}{2}$, $K_{\text{IBF}}(y \mid x) = q^2(1-q)^{n-2}$ and $K_{\text{sym}}(y \mid x) = K_{\text{IBF}}(y \mid x)/Z_x(q)$. Since $\partial_\alpha K_{\text{SA}}$ is independent of α and only $(q/\beta)\partial_q K_{\text{SA}}$ contributes the α -dependent mixing, we factor:

$$R_x(y) = R_0(q, n, w) + (\alpha/\beta) R_1(q, n, w),$$

where $R_0 := (K_{\text{sym}} - K_{\text{IBF}}) + q \partial_q K_{\text{IBF}}$ and $R_1 := q \partial_q K_{\text{sym}}$ are rational functions of q (with $Z_x(q)$ in the denominator of both R_0 and R_1) that depend only on (q, n, w) . Direct Taylor expansion through order q^4 , using $K_{\text{sym}} = K_{\text{IBF}}/Z_x$ and $1/Z_x = 1+nq+[n^2-C-c_x]q^2+\mathcal{O}(q^3)$, yields

$$K_{\text{IBF}} = q^2 - (n-2)q^3 + \frac{(n-2)(n-3)}{2} q^4 + \mathcal{O}(q^5),$$

$$K_{\text{sym}} = q^2 + 2q^3 + (3-c_x)q^4 + \mathcal{O}(q^5),$$

$$\partial_q K_{\text{IBF}} = 2q - 3(n-2)q^2 + 2(n-2)(n-3)q^3 + \mathcal{O}(q^4),$$

$$\partial_q K_{\text{sym}} = 2q + 6q^2 + 4(3-c_x)q^3 + \mathcal{O}(q^4),$$

giving

$$R_0(q) = 2q^2 + (6-2n)q^3 + [(3-c_x) + \frac{3}{2}(n-2)(n-3)]q^4 + \mathcal{O}(q^5),$$

$$R_1(q) = 2q^2 + 6q^3 + 4(3-c_x)q^4 + \mathcal{O}(q^5).$$

Combining with the factor $\alpha/\beta = (1-\beta)/\beta = 1/\beta - 1$:

$$R_x(y) = \frac{2q^2}{\beta} + \frac{6-2n\beta}{\beta} q^3 + \left[(3-c_x) \left(1 + \frac{4\alpha}{\beta} \right) + \frac{3}{2}(n-2)(n-3) \right] q^4 + \mathcal{O}(q^5).$$

Finite- q tail bound from the certified pointwise bounds. The rational-function form $R_x(y) = R_0 + (\alpha/\beta)R_1$ and the bound $Z_x(q) \geq 9/16$ on $q \in [0, 1/(2n)]$ (Lemma S1(ii)) reduce the tail estimate to the pointwise bounds

$$|R_0(q, n, w, 2)| \leq 3n^2q^2, \quad |R_1(q, n, w, 2)| \leq 3n^2q^2 \quad \text{on } q \in [0, 1/(2n)],$$

which are certified over the continuum for $n \in \{2, \dots, N_{\text{verified}}\}$ by the interval-arithmetic verification of Remark S1 (8,000 subintervals at 50-digit precision).

Combining $|R_0|, |R_1| \leq 3n^2q^2$ on $q \in [0, 1/(2n)]$:

$$|R_x(y)| \leq |R_0(q)| + (\alpha/\beta)|R_1(q)| \leq 3n^2q^2 \cdot (1 + \alpha/\beta) = 3n^2q^2/\beta \quad \text{on } q \in [0, 1/(2n)].$$

This gives a *global* finite- q envelope. For the Lemma's tight conclusion on $q \leq 1/(8n)$, we use the auxiliary tail-coefficient bound

$$|R_i(q) - 2q^2|/q^3 \leq 4n \quad \text{on } q \in [0, 1/(8n)], \quad i \in \{0, 1\}, \quad (*)$$

which is certified over the continuum per Remark S1 (interval-sup $\leq 0.812 \cdot 4n$ across $n \in \{2, \dots, 12\}$).

From (*): let $S_i(q) := R_i(q) - 2q^2$. Using $[q^2]R_0 = [q^2]R_1 = 2$ (computed above) and $\beta + \alpha = 1$:

$$\beta R_x(y) - 2q^2 = \beta(R_0 - 2q^2) + \alpha(R_1 - 2q^2) = \beta S_0 + \alpha S_1.$$

By the convex-combination bound $|\beta S_0 + \alpha S_1| \leq \max(|S_0|, |S_1|) \leq 4nq^3$ on $q \leq 1/(8n)$ (from (*)), which is interval-arithmetic certified per Remark S1), we obtain

$$|R_x(y) - \frac{2q^2}{\beta}| = |\beta R_x(y) - 2q^2|/\beta \leq \frac{4n}{\beta}q^3.$$

Finally, $|R_x(y)| \leq (2 + 4nq)q^2/\beta \leq (2 + 1/2)q^2/\beta \leq 3q^2/\beta$ on $q \leq 1/(8n)$. $\square$

Remark S1 (Interval-arithmetic certification of the four pointwise-bound families) We certify the four pointwise-bound families of Lemma S3 and Theorem 1(b) by interval arithmetic.

Each R_i is a rational function $P_i(q)/Z_x(q)^2$, where P_i is a polynomial in q of degree $\leq 3n$ with coefficients that are integer combinations of $\binom{w}{k}, \binom{n-w}{k}, \binom{n}{k}$. The normalizer satisfies $Z_x(q) \geq 9/16$ on $q \in [0, 1/(2n)]$ for $n \geq 2$ (Lemma S1), so the denominator Z_x^2 is bounded below by $(9/16)^2 > 0$, and each ratio P_i/Z_x^2 admits bounded interval enclosures.

To verify the bounds, we construct R_0, R_1 as explicit rational functions of q and evaluate each ratio as an interval enclosure on 8,000 subintervals partitioning the relevant q -interval at 50-digit precision. On each subinterval $[t_k, t_{k+1}]$ the enclosure upper-bounds $|\text{ratio}|$, and the supremum over all subintervals and all (n, w, d) gives a continuum upper bound. At $N_{\text{verified}} = 12$, 315/315 cases pass:

- (i) $d = 0$: $|R_0|/q^2, |R_1|/q^2 \leq 3n^2$ on $q \in [0, 1/(2n)]$; interval-sup $\leq 0.320 \cdot 3n^2$.
- (ii)a $d = 2$ in-sector global: $|R_0|/q^2, |R_1|/q^2 \leq 3n^2$ on $q \in [0, 1/(2n)]$; interval-sup $\leq 0.320 \cdot 3n^2$.
- (ii)b $d = 2$ in-sector tail (used in Lemma S3): $|R_i - 2q^2|/q^3 \leq 4n$ on $q \in [0, 1/(8n)]$, $i \in \{0, 1\}$; interval-sup $\leq 0.812 \cdot 4n$.
- (iii) $d \in \{4, 6, \dots, 2\min(w, n-w)\}$ in-sector: $|R_0|/q^d, |R_1|/q^d \leq 2n$ on $q \in [0, 1/(2n)]$; interval-sup $\leq 0.867 \cdot 2n$.

Range of the certification. Theorem 1(b) and Lemma S3 are *computer-assisted theorems* for $n \in \{2, \dots, 12\} =: \{2, \dots, N_{\text{verified}}\}$. The four supporting pointwise bounds are certified over the continuum q -interval by the interval-arithmetic verification above, at 50-digit precision with outward-rounded endpoints and exact-rational comparison against the claims. Across all 315 certified cases the largest ratio of interval supremum to claimed bound is at most 0.867 (the $d \geq 4$ in-sector family), so every bound holds with a relative slack of at least 0.133. The

slack per family is at least 0.680 (i), 0.680 (ii)a, 0.188 (ii)b, and 0.133 (iii). The experimental range of this paper, $n = n_q \in \{4, \dots, 7\}$, is covered. Larger n requires rerunning the same finite-range certification. The number of certified (w, d) cases at each n grows as $O(n^2)$, and each case requires expanding $R_0 \cdot Z_x^2$ and $R_1 \cdot Z_x^2$ (q -degree $\leq 3n$) and 8,000 interval evaluations.

Supplementary Note 4: Proofs of Theorems 1–3

Fix $\epsilon \in (0, 1)$. The open joint domain

$$\mathcal{D}_\epsilon := (0, 1 - \epsilon) \times (0, 1/2) \times \text{Int}(\Delta_{2^n}) \subset \mathbb{R}^{2+(2^n-1)}$$

is connected and open in $\mathbb{R}^{2+(2^n-1)}$ (product of nonempty connected opens). We parameterize $\text{Int}(\Delta_{2^n})$ by $2^n - 1$ affine coordinates. Define the Jacobian

$$J(\alpha, q, P_{\text{ideal}}) := [\partial_\alpha \hat{P}, \partial_q \hat{P}] \in \mathbb{R}^{2^n \times 2};$$

each column is a function on $\{0, 1\}^n$, and J is P_{ideal} -linear. We show Theorem 1 parts (a) and (b) below, then a joint-generic rank lemma used in the proof of Theorem 2.

