MEASUREMENT AS COLD DIFFUSION: PRICING THE SHOT FRONTIER

Anonymous authors
Paper under double-blind review

ABSTRACT

Pipelines that reconstruct a quantum state's full Born distribution are usually built by importing classical Gaussian diffusion onto the device, which assumes away the cost that dominates them: the distribution must be read out by measurement, and estimating it over 2^n outcomes converges only as $1/\sqrt{m}$ in the shot budget m. We instead take the measurement to *be* the forward process, making the shot count the diffusion time. We prove that its control parameter is the only one among standard cold-diffusion corruptions that is a genuine resource—costing $\Theta(2^n)$ shots to hold the error fixed, while the Gaussian, blur and mask knobs are free—with a predicted constant confirmed to 2.1% and the scaling verified over a $256\times$ range in D. A learned prior *bends* that frontier, flattening it and saving up to $47.6\times$ of inference-time readout only for fixed-complexity data and only once the prior's acquisition cost is amortized; on ground-state Born distributions of an Ising chain the flattening fails. The corruption is also not interchangeable: restorers move between multinomial and Poisson readout at a cost of 0–3% but lose 57–229% moving from Gaussian or blur onto real readout. Finally, the prior does not appear able to be quantum. We characterize the class a postselecting circuit restorer realizes—the projectivization of a linear map, closed under coherent composition—and find classical controls with *fewer* parameters beating it at every scale, with logged curves showing it at a floor rather than out of steps. We are explicit that this is a characterization, not an impossibility proof. On `ibm_kawasaki` the idealized forward model tracks the device only while shots bind, yet a restorer trained solely on it still halves device error at the state-preparation floor.

1 INTRODUCTION

A diffusion model is a forward corruption process plus a learned reverse process. When the construction is carried onto quantum hardware, the forward process is almost always inherited unchanged from the classical setting: existing quantum diffusion models use a variety of engineered forward processes—random-unitary scrambling, a non-unitary channel driving the state to the maximally mixed state, and Gaussian-noise-assisted scrambling (Zhang et al., 2024; Chen et al., 2026; Li et al., 2025)—and a parameterized quantum circuit (PQC) is trained to reverse them. The move is natural, and it is physically wrong in a way that matters.

For a pipeline that requires an explicit Born distribution, the unavoidable degradation in accessing that distribution is not injected Gaussian noise but finite-shot *measurement*. The only way to obtain a distribution p over 2^n computational basis states is to sample it, and with a budget of m shots one observes an empirical histogram whose distance to p shrinks only as $1/\sqrt{m}$. This is not an implementation detail to be abstracted away; it is the dominant cost of any pipeline that must reconstruct the distribution, and it is why pure-quantum generative models have been criticized for reporting image quality at a qubit count where the cost is still invisible. A forward process built on artificial Gaussian noise cannot even express this cost, because its control knob is free: one may dial σ to any value at no price, whereas one cannot dial m without paying for it.

Measurement as the forward process. Let $p \in \Delta^{D-1}$, $D = 2^n$, be the Born distribution of a state over the computational basis and let

$$\hat{p}_m \;=\; \tfrac{1}{m}\,\mathrm{Multinomial}(m, p), \tag{1}$$

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

so that a small shot budget is a heavily degraded observation and a large budget a nearly clean one. In the language of cold diffusion (Bansal et al., 2023), Eq. (1) is a physically realized degradation operator and m plays the role of diffusion time; the reverse process is a learned restoration $R_\theta(\hat{p}_m, m) \approx p$. Nothing artificial is injected, and the diffusion time is now denominated in the same units as the hardware bill. Figure 1 shows the pipeline.

This reframing lets us pose three questions that the Gaussian formulation cannot, and we answer all three.

Contributions.

1. **A cost calculus for cold-diffusion control parameters** (§4). Holding the error at ε costs $\Theta(2^n)$ shots (Proposition 1), whereas the mask rate achieving the same error is *exactly* ε, free of D and p (Proposition 3), and the Gaussian scale *shrinks* as $1/D$ (Proposition 4). Only the measurement knob is a budget; the predicted constant matches measurement to 2.1%.
2. **The shot frontier** (§5). Naive readout pays the $\Theta(2^n)$ law; a learned restoration reaches the same error at a budget flat in n, saving $4.1\times$/$16.9\times$/$47.6\times$ at $n = 8, 10, 12$—but only for data whose intrinsic complexity does not grow with the grid, which we establish by exhibiting families on both sides, including genuine Born distributions where it fails.
3. **The forward process is not interchangeable** (§6). Under exact matching of degradation level, restorers move between multinomial and Poisson readout at a cost of 0–3% but lose 57–229% moving from Gaussian or blur onto readout. Counting corruption is its own equivalence class.
4. **What class a circuit restorer realizes** (§7). The tested restorer measures its ancillas and renormalizes, so it is *not* linear in ρ; we characterize what it is—the projectivization of a linear map, closed under coherent composition (Propositions 7, 8)—and derive a hypothesis that only repeated classical re-encoding should help, which holds. Remark 7.1 states what this does not establish.

2 Related Work

Cold diffusion. Bansal et al. (2023) built diffusion-style models on deterministic, non-Gaussian degradations such as blur and masking, reporting that generative behaviour is not strongly dependent on the choice of degradation; Hoogeboom & Salimans (2023) develop the blur case. Eq. (1) is a cold-diffusion forward process in that sense, with the degradation supplied by physics rather than chosen. §6 refines rather than contradicts their observation: each family is learnable when trained on itself, but the restorers are not interchangeable at deployment. The reverse-process backbone follows standard practice (Ho et al., 2020; Song et al., 2021).

Quantum generative models. Born machines generate by measuring a parameterized state (Benedetti et al., 2019; Liu & Wang, 2018), PQCs are universal generative models for suitable targets (Barthe et al., 2025), and Rudolph et al. (2024) map trainability barriers specific to this setting. Quantum diffusion constructions engineer a forward process—random-unitary scrambling (Zhang et al., 2024), a non-unitary channel driving the state to the maximally mixed state (Chen et al., 2026), or Gaussian-noise-assisted scrambling (Li et al., 2025)—and train a backward circuit to reverse it; our prior work accelerates that loop (Xu et al., 2026) and supplies the circuit we test. In all of them the readout cost is outside the analysis, which is the gap §4 addresses.

Expressivity, and shot-efficient readout. On classical inputs a re-uploading PQC is a partial Fourier series in the input (Pérez-Salinas et al., 2020; Schuld et al., 2021), highly nonlinear in it; that is not our regime. For quantum inputs the relevant limit is that pre-measurement transformations are linear in the state (Wang et al., 2025; Rattew et al., 2026), while Zhang et al. (2026) give copy-complexity bounds for estimating degree-k functionals; §7 connects these. Prior work studies trainability and classical simulability rather than the specific projective-linear function class considered here (McClean et al., 2018; Cerezo et al., 2025). Classical shadows (Huang et al., 2020) attack the same obstacle from the opposite side, changing *what* is estimated where we keep the full distribution and change the prior applied to it.

3 MEASUREMENT AS A COLD-DIFFUSION FORWARD PROCESS

Setup. A datum is a Born distribution over $D = 2^n$ computational basis states. A grayscale image with D pixels is mapped to $p \in \Delta^{D-1}$ by clipping to non-negative values, adding 10^{-6} and normalizing, so intensities are read as $|\text{amplitude}|^2$; a state preparing p has amplitudes $\sqrt{p}$. Experiments run at $n \in \{8, 10, 12\}$, extended to $n = 16$