4.1 Proof of Theorem 1(a): small- q local identifiability

Proof Except for the boundary case $n = 2$, $w = 1$ (treated separately in Step 4 below, where $w := H(x_a)$), pick any $x_a \in \{0, 1\}^n$ with $\max(w, n - w) \geq 2$, and any y_b at Hamming distance 2 off-sector from x_a (flipping two 0-bits to 1 or two 1-bits to 0 produces a bitstring at distance 2 with weight shifted by ± 2 ; the condition $\max(w, n - w) \geq 2$ guarantees such a y_b exists).

Step 1 (constructive witness). At $P_{\text{ideal}} = e_{x_a}$, with $P := P_{\text{ideal}}(x_a) = 1$, direct Taylor expansion of the kernel definitions gives

$$\begin{aligned} \partial_\alpha \hat{P}(x_a) &= nq + \mathcal{O}(q^2), & \partial_q \hat{P}(x_a) &= -n(1 - \alpha) + \mathcal{O}(q), \\ \partial_\alpha \hat{P}(y_b) &= -q^2 + \mathcal{O}(q^3), & \partial_q \hat{P}(y_b) &= 2q(1 - \alpha) + \mathcal{O}(q^2). \end{aligned}$$

Indeed, $K_{\text{sym}}(y_b | x_a; q) = 0$ and $K_{\text{IBF}}(y_b | x_a; q) = q^2 + \mathcal{O}(q^3)$ for the chosen off-sector $d = 2$ witness, while $K_{\text{sym}}(x_a | x_a; q) = 1 + \mathcal{O}(q^2)$ and $K_{\text{IBF}}(x_a | x_a; q) = 1 - nq + \mathcal{O}(q^2)$. The 2×2 Jacobian on rows $\{x_a, y_b\}$ has determinant

$$\begin{aligned} F(\alpha, q) &:= \det J_{ab}(\alpha, q, e_{x_a}) \\ &= (nq) \cdot 2q(1 - \alpha) - (-q^2) \cdot (-n(1 - \alpha)) + \mathcal{O}(q^3) \\ &= n(1 - \alpha)q^2 + \mathcal{O}(q^3). \end{aligned}$$

Step 2 (compactness and threshold via Hadamard's lemma). By the kernel definitions and Lemma S1(ii), $F(\alpha, q)$ is real-analytic in q on an open neighborhood of $[0, 1/2)$ and polynomial in α . Define $G(\alpha, q) := F(\alpha, q) - n(1 - \alpha)q^2$. The expansions in Step 1 give $G(\alpha, 0) = 0$ and $\partial_q G(\alpha, 0) = \partial_q^2 G(\alpha, 0) = 0$, with these three vanishing identities holding as polynomial identities in α . By the real-analytic Hadamard lemma, or equivalently by direct Taylor division in the q variable on an open neighborhood of $[0, 1 - \epsilon] \times [0, 1/(2n)]$, there exists $\tilde{G}$ real-analytic on that open neighborhood with

$$G(\alpha, q) = q^3 \tilde{G}(\alpha, q) \quad \text{on } [0, 1 - \epsilon] \times [0, 1/(2n)].$$

Define $M := \sup_{[0,1-\epsilon] \times [0,1/(2n)]} |\tilde{G}(\alpha, q)|$. By compactness of this closed rectangle and continuity of $\tilde{G}$, $M < \infty$ is a finite constant depending only on (n, ϵ) . Hence

$$|F(\alpha, q) - n(1 - \alpha)q^2| \leq Mq^3 \quad \text{on } [0, 1 - \epsilon] \times [0, 1/(2n)].$$

Define $q_1(n, \epsilon) := \min(1/(2n), n\epsilon/(2 \max(M, 1))) > 0$. For $(\alpha, q) \in [0, 1 - \epsilon] \times (0, q_1)$,

$$F(\alpha, q) \geq n(1 - \alpha)q^2 - Mq^3 \geq n\epsilon q^2 - Mq^3 = q^2(n\epsilon - Mq) \geq \frac{1}{2}n\epsilon q^2 > 0.$$

Step 3 (identity theorem over P_{ideal}). The 2×2 minor $M_{ab}(\alpha, q, P_{\text{ideal}})$ is P_{ideal} -polynomial of total degree ≤ 2 (each entry of J is P_{ideal} -linear), so it extends continuously to the closed simplex Δ_{2^n} . Step 2 gives $M_{ab}(\alpha, q, e_{x_a}) = F(\alpha, q) > 0$ at the vertex $e_{x_a} \in \partial\Delta_{2^n}$ (the point mass at x_a); by continuity, $M_{ab}(\alpha, q, P_0) > 0$ for every $P_0 \in \text{Int}(\Delta_{2^n})$ in a total-variation neighborhood of e_{x_a} . Hence $M_{ab}(\alpha, q, \cdot)$ is not identically zero on $\text{Int}(\Delta_{2^n})$. Let $\mathcal{V}(\alpha, q)$ be the rank-deficient set in the simplex interior. Then

$$\mathcal{V}(\alpha, q) = \bigcap_{r < s} \{P_{\text{ideal}} : M_{rs}(\alpha, q, P_{\text{ideal}}) = 0\},$$

where the intersection ranges over all two-row 2×2 minors. This finite common zero set is a real-analytic subvariety, and it is proper because it is contained in the proper zero set of the nontrivial minor M_{ab} . On its open dense complement, J has full column rank 2, so the analytic extension of $(\alpha, q) \mapsto \hat{P}$ is an immersion into $\mathbb{R}^{2^n}$; by the constant-rank theorem, it is locally injective, and at $\alpha = 0$ this gives local injectivity relative to the stated parameter domain.

Step 4 (boundary case $n = 2, w = 1$). For $x_a \in \{01, 10\}$, pick y_b at $d = 1$ off-sector from x_a (this gives a weight-0 or weight-2 string). Using $C = \binom{2}{2} = 1$, $c_x = w(n - w) = 1$, and direct Taylor expansion at $P_{\text{ideal}} = e_{x_a}$ (carried out explicitly), one finds

$$\det J_{ab}(\alpha, q, e_{x_a}) = -[C(1 - \alpha) + c_x(1 + \alpha)]q^2 + \mathcal{O}(q^3) = -2q^2 + \mathcal{O}(q^3),$$

uniformly in $\alpha \in [0, 1 - \epsilon]$. The compactness argument of Step 2 applied to $F'(\alpha, q) := \det J_{ab}(\alpha, q, e_{x_a}) + 2q^2$ gives a threshold $q'_1(n, \epsilon) > 0$ such that $|\det J_{ab}| \geq q^2$ on $[0, 1 - \epsilon] \times (0, q'_1)$. Step 3 then applies with the replacement $q_1 \leftarrow \min(q_1, q'_1)$.

Step 5 (uniform distribution is in $\mathcal{V}(\alpha, q)$). $P_{\text{ideal}} = 2^{-n}\mathbf{1}$ is fixed by both K_{IBF} and K_{sym} : $\sum_x 2^{-n} K_{\text{IBF}}(y | x; q) = 2^{-n} \sum_{d=0}^n \binom{n}{d} q^d (1 - q)^{n-d} = 2^{-n}$, and similarly for K_{sym} by sector-wise doubly stochasticity. Hence $\hat{P} \equiv 2^{-n}\mathbf{1}$ independent of (α, q) , and $J \equiv 0$: rank $0 < 2$. So $2^{-n}\mathbf{1} \in \mathcal{V}(\alpha, q)$ for every (α, q) . $\square$

4.2 Proof of Theorem 1(b): Fisher upper bound and Cramér–Rao divergence

Proof Write $u(y) := \partial_q \hat{P}(y)$, $v(y) := \partial_\alpha \hat{P}(y)$. The residual identity of Lemma S2 gives $v = cu + R$ with $c = -q/(1 - \alpha)$, where $R(y) := (\mathcal{L}_\alpha \hat{P})(y) = \sum_x P_{\text{ideal}}(x) R_x(y)$ by linearity of $\mathcal{L}_\alpha$. Under N -shot multinomial observations of $\hat{P}$, the Fisher matrix entries are $I_{jk} = N \sum_y (\partial_j \hat{P})(\partial_k \hat{P}) / \hat{P}$. Define the $\chi^2(\hat{P})$ -inner product

$$\langle f, g \rangle := N \sum_y \frac{f(y)g(y)}{\hat{P}(y)}.$$

Schur-complement bound (projection form). The marginal Fisher information $I_{\alpha|q}$ equals the squared $\chi^2(\hat{P})$ -distance from v to $\text{span}(u)$:

$$I_{\alpha|q} := \inf_{c' \in \mathbb{R}} \|v - c'u\|^2 = \|v\|^2 - \frac{\langle u, v \rangle^2}{\langle u, u \rangle} \quad (\text{when } \langle u, u \rangle > 0),$$

with the convention $I_{\alpha|q} := \|v\|^2$ when $\langle u, u \rangle = 0$ (i.e., $u \equiv 0$ in $L^2(\hat{P})$, so $\text{span}(u) = \{0\}$ and $I_{qq} = 0$). Using $v = cu + R$ with $c = -q/(1 - \alpha)$:

- If $\langle u, u \rangle > 0$: expanding, $\langle v, v \rangle = c^2 \langle u, u \rangle + 2c \langle u, R \rangle + \langle R, R \rangle$ and $\langle u, v \rangle = c \langle u, u \rangle + \langle u, R \rangle$, so the Schur subtraction telescopes:

$$I_{\alpha|q} = \langle R, R \rangle - \frac{\langle u, R \rangle^2}{\langle u, u \rangle} \leq \langle R, R \rangle$$

(Cauchy–Schwarz: the subtracted term is nonnegative).

- If $\langle u, u \rangle = 0$: $u \equiv 0$ in $L^2(\hat{P})$, so $v = R$ and $I_{\alpha|q} = \|v\|^2 = \langle R, R \rangle$.

In both cases, $I_{\alpha|q} \leq \langle R, R \rangle$.

Mixture inequality. For each y , apply the Cauchy–Schwarz inequality

$$\left(\sum_x a_x b_x \right)^2 \leq \left(\sum_x a_x^2 / w_x \right) \left(\sum_x b_x^2 w_x \right)$$

with $a_x = P_{\text{ideal}}(x) R_x(y)$, $b_x = 1$, and strictly positive weights $w_x := P_{\text{ideal}}(x) K_{\text{SA}}(y | x) > 0$ (Lemma S1; the symbol $c_x = H(x)(n - H(x))$ is reserved for the Hamming combinatorial factor of Lemma S2):

$$\begin{aligned} \left(\sum_x P_{\text{ideal}}(x) R_x(y) \right)^2 &\leq \left(\sum_x \frac{P_{\text{ideal}}(x)^2 R_x(y)^2}{P_{\text{ideal}}(x) K_{\text{SA}}(y | x)} \right) \left(\sum_x P_{\text{ideal}}(x) K_{\text{SA}}(y | x) \right) \\ &= \hat{P}(y) \sum_x P_{\text{ideal}}(x) \frac{R_x(y)^2}{K_{\text{SA}}(y | x)}. \end{aligned}$$

Dividing by $\hat{P}(y)$:

$$\frac{R(y)^2}{\hat{P}(y)} \leq \sum_x P_{\text{ideal}}(x) \frac{R_x(y)^2}{K_{\text{SA}}(y | x)}.$$

Per-source bound. Define $F(x; n, \alpha, q) := \sum_y R_x(y)^2 / K_{\text{SA}}(y | x)$. Using $K_{\text{SA}}(y | x) \geq (1 - \alpha)q^d(1 - q)^{n-d} \geq (1 - \alpha)q^d/2$ for $q \leq 1/(8n)$ (with $(1 - q)^{n-d} \geq (1 - 1/(8n))^n \geq 1/2$ for $n \geq 2$), and the finite- q bounds summarized below, we partition $\sum_y$ by $d = d(x, y)$.

Finite- q residual bounds for each Hamming distance. For $\alpha \in [0, 1 - \epsilon]$ and $q \in (0, 1/(8n)]$, setting $\beta := 1 - \alpha$, write $R_x(y) = R_0(q, n, w, d) + (\alpha/\beta)R_1(q, n, w, d)$ with $R_0 = (K_{\text{sym}} - K_{\text{IBF}}) + q\partial_q K_{\text{IBF}}$ and $R_1 = q\partial_q K_{\text{sym}}$ (both rational in q and independent of α):

- (i) $d = 0$: $R_0(x|x) = (1 - q)^n(1 - Z_x)/Z_x - nq(1 - q)^{n-1}$, $R_1(x|x) = -nq(1 - q)^{n-1}/Z_x - q(1 - q)^n \partial_q Z_x / Z_x^2$. Leading Taylor coefficients (direct expansion via $K_{\text{sym}}(x|x) = (1 - q)^n/Z_x = 1 - c_x q^2 + \mathcal{O}(q^3)$ and $K_{\text{IBF}}(x|x) = (1 - q)^n = 1 - nq + Cq^2 + \mathcal{O}(q^3)$): $[q^2]R_0 = C - c_x$, $[q^2]R_1 = -2c_x$; combined, $[q^2]R_x(x) = (C - c_x) - (2\alpha/\beta)c_x = C - c_x(1 + \alpha)/\beta$. Using $Z_x \geq 9/16$ on $q \leq 1/(2n)$ (Lemma S1) for the rational-function denominator, the interval-arithmetic continuum certification of Remark S1 family (i) gives $|R_0|, |R_1| \leq 3n^2 q^2$ on $q \leq 1/(2n)$ for $n \in \{2, \dots, N_{\text{verified}}\}$. Hence $|R_x(x)| \leq |R_0| + (\alpha/\beta)|R_1| \leq 3n^2 q^2(1 + \alpha/\beta) = \frac{3n^2}{\beta} q^2$ (using $\alpha + \beta = 1$), matching the $d = 0$ entry of the display below.
- (ii) $d = 1$ off-sector: $K_{\text{sym}}(y | x) = 0$, so $R_0 = -K_{\text{IBF}} + q\partial_q K_{\text{IBF}}$ and $R_1 = 0$. Substituting $K_{\text{IBF}} = q(1 - q)^{n-1}$ and $\partial_q K_{\text{IBF}} = (1 - q)^{n-1} - (n - 1)q(1 - q)^{n-2}$: $R_x(y) = R_0(q) = -q(1 - q)^{n-1} + q(1 - q)^{n-1} - (n - 1)q^2(1 - q)^{n-2} = -(n - 1)q^2(1 - q)^{n-2}$. Hence $|R_x(y)| \leq (n - 1)q^2 \leq nq^2$ for all $q \in [0, 1]$, *no $1/\beta$ factor*.
- (iii) $d = 2$ in-sector: $|R_x(y)| \leq 3q^2/\beta$ by Lemma S3.
- (iv) $d = 2$ off-sector: $K_{\text{sym}} = 0$, $R_1 = 0$, $R_0 = -K_{\text{IBF}} + q\partial_q K_{\text{IBF}}$. For $K_{\text{IBF}} = q^2(1 - q)^{n-2}$ and $\partial_q K_{\text{IBF}} = 2q(1 - q)^{n-2} - (n - 2)q^2(1 - q)^{n-3}$: $R_x(y) = -q^2(1 - q)^{n-2} + 2q^2(1 - q)^{n-2} - (n - 2)q^3(1 - q)^{n-3} = q^2(1 - q)^{n-3}[1 - (n - 1)q]$. On $q \leq 1/(8n)$, $(1 - q)^{n-3} \leq 1$ for $n \geq 3$ and $|1 - (n - 1)q| \leq 1$, so $|R_x(y)| \leq q^2$, *no $1/\beta$ factor*. For $n = 2$ the formula simplifies directly: $q^2(1 - q)^{-1} \cdot [1 - (n - 1)q]$ with $(n - 1)q = q$ gives $R_x(y) = q^2(1 - q)^{-1}(1 - q) = q^2$.

- (v) $d \geq 3$ off-sector: $K_{\text{sym}} = 0$, $R_1 = 0$. Same computation: $R_x(y) = q^d(1-q)^{n-d-1}[(d-1)(1-q) - (n-d)q]$. For $3 \leq d < n$, $(1-q)^{n-d-1} \leq 1$ on $q \in [0, 1]$ and on $q \leq 1/(8n)$, $|(d-1)(1-q) - (n-d)q| \leq (d-1) + n/(8n) = d-1 + 1/8 \leq n$, so $|R_x(y)| \leq n \cdot q^d$. For $d = n$ (which occurs when $H(x) \neq n/2$, so that $y = \bar{x}$ is off-sector), the $(n-d)q$ term vanishes and the $(1-q)^{n-d-1} = (1-q)^{-1}$ factor cancels the bracket's $(1-q)$: $R_x(y) = q^n(1-q)^{-1} \cdot (n-1)(1-q) = (n-1)q^n$, giving $|R_x(y)| \leq n \cdot q^n$ directly. In both subcases $|R_x(y)| \leq n \cdot q^d$, with no $1/\beta$ factor.
- (vi) $d \geq 4$ in-sector (even d , requires $2 \min(w, n-w) \geq d$): $K_{\text{sym}} \neq 0$ and $R_1 \neq 0$. The bound $|R_0|, |R_1| \leq 2nq^d$ on $q \in [0, 1/(2n)]$ is certified over the continuum per Remark S1 via interval arithmetic on $d \in \{4, 6, \dots, 2 \min(w, n-w)\}$ with interval-sup $\leq 0.867 \cdot 2n$ across $n \in \{2, \dots, N_{\text{verified}}\}$. The interval-arithmetic certificate supplies the uniform $2n$ coefficient, including the in-sector saturation boundary $d = 2 \min(w, n-w) = n$ reached for even n with $w = n/2$ (e.g., $(n, w, d) = (4, 2, 4)$ and $(6, 3, 6)$). Combining, $|R_x(y)| \leq 2n \cdot q^d \cdot (1 + \alpha/\beta) \leq 2n \cdot q^d/\beta$.

$$\begin{aligned}
d = 0: & \quad 1 \text{ term, } |R_x|^2 \leq \frac{9n^4}{(1-\alpha)^2} q^4, \quad K_{\text{SA}} \geq (1-\alpha)/2: \quad \frac{18n^4 q^4}{(1-\alpha)^3} \leq \frac{9n^2 q^2}{32(1-\alpha)^3}. \\
d = 1 \text{ off:} & \quad n \text{ terms, } |R_x|^2 \leq n^2 q^4, \quad K_{\text{SA}} \geq (1-\alpha)q/2: \quad \text{total} \leq \frac{2n^3 q^3}{1-\alpha} \leq \frac{n^2 q^2}{4(1-\alpha)}. \\
d = 2 \text{ in:} & \quad \leq \frac{n^2}{4} \text{ terms, } |R_x|^2 \leq \frac{9}{(1-\alpha)^2} q^4, \quad K_{\text{SA}} \geq (1-\alpha)q^2/2: \quad \text{total} \leq \frac{9n^2 q^2}{2(1-\alpha)^3}. \\
d = 2 \text{ off:} & \quad \leq \frac{n^2}{2} \text{ terms, } |R_x|^2 \leq q^4, \quad K_{\text{SA}} \geq (1-\alpha)q^2/2: \quad \text{total} \leq \frac{n^2 q^2}{1-\alpha}. \\
d \geq 3: & \quad |R_x|^2 \leq \frac{4n^2}{(1-\alpha)^2} q^{2d}, \quad K_{\text{SA}} \geq (1-\alpha)q^d/2; \quad \text{total} \leq \frac{8n^2}{(1-\alpha)^3} \sum_{d \geq 3} \binom{n}{d} q^d \leq \frac{n^4 q^2}{(1-\alpha)^3}.
\end{aligned}$$

(The $d \geq 3$ tail uses $\sum_{d \geq 3} \binom{n}{d} q^d \leq \frac{(nq)^3}{3!} e^{nq} \leq \frac{(nq)^3}{5}$ on $q \leq 1/(8n)$, via $\binom{n}{d} \leq n^d/d!$, the Taylor tail $\sum_{d \geq 3} (nq)^d/d! \leq e^{nq}(nq)^3/6$, and $e^{1/8} < 6/5$; then $\frac{8n^2}{(1-\alpha)^3} \cdot \frac{(nq)^3}{5} = \frac{8n^5 q^3}{5(1-\alpha)^3} \leq \frac{n^4 q^2}{(1-\alpha)^3}$ since $8nq/5 \leq 1$ on $q \leq 1/(8n)$.) Summing (using $1/(1-\alpha) \leq 1/(1-\alpha)^5$ for $\alpha \in [0, 1-\epsilon]$):

$$F(x; n, \alpha, q) \leq \frac{q^2}{(1-\alpha)^5} \left[\frac{9n^2}{32} + \frac{n^2}{4} + \frac{9n^2}{2} + n^2 + n^4 \right] \leq \frac{10n^4 q^2}{(1-\alpha)^5}.$$

(The last step uses $9n^2/32 + n^2/4 + 9n^2/2 + n^2 = \frac{193}{32}n^2 < 7n^2$ and $7n^2 + n^4 \leq 10n^4$ for $n \geq 2$.)

Conclusion. Summing over y and using $\sum_x P_{\text{ideal}}(x) = 1$:

$$\langle R, R \rangle = N \sum_y \frac{R(y)^2}{\hat{P}(y)} \leq N \sum_x P_{\text{ideal}}(x) F(x) \leq \sup_x F(x) \cdot N \leq \frac{10n^4}{(1-\alpha)^5} N q^2.$$

By the projection-form Schur bound (covering both $I_{qq} > 0$ and $I_{qq} = 0$), $I_{\alpha|q} \leq \langle R, R \rangle \leq c_+(n, \alpha) N q^2$ with $c_+ = 10n^4/(1-\alpha)^5$. If $I_{\alpha|q} > 0$, the Cramér–Rao inequality gives $\text{Var}(\hat{\alpha}) \geq 1/I_{\alpha|q} \geq 1/(c_+ N q^2)$ for regular unbiased estimators at two-sided interior points. If $I_{\alpha|q} = 0$, the efficient information for α after accounting for q is zero and the local Cramér–Rao lower bound is $+\infty$. $\square$

4.3 Joint-generic rank lemma

Lemma S4 (Joint-generic rank) *Under Assumption 1, the rank-deficient set $\mathcal{E}_\epsilon := \{(\alpha, q, P_{\text{ideal}}) \in \mathcal{D}_\epsilon : \text{rank } J < 2\}$ equals the finite intersection $\bigcap_{(a,b)} \{\det J_{ab} = 0\}$ over*

all 2×2 minors of J , is contained in the zero-locus of any single nontrivial minor, and is Lebesgue-null in the affine $(\alpha, q, P_{\text{ideal}})$ coordinates on $\mathcal{D}_\epsilon$ [2, 3].

Proof The proof has four steps: joint real-analyticity of a minor, non-vanishing at a witness point, Krantz–Parks identity theorem, and containment of the rank-deficient set.

Step A'.1 (joint real-analyticity). On $\mathcal{D}_\epsilon$, each entry of $J(\alpha, q, P_{\text{ideal}})$ is P_{ideal} -linear with coefficients that are finite sums of products of α , $(1 - \alpha)$, $q^d(1 - q)^{n-d}$, $1/Z_x(q)$, and their derivatives in α, q . By Lemma S1(ii), $1/Z_x(q)$ is real-analytic on $(0, 1) \supset (0, 1/2)$. Products, sums, and P_{ideal} -linear combinations of real-analytic functions are real-analytic on the connected open set $\mathcal{D}_\epsilon$. Hence each 2×2 minor $\det J_{ab}$ is real-analytic on $\mathcal{D}_\epsilon$.

Step A'.2 (non-vanishing witness). Part (a) Step 1 provides $(\alpha_\circ, q_\circ) \in (0, 1 - \epsilon) \times (0, q_1)$ with $\det J_{a_0 b_0}(\alpha_\circ, q_\circ, e_{x_a}) > 0$ for some specific minor indices (a_0, b_0) . Since $\det J_{a_0 b_0}$ is P_{ideal} -polynomial of total degree ≤ 2 , it extends continuously to the closed simplex. For any $P_\circ \in \text{Int}(\Delta_{2^n})$ sufficiently close to e_{x_a} (in total variation), $\det J_{a_0 b_0}(\alpha_\circ, q_\circ, P_\circ) > 0$ by continuity. Hence $\det J_{a_0 b_0}$ is not identically zero on $\mathcal{D}_\epsilon$.

Step A'.3 (real-analytic identity theorem). By the real-analytic identity theorem [2], if $f : \Omega \rightarrow \mathbb{R}$ is real-analytic on a connected open subset $\Omega \subset \mathbb{R}^m$ and not identically zero, then $\{x \in \Omega : f(x) = 0\}$ is a proper real-analytic subvariety of Ω , hence Lebesgue-null in $\mathbb{R}^m$ under the standard Lebesgue-measure conventions [3].

Applied to $\det J_{a_0 b_0}$ on $\mathcal{D}_\epsilon \subset \mathbb{R}^{2+(2^n-1)}$: the zero set $\mathcal{E}_{a_0 b_0}$ is a proper real-analytic subvariety, Lebesgue-null.

Step A'.4 (containment). The rank-deficient set $\mathcal{E}_\epsilon = \{(\alpha, q, P_{\text{ideal}}) \in \mathcal{D}_\epsilon : \text{rank } J < 2\}$ equals the intersection $\bigcap_{(a,b)} \mathcal{E}_{ab}$ over all 2×2 minors. In particular, $\mathcal{E}_\epsilon \subset \mathcal{E}_{a_0 b_0}$. Since $\mathcal{E}_{a_0 b_0}$ is Lebesgue-null, so is $\mathcal{E}_\epsilon$. $\square$

Remark S2 (Fixed-parameter slices) Lemma S4 gives a joint measure-zero exceptional set. By Fubini [3], for Lebesgue-almost-every $(\alpha_\star, q_\star) \in (0, 1 - \epsilon) \times (0, 1/2)$ the slice $\{P_{\text{ideal}} : (\alpha_\star, q_\star, P_{\text{ideal}}) \in \mathcal{E}_\epsilon\}$ is Lebesgue-null in $\text{Int}(\Delta_{2^n})$, and the rank diagnostic of the main text checks the rank at a given fit.

4.4 Proof of Theorem 2

Fix a nonempty support $T \subseteq \{0, 1\}^n$ and write

$$\begin{aligned} \Delta_T &:= \{P \in \Delta_{2^n} : P(x) = 0 \text{ for } x \notin T\}, \\ \text{relint } \Delta_T &:= \{P \in \Delta_T : P(x) > 0 \text{ for } x \in T\}. \end{aligned}$$

The proof is the same analytic-minor argument as Supplementary Notes 4.1 and 4.3, read in the relative affine coordinates of the support face.

Proof Pick any $x_a \in T$ and set $w = H(x_a)$. If $\max(w, n - w) \geq 2$, choose y_b by flipping two zero bits of x_a to ones or two one bits to zeros, so that $d(x_a, y_b) = 2$ and $H(y_b) \neq H(x_a)$. In the exceptional case $n = 2, w = 1$, choose a distance-one off-sector string. These choices are always possible for $n \geq 2$.

At the face vertex $P = e_{x_a}$, the same two-row Jacobian minor used in Supplementary Note 4.1 is nonzero for small q , uniformly over $\alpha \in [0, 1 - \epsilon]$: in the distance-two case,

$$\det J_{\{x_a, y_b\}}(\alpha, q, e_{x_a}) = n(1 - \alpha)q^2 + \mathcal{O}(q^3),$$

and in the $n = 2, w = 1$ case the determinant is $-2q^2 + \mathcal{O}(q^3)$. Compactness of $\alpha \in [0, 1 - \epsilon]$ gives a threshold $q_x(n, \epsilon) > 0$ for this chosen source string. Taking the minimum over the finite set of possible source strings gives $q_{\text{face}}(n, \epsilon) > 0$ independent of the particular support T .

Now fix $(\alpha, q) \in [0, 1 - \epsilon] \times (0, q_{\text{face}})$. Each entry of $J(\alpha, q, P)$ is linear in P , so every 2×2 minor is a polynomial of degree at most two in the relative affine coordinates of Δ_T . The selected minor is nonzero at the closure point $e_{x_a} \in \Delta_T$, so it is not the zero polynomial on the affine hull of Δ_T . Its zero set inside $\text{relint } \Delta_T$ is therefore either empty or a proper real-analytic subvariety. The rank-deficient set equals the finite intersection of the zero sets of all two-row 2×2 minors; hence it is a real-analytic subvariety, and it is proper because it is contained in the selected proper zero set. On the relative open dense complement, J has full column rank, so the analytic extension of $(\alpha, q) \mapsto \hat{P}$ is locally injective by the constant-rank theorem; at $\alpha = 0$ this conclusion is relative to the stated parameter domain.

For the full- q joint statement, use the same minor as a real-analytic function on the connected relative domain

$$(0, 1 - \epsilon) \times (0, 1/2) \times \text{relint } \Delta_T.$$

This minor is not identically zero on this relative domain: for $|T| = 1$, the witness e_{x_a} lies in $\text{relint } \Delta_T$; for $|T| > 1$, if the selected minor vanished on $\text{relint } \Delta_T$ at the small- q witness (α, q) above, its polynomial extension in the face coordinates would vanish on the closure Δ_T , contradicting its nonzero value at e_{x_a} . Thus some point of the relative domain has a nonzero value of the selected minor. The real-analytic identity theorem therefore makes its zero set relative-Lebesgue-null, and the joint rank-deficient set is contained in that null set. $\square$

Remark S3 (Relation to Theorem 3) Theorem 2 assumes a fixed known support T and makes all genericity statements in the relative face coordinates. Hardware fits at high $\hat{q}$ or on sparse supports are covered by Theorem 3 through the diagnostic of Definition 1.

4.5 Proof of Theorem 3

Proof Fix $P_0 \in \text{relint } \Delta_T$ and define

$$\Phi_{P_0}(\alpha, q) := \hat{P}(\cdot; \alpha, q, P_0) \in \mathbb{R}^{2^n}.$$

For $q \in (0, 1/2)$, every denominator $Z_x(q)$ in K_{sym} is strictly positive because the term $y = x$ contributes $(1 - q)^n > 0$. Since $\alpha_0 \in (0, 1)$ and $q_0 \in (0, 1/2)$, there is an open rectangle around (α_0, q_0) contained in $(0, 1) \times (0, 1/2)$; on this rectangle Φ_{P_0} is real-analytic in (α, q) .

Since $J(\alpha_0, q_0, P_0) = D\Phi_{P_0}(\alpha_0, q_0)$ has column rank 2, there exist two output coordinates $a \neq b$ such that the 2×2 minor $M_{ab}(\alpha_0, q_0, P_0)$ is nonzero. Let $\pi_{ab} : \mathbb{R}^{2^n} \rightarrow \mathbb{R}^2$ project onto those two coordinates and set $G_{ab} := \pi_{ab} \circ \Phi_{P_0}$. The Jacobian of G_{ab} at (α_0, q_0) has determinant $M_{ab}(\alpha_0, q_0, P_0) \neq 0$. By the inverse function theorem, G_{ab} is one-to-one on some open neighborhood U of (α_0, q_0) . If two points in U have the same full prediction Φ_{P_0} , then their two selected coordinates under G_{ab} are also equal; the injectivity of G_{ab} on U forces the two parameter pairs to be identical. Therefore Φ_{P_0} is locally injective at (α_0, q_0) . $\square$

Supplementary Note 5: Rank diagnostic

This note reports the diagnostic of Definition 1 at the seven representative operating points and for the fit of every circuit in the main data set, and compares its standard errors with the spread of refitted parameters.

5.1 Representative configurations

Supplementary Table 1 reports the rank diagnostic of the main text for each configuration, with the singular values and condition number of J and both Cramér–Rao standard errors at each representative operating point, for 30,000 shots. Every row passes the diagnostic.

Supplementary Table 1 Rank diagnostic of Definition 1 at the seven representative operating points, each pairing the median fit of a configuration with the ideal distribution of its first circuit.

Family	n_q	$\hat{\alpha}$	$\hat{q}$	$\sigma_{\min}(J)$	$\sigma_{\max}(J)$	$\kappa(J)$	seCRLB $_{\alpha}$	seCRLB $_q$
UCCSD	4	0.059	0.225	2.33×10^{-1}	1.96	8.4	0.0062	0.0014
Magnetization	4	0.351	0.070	3.05×10^{-2}	1.58	52	0.0182	0.0021
Magnetization	6	0.262	0.045	1.86×10^{-2}	2.45	1.3×10^2	0.0201	0.0013
MolQAE	4	0.134	0.462	5.59×10^{-1}	1.14	2.0	0.0028	0.0019
MolQAE	5	0.016	0.499	6.41×10^{-1}	1.05	1.6	0.0014	0.0014
MolQAE	6	0.013	0.479	5.39×10^{-1}	1.00	1.9	0.0011	0.0013
MolQAE	7	0.077	0.463	4.07×10^{-1}	0.98	2.4	0.0017	0.0012

5.2 Fits to individual circuits

We apply the diagnostic of Supplementary Note 5.1 to the fits of all 230 circuits in the main data set (Table 1). For each fit $(\hat{\alpha}_i, \hat{q}_i, P_{\text{ideal}}^{(i)})$ we evaluate the analytic Jacobian $J_i = [\partial_{\alpha} \hat{P}, \partial_q \hat{P}]$ at the circuit’s own operating point and the multinomial Fisher information $I_i = J_i^{\top} \text{diag}(N/\hat{p}_i) J_i$ for $N=30,000$ shots, and we apply the thresholds of Definition 1. The diagnostic *passes* if and only if (i) $\min_y \hat{p}_i(y) > 0$, (ii) I_i is nonsingular, (iii) $\kappa(J_i) < 10^3$, and (iv) $\max(\text{seCRLB}_{\alpha,i}, \text{seCRLB}_{q,i}) < 0.1$. The fits also satisfy $\hat{\alpha}_i \in [0.0070, 0.596]$ and $\hat{q}_i \in [0.0293, 0.499]$, so they lie in the two-sided parameter interior required by Theorem 3. Aggregate counts and conditioning statistics are in Supplementary Table 2.

Because I_i is linear in N , the 15,000-shot training histograms of the main data set multiply each seCRLB by $\sqrt{2}$ and leave $\kappa(J_i)$ unchanged. With these histograms all 230 fits also pass, with $\max(\text{seCRLB})$ at most 0.057 (Magnetization $n_q=4$) and 0.051 (Magnetization $n_q=6$), and the representative configurations of Supplementary Table 1 pass with $\text{seCRLB}_{\alpha} \leq 0.029$. Each fit therefore satisfies the rank condition of Theorem 3, with Fisher conditioning inside the stated standard-error tolerance.

Supplementary Table 2 Rank diagnostic for each circuit of the main data set, evaluated at the circuit’s own fit. “pass” counts circuits that pass Definition 1. $\kappa(J)$ is summarized by the median, interquartile range (IQR), and maximum over circuits, and $\max(\text{seCRLB})$ is the larger of the two standard errors of each circuit. The bold row gives the total over the seven configurations.

Family	n_q	N_{circ}	pass / N_{circ}	$\kappa(J)$			$\max(\text{seCRLB})$	
				median	IQR	max	median	max
UCCSD	4	30	30/30	8.4	[7.7, 8.8]	9.9	0.006	0.007
Magnetization	4	60	60/60	54	[37, 88]	272	0.025	0.040
Magnetization	6	40	40/40	134	[77, 203]	445	0.023	0.036
MolQAE	4	20	20/20	2.0	[2.0, 2.2]	2.5	0.003	0.003
MolQAE	5	20	20/20	1.6	[1.6, 1.6]	1.7	0.0015	0.0015
MolQAE	6	20	20/20	1.9	[1.9, 1.9]	1.9	0.0013	0.0013
MolQAE	7	40	40/40	2.4	[2.4, 2.5]	2.6	0.002	0.002
Total	—	230	230/230	—	—	—	—	—

5.3 Parameter recovery

For ten circuits of each configuration, we draw 100 multinomial samples of 15,000 shots from the fitted SA-IBF prediction and 100 from the measured 30,000-shot histogram, refit SA-IBF to each sample, and compare the spread of the refitted parameters with the Cramér–Rao standard errors of Definition 1 for 15,000 shots (Supplementary Table 3). In every configuration, the standard deviation of $\hat{\alpha}$ under both kinds of resampling is close to seCRLB_α , and the median bias of $\hat{\alpha}$ is smaller than its standard deviation. The refitted predictions differ from the original fit by a median Hellinger distance below 0.004. On MolQAE $n_q=5$, most resamples keep $\hat{q}$ at its upper bound, which narrows the spread of $\hat{q}$.

Supplementary Table 3 Parameter recovery for ten circuits of each configuration of the main data set. SD is the standard deviation of a refitted parameter over 100 resamples of 15,000 shots, drawn from the fitted SA-IBF prediction (model) or from the measured histogram (hardware), and seCRLB is the Cramér–Rao standard error for 15,000 shots. |bias| is the absolute difference between the mean refitted $\hat{\alpha}$ and the fitted value under model resampling. All entries are medians over circuits.

Family	n_q	$\hat{\alpha}$	$\hat{\alpha}$				$\hat{q}$	
			SD, model	SD, hardware	seCRLB_α	bias	SD, hardware	seCRLB_q
UCCSD	4	0.069	0.0083	0.0091	0.0087	0.0007	0.0022	0.0020
Magnetization	4	0.311	0.0351	0.0297	0.0367	0.0065	0.0028	0.0032
Magnetization	6	0.288	0.0328	0.0287	0.0343	0.0170	0.0020	0.0021
MolQAE	4	0.125	0.0041	0.0042	0.0041	0.0003	0.0028	0.0027
MolQAE	5	0.016	0.0018	0.0018	0.0019	0.0002	0.0003	0.0021
MolQAE	6	0.013	0.0015	0.0016	0.0016	0.0001	0.0019	0.0018
MolQAE	7	0.059	0.0021	0.0022	0.0022	0.0001	0.0021	0.0017

Supplementary Note 6: Baselines and robustness

This note defines the baseline noise models of Table 1 and reports robustness checks of the SA-IBF fit: a simulation with readout error only, alternative within-sector components, and the search bound on q .

6.1 Baseline noise models

The baselines of Table 1 are the following Markov kernels. Each is fitted by the Hellinger objective of the main text on the same training and test split as SA-IBF.

Definition S1 (Baseline noise models) For θ in the natural range of each kernel, define:

- (i) *Depolarization toward uniform* (one-parameter): $f_{\theta}^{(1)}(y | x) := (1 - \theta)\mathbf{1}[y = x] + \theta 2^{-n}$, $\theta \in [0, 1]$.
- (ii) *Concentration shrinkage toward a P_{ideal} -peak* (one-parameter): $f_{\theta}^{(2)}(y | x) := (1 - \theta)\mathbf{1}[y = x] + \theta \mathbf{1}[y = x_P^*]$, where $x_P^* := \arg \max_x P_{\text{ideal}}(x)$, $\theta \in [0, 1]$.
- (iii) *Hamming-weight shrinkage toward a geometric envelope* (two-parameter): $f_{F,\lambda}^{(3)}(y | x) := (1 - F)\mathbf{1}[y = x] + F g_{\lambda}(y)$, where $g_{\lambda}(y) \propto e^{-\lambda H(y)}$ normalizes over $y \in \{0, 1\}^n$, $F \in [0, 1]$, $\lambda \geq 0$.
- (iv) *Independent bit-flip* (one-parameter): $f_{\theta}^{(4)}(y | x) := K_{\text{IBF}}(y | x; \theta)$, $\theta = q \in (0, 1/2)$.

For each family, let $\theta_{\star}^{(i)}(P_{\text{ideal}}, P_{\text{obs}}) \in \arg \min_{\theta} D_{\text{H}}(\sum_x P_{\text{ideal}}(x) f_{\theta}^{(i)}(\cdot | x), P_{\text{obs}})$ denote the Hellinger-optimal fit of its parameter vector on the training histogram P_{obs} . D_{H} is then evaluated on the test histogram, as for SA-IBF. Families (i), (ii), (iv) have a single scalar parameter; family (iii) has two.

6.2 Simulation with readout error only

Supplementary Table 4 reports, for each of the seven configurations of the main data set, the median hardware fit $\hat{\alpha}_{\text{obs}}$ and the median $\hat{\alpha}_{\text{RO}}$ from the simulation with readout error only described in Methods. The median signed asymmetry of the per-qubit readout error rates is $a_{\text{RO}} := \text{median}_j(p_{01,j} - p_{10,j}) \approx 0.0017$. The difference $\hat{\alpha}_{\text{obs}} - \hat{\alpha}_{\text{RO}}$ is computed for each circuit and summarized by its median and interquartile range across circuits. Labels use the following rule: *positive* means the interquartile range is strictly positive with median difference > 0.05 , *small positive* means it is strictly positive with median difference ≤ 0.05 , and *IQR includes 0* marks the remaining rows. The readout-only fits give a median $\hat{\alpha}_{\text{RO}}$ that rounds to 0.000 in all seven configurations.

6.3 K_{sym} construction and optimizer robustness

The alternative components below are evaluated with the same MolQAE $n_q=4$, $N_{\text{circ}}=100$ five-fold cross-validation as Supplementary Table 12. Supplementary Table 5 compares the row-normalized sector component of SA-IBF with four alternatives: column normalization, row/column averaging, a q -independent uniform distribution within each sector, and the identity within each sector. Column normalization and row/column averaging coincide with row normalization exactly, because K_{IBF} is symmetric in its two arguments and Z_x depends only on $H(x)$. The uniform component

Supplementary Table 4 Hardware and readout-only fits of $\hat{\alpha}$ for each configuration of the main data set. $\hat{\alpha}_{\text{obs}}$ and $\hat{\alpha}_{\text{RO}}$ are configuration medians of the hardware and readout-only fits, and the difference column is the median over circuits of $\hat{\alpha}_{\text{obs},i} - \hat{\alpha}_{\text{RO},i}$, with its interquartile range (IQR) over circuits. The label is positive when the IQR lies above 0 and the median difference exceeds 0.05, small positive when the IQR lies above 0 and the median difference is at most 0.05, and IQR includes 0 for all other rows.

Family	n_q	N_{circ}	$\hat{\alpha}_{\text{obs}}$	$\hat{\alpha}_{\text{RO}}$	difference	IQR	Label
UCCSD	4	30	0.059	0.000	+0.053	[+0.037, +0.071]	positive
Magnetization	4	60	0.351	0.000	+0.293	[+0.105, +0.390]	positive
Magnetization	6	40	0.262	0.000	+0.261	[+0.169, +0.347]	positive
MolQAE	4	20	0.134	0.000	+0.086	[-0.323, +0.135]	IQR includes 0
MolQAE	5	20	0.016	0.000	+0.014	[-0.186, +0.017]	IQR includes 0
MolQAE	6	20	0.013	0.000	+0.013	[+0.011, +0.015]	small positive
MolQAE	7	40	0.077	0.000	+0.072	[+0.044, +0.093]	positive

changes the mean D_{H} by less than 10^{-4} , and the identity component raises it by 3.7×10^{-4} . At fixed parameters, all stochastic sector-preserving components give the same sector population as Eq. (5) and differ only in the within-sector shape, which therefore has little effect on D_{H} here.

Supplementary Table 5 Alternative within-sector components and an optimizer-grid check under five-fold cross-validation on MolQAE $n_q=4$. $\bar{\alpha}$ and $\bar{q}$ are fold means, and lower D_{H} is better. Shading marks the sector component of SA-IBF.

Variant	$\bar{\alpha}$	$\bar{q}$	mean D_{H}
Row-normalized K_{sym} (SA-IBF)	0.40618	0.42085	0.151460
Column-normalized K_{sym}	0.40618	0.42085	0.151460
Row/column averaged K_{sym}	0.40618	0.42085	0.151460
Uniform within each sector	0.40620	0.42065	0.151397
Identity within each sector	0.40546	0.42067	0.151827
Fine-grid optimizer check	0.40620	0.42086	0.151461

6.4 Search bound on q

We refit each of the seven configurations of the main data set with $q_{\text{max}} \in \{0.499, 0.6, 0.7, 0.99\}$, using the same optimizer as in Methods and keeping all other settings fixed. Six of the seven configurations have no fit within 0.001 of $q_{\text{max}}=0.499$, and their D_{H} is the same under all four bounds. On MolQAE $n_q=5$, 85% of the fits lie within 0.001 of $q_{\text{max}}=0.499$. Raising the bound to $q_{\text{max}}=0.99$ lowers the median D_{H} by 7×10^{-5} , shifts the median $\hat{\alpha}$ by less than 0.002, and leaves every regime label, conditioning verdict, and Ji-BO verdict unchanged.

Supplementary Note 7: Processors, registers, time, and circuit families

7.1 Processor and temporal checks

Supplementary Table 6 summarizes the checks across processors and over time. SA-IBF is more accurate than IBF on every processor, while $\hat{\alpha}$ varies between processors.

Supplementary Table 6 Checks across processors and over time. Each row states the design and the result. Values after $\pm$ are standard errors.

Data set	Design	Result
MolQAE $n_q=4$ four-processor data set	Same 100 circuits on <code>ibm_boston</code> , <code>ibm_kingston</code> , <code>ibm_pittsburgh</code> , and <code>ibm_marrakesh</code> .	SA-IBF has lower D_H than IBF on 400/400 executions, and $\hat{\alpha}$ varies between processors (0.207–0.495).
Pittsburgh Magnetization $n_q=4$	Comparison of Boston and Pittsburgh.	SA-IBF has lower D_H than IBF on 30/30 Pittsburgh circuits. Boston and Pittsburgh $\hat{\alpha}$ differ by 0.300 ± 0.021 .
Second Kingston run	Same processor and same four physical qubits about 10 days apart, with a different molecule set in the second run.	The mean D_H of SA-IBF changes by $+7.2 \times 10^{-4} \pm 3.9 \times 10^{-3}$, consistent with zero.

7.2 MolQAE $n_q = 4$ four-processor data set

Supplementary Table 7 summarizes the 400 executions of the MolQAE $n_q=4$ four-processor data set: the same 100 circuits are run on each of four IBM Heron processors with 30,000 shots per circuit and evaluated by five-fold cross-validation. SA-IBF has lower D_H than IBF on all 400 executions, with mean paired differences (IBF minus SA-IBF) from 0.074 ± 0.002 (`ibm_marrakesh`) to 0.131 ± 0.002 (`ibm_kingston`). The fitted $\hat{\alpha}$ varies between processors, from 0.207 on `ibm_marrakesh` to 0.495 on `ibm_kingston`.

Supplementary Table 7 MolQAE four-processor data set on IBM Heron processors. D_H entries are means $\pm$ standard errors across circuits, $\hat{\alpha}$ carries its standard error from resampling circuits, ΔD_H is the mean paired difference IBF minus SA-IBF, and wins count circuits where SA-IBF has the lower D_H . Shading marks SA-IBF.

Processor	$\hat{\alpha}$	$\hat{q}$	SA-IBF D_H	IBF D_H	ΔD_H	Wins
<code>ibm_boston</code>	0.406 ± 0.010	0.421	0.1514 ± 0.0036	0.2396 ± 0.0052	$+0.088 \pm 0.002$	100/100
<code>ibm_kingston</code>	0.495 ± 0.003	0.451	0.1248 ± 0.0013	0.2557 ± 0.0027	$+0.131 \pm 0.002$	100/100
<code>ibm_pittsburgh</code>	0.364 ± 0.006	0.436	0.1007 ± 0.0042	0.2067 ± 0.0049	$+0.106 \pm 0.002$	100/100
<code>ibm_marrakesh</code>	0.207 ± 0.003	0.481	0.1036 ± 0.0028	0.1777 ± 0.0042	$+0.074 \pm 0.002$	100/100

7.3 Second Kingston run

We compare 30 MolQAE $n_q = 4$ circuits run on `ibm_kingston` with the 100 circuits of the Kingston run of the four-processor data set, taken about ten days earlier. The two runs use the same four physical qubits [20, 21, 22, 36] and the same logical-to-physical layout. The mean D_H of the SA-IBF fit changes by $+7.2 \times 10^{-4} \pm 3.9 \times 10^{-3}$ between the two runs, consistent with zero.

The fitted weight changes between the two runs. Mean $\hat{\alpha}$ decreases by 0.143 ± 0.007 (second minus first), and the measured distributions of the second run lie farther from their ideal distributions, with a mean Hellinger distance larger by 0.084 ± 0.004 . The two runs differ in time and in molecule set.

7.4 $\hat{\alpha}$ across processors

We compare the fitted $\hat{\alpha}$ of the same circuit family on two processors and report the mean difference $\pm$ standard error. For Magnetization $n_q=4$, Boston and Kingston differ by 0.013 ± 0.023 . For MolQAE $n_q=6$, Boston and Kingston differ by -0.240 ± 0.003 , and for Magnetization $n_q=4$, Boston and Pittsburgh differ by $+0.300 \pm 0.021$. The fitted $\hat{\alpha}$ can therefore differ substantially between processors.

7.5 Register dependence

Supplementary Table 8 groups the MolQAE $n_q=4$ circuits by the physical register recorded for each execution and applies five-fold cross-validation within each group. SA-IBF has lower D_H than IBF in every group and on 82 of 85 circuits, while $\hat{\alpha}$ ranges from 0.14 to 0.48 across groups. The transfer test (Methods) is evaluated between groups on the same processor, which gives 14 ordered pairs. The transferred parameters give the lower mean D_H in 12 of the 14 transfers. The two exceptions are transfers from the July Boston groups J1 and J2 into group A2, the MolQAE circuits of the main data set on register [20, 21, 22, 36].

Supplementary Table 8 Register groups of MolQAE circuits. J1 and J2 are the Boston circuits of the three-scale gate-folding experiment at scale 1. A1 and A2 are the Boston circuits of the two-scale experiment at the original depth and the MolQAE $n_q=4$ circuits of the main data set. K1 and K2 are the Kingston circuits of the three-scale experiment. $\hat{\alpha}$ is the pooled fit with a 95% interval from resampling circuits, D_H entries are means under five-fold cross-validation, and wins count circuits where SA-IBF has the lower D_H . Shading marks SA-IBF.

Code	Group	Register	N	$\hat{\alpha}$	95% interval	SA-IBF	IBF	Wins
J1	Boston, Jul.	[60, 61, 62, 76]	20	0.364	[0.312, 0.418]	0.2186	0.2869	20/20
J2	Boston, Jul.	[151, 152, 153, 154]	15	0.446	[0.333, 0.539]	0.1847	0.2845	15/15
A1	Boston, Apr.	[20, 21, 22, 36]	15	0.259	[0.238, 0.282]	0.2340	0.2739	15/15
A2	Boston, Apr.	[20, 21, 22, 36]	20	0.139	[0.117, 0.160]	0.2192	0.2387	17/20
K1	Kingston	[136, 141, 142, 143]	6	0.483	[0.469, 0.494]	0.1199	0.2452	6/6
K2	Kingston	[138, 150, 151, 152]	9	0.479	[0.362, 0.550]	0.1753	0.2798	9/9

7.6 Diagnostic circuits on Boston

The 120 Boston diagnostic circuits form two checks that are part of the hardware data but separate from the main data set (Supplementary Table 9). First, 90 circuits test the GHZ, Clifford, and near-boundary families at 30 circuits each. GHZ circuits give the largest fitted weight of the three families ($\hat{\alpha} = 0.808$), and SA-IBF has the lower D_H on all of them (median $\Delta = +4.6 \times 10^{-3}$, 30/30 wins). Near-boundary circuits give a smaller improvement ($\hat{\alpha} = 0.388$, median $\Delta = +1.2 \times 10^{-3}$, 23/30 wins), and Clifford circuits give lower SA-IBF D_H on 27/30 circuits with a median difference of ≈ 0 . Second, 30 circuits form the 15 MolQAE pairs of the two-scale gate-folding experiment, each measured at the original depth and with every two-qubit gate folded threefold. SA-IBF and IBF are fitted at each depth, and their predictions are extrapolated to zero noise (Methods). The extrapolated SA-IBF prediction is closer to P_{ideal} for all 15 pairs, with median clipped D_H of 0.498 versus 0.609 for IBF.

Supplementary Table 9 Boston diagnostic circuits in the hardware data. For the diagnostic families, Δ is the median D_H^{IBF} minus the median $D_H^{\text{SA-IBF}}$ across circuits. For the two-scale gate-folding experiment, Δ is the median paired difference, IBF minus SA-IBF, in the clipped D_H between the extrapolated prediction and P_{ideal} . Positive Δ favors SA-IBF, and wins count cases where SA-IBF has the lower D_H .

Family	Circuits	Fit summary	Δ	Wins
<i>Diagnostic families</i>				
GHZ	30	$\hat{\alpha} = 0.808$, $\hat{q} = 0.068$	$+4.6 \times 10^{-3}$	30/30
Clifford	30	$\hat{\alpha} = 0.333$, $\hat{q} = 0.152$	≈ 0	27/30
Near-boundary	30	$\hat{\alpha} = 0.388$, $\hat{q} = 0.054$	$+1.2 \times 10^{-3}$	23/30
<i>Two-scale gate folding</i>				
MolQAE	30 (15 pairs)	clipped D_H : 0.498 vs. 0.609	$+1.19 \times 10^{-1}$	15/15

Supplementary Note 8: Sector observables and ZNE

8.1 MolQAE sector observables

For the MolQAE configurations, we use the same 50/50 binomial-thinning split of each bitstring count as in the main data set and compare absolute errors of diagonal observables computed from the predicted and test histograms. Supplementary Table 10 reports median error differences, IBF minus SA-IBF, so positive entries favor SA-IBF.

8.2 Zero-noise extrapolation of Π

Supplementary Table 11 gives the zero-noise-extrapolated probability that qubits 0 and 1 both read 0, $\Pi(P) = \sum_{y: y_0=y_1=0} P(y)$, for all gate-folding circuits. The 50 Boston circuits comprise the 15 pairs of the two-scale experiment and the 35 circuits of the three-scale experiment. At every scale, SA-IBF and IBF are fitted to the full histogram of each circuit, the value of Π on each fitted prediction is extrapolated to zero noise,

Supplementary Table 10 MolQAE sector observables. “Ideal sector” is the population $P(H=0)$ of the ideal sector, the all-zero probability, and “sector L1” is the L1 error of the sector populations. Positive values favor SA-IBF, and wins count circuits where SA-IBF has the lower L1 error.

Family	n_q	Δ ideal sector	Δ sector L1	Sector L1 wins
MolQAE	4	$+2.89 \times 10^{-2}$	$+7.00 \times 10^{-2}$	20/20
MolQAE	5	$+1.18 \times 10^{-2}$	$+2.00 \times 10^{-2}$	20/20
MolQAE	6	$+9.20 \times 10^{-3}$	$+1.26 \times 10^{-2}$	16/20
MolQAE	7	$+5.88 \times 10^{-2}$	$+1.08 \times 10^{-1}$	40/40

and the error is measured against $\Pi(P_{\text{ideal}})$. Sector-aware Ji-BO uses the calibration recorded on the day of execution, which is available for the 35 Boston circuits of the three-scale experiment. On the Kingston circuits, SA-IBF has the lower median error under both extrapolations, and its error is smaller than that of IBF on 7 and 10 of the 15 circuits under two-point and three-point extrapolation.

Supplementary Table 11 Median absolute error of the zero-noise-extrapolated probability Π that qubits 0 and 1 both read 0, under two-point and three-point extrapolation (Methods). Wins count circuits where SA-IBF has the smaller error. Shading marks SA-IBF.

Processor	N	Extrapolation	SA-IBF	IBF	Ji-BO	Wins vs IBF	Wins vs Ji-BO
Boston	50	two-point	0.355	0.410	—	46/50	—
Boston, three-scale	35	two-point	0.316	0.346	0.353	31/35	25/35
Boston, three-scale	35	three-point	0.214	0.303	—	26/35	—
Kingston	15	two-point	0.229	0.247	—	7/15	—
Kingston	15	three-point	0.111	0.254	—	10/15	—

Supplementary Note 9: Comparison with Ji-BO

This note gives the full Ji-BO comparison at $n_q=4$, three-parameter variants at $n_q=6$, the shot-noise floor behind the tolerance, and the convergence of Ji-BO.

9.1 Full comparison at $n_q=4$

Supplementary Table 12 reports the full $n_q=4$ Ji-BO comparison on MolQAE and Magnetization summarized in the main text. For each family the highlighted row is SA-IBF, with 2 parameters and a closed-form prediction. “s/fold” is the wall-clock time per fold, Δ_{Ji} is Ji-BO minus SA-IBF fold-mean D_{H} , and wins count test circuits where SA-IBF has the lower D_{H} . For the MolQAE sector-aware Ji-BO row with 100 trials, the 95% confidence interval over folds of the difference in fold-mean D_{H} is $[-0.006, +0.004]$, and the interval over paired circuits is $[-0.0026, +0.0005]$, both inside the tolerance $\delta=0.01$ of the main text.

CPU timing. Supplementary Table 12 reports batched Ji-BO with GPU timing in Qiskit Aer. A CPU rerun with the same implementation on the MolQAE $n_q=4$ data

with $N_{\text{circ}}=100$ gives comparable wall-clock times per fold (536 s for the 20-parameter and 520 s for the sector-aware variant, against 515 and 520 s on the GPU).

Supplementary Table 12 Ji-BO comparison on MolQAE and Magnetization under five-fold cross-validation. Lower D_H is better, Δ_{Ji} is the listed method minus SA-IBF, and wins count circuits where SA-IBF has the lower D_H . Trial budgets and wall-clock times per fold are listed for each row. SA-IBF runs on a CPU and Ji-BO with Aer on a GPU. ^aBest of four warm-started search strategies. Shading marks SA-IBF, and bold marks the smallest D_H in each family.

Family (n_q, N_{circ})	Method	Par.	Trials	D_H	s/fold	Δ_{Ji}	Wins
MolQAE (4, 100)	SA-IBF	2	—	0.15146	1.2	—	—
	Ji-BO	20	100	0.20736	515	+0.0559	87/100
	Sector-aware Ji-BO	21	100	0.15035	520	−0.0011	49/100
Magnetization (4, 60)	SA-IBF	2	—	0.0776	0.99	—	—
	Ji-BO	20	100	0.0933	136	+0.016	45/60
	Sector-aware Ji-BO	21	100	0.1285	123	+0.051	59/60
	Sector-aware Ji-BO, warm-started	21	200–500 ^a	0.0920	—	+0.014	—

9.2 Three-parameter variants at $n_q=6$

Supplementary Table 13 lists three-parameter variants of SA-IBF for the MolQAE $n_q=6$ comparison of Table 2.

Supplementary Table 13 Three-parameter variants of SA-IBF on the MolQAE $n_q=6$ comparison under the same five-fold cross-validation. Lower D_H is better. Δ is sector-aware Ji-BO minus the listed model, as in Table 2. Shading marks SA-IBF.

Model	Par.	D_H	Δ
SA-IBF	2	0.08492	−0.01978
Separate q for the two components	3	0.08488	−0.01974
Two-block logical- q grouping	3	0.08450	−0.01935
Even/odd logical- q grouping	3	0.08494	−0.01980

9.3 Tolerance and shot noise

The MolQAE Ji-BO comparison evaluates each test circuit against its full 30,000-shot hardware histogram, without a shot split. The 15,000-shot column below corresponds to the 50/50 test histograms of the main data set. To isolate multinomial noise in the test data, we fix the five-fold SA-IBF and sector-aware Ji-BO predictions and resample only the test histogram, $\hat{P}_i^{(b)} \sim \text{Multinomial}(N, P_{\text{obs},i})/N$. The observed

mean $\Delta D_H := D_H(P_{Ji}, P_{obs}) - D_H(P_{SA}, P_{obs})$ is -1.11×10^{-3} . We center each of 20,000 Monte Carlo realizations at this mean, so the half-widths in Supplementary Table 14 describe multinomial fluctuation alone.

Supplementary Table 14 Multinomial shot-noise floor for the MolQAE Ji-BO comparison with fixed predictions. ΔD_H is sector-aware Ji-BO minus SA-IBF before centering. Entries are central-95% half-widths of the centered ΔD_H distribution over Monte Carlo resamples of the test histograms.

Aggregation level	30,000 shots	15,000 shots
100-circuit paired mean	1.13×10^{-4}	1.60×10^{-4}
Fold mean (median over 5 folds)	2.61×10^{-4}	3.72×10^{-4}
Per circuit	1.14×10^{-3}	1.60×10^{-3}

9.4 Convergence of the Ji-BO search on MolQAE

Supplementary Table 15 reports the running-best training-fold D_H at trials 10, 25, 50, 75, and 100 for both Ji-BO variants on each of the five MolQAE $n_q=4$ folds with $N_{circ}=100$. The 20-parameter Ji-BO reaches its final running-best value by trial 10 on every fold and does not improve between trials 10 and 100. Sector-aware 21-parameter Ji-BO plateaus by trial 50 on 2 of 5 folds, one of them already by trial 25. On the remaining three folds the running best continues to improve between trials 50 and 100 (relative reduction 3.5%, 7.3%, and 2.5% on folds 2, 3, and 5). With 100 trials, the 95% confidence interval over folds is $[-0.006, +0.004]$ around a mean difference of -0.001 , inside the tolerance $\delta=0.01$.

Supplementary Table 15 Running-best training-fold D_H over the Optuna trials of the five-fold Ji-BO runs on MolQAE. Lower is better. “Plateaued” marks runs whose running best improves by less than 1% between the midpoint and the end of the search.

Variant	Fold	Trial 10	Trial 25	Trial 50	Trial 75	Trial 100	Plateaued?
Ji-BO	1	0.20602	0.20602	0.20602	0.20602	0.20602	yes
Ji-BO	2	0.19829	0.19829	0.19829	0.19829	0.19829	yes
Ji-BO	3	0.20467	0.20467	0.20467	0.20467	0.20467	yes
Ji-BO	4	0.20522	0.20522	0.20522	0.20522	0.20522	yes
Ji-BO	5	0.20620	0.20620	0.20620	0.20620	0.20620	yes
Sector-aware Ji-BO	1	0.16048	0.15793	0.15105	0.15105	0.15105	yes
Sector-aware Ji-BO	2	0.15710	0.15710	0.15651	0.15431	0.15097	no (−3.5%)
Sector-aware Ji-BO	3	0.17066	0.16240	0.16181	0.15090	0.14994	no (−7.3%)
Sector-aware Ji-BO	4	0.16547	0.15546	0.15546	0.15546	0.15546	yes
Sector-aware Ji-BO	5	0.16490	0.16377	0.14723	0.14358	0.14358	no (−2.5%)

Supplementary References

- [1] Wallman, J. J. & Emerson, J. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A* **94**, 052325 (2016).
- [2] Krantz, S. G. & Parks, H. R. *A Primer of Real Analytic Functions* 2nd edn (Birkhäuser, 2002).
- [3] Folland, G. B. *Real Analysis: Modern Techniques and Their Applications* 2nd edn (John Wiley & Sons, 1999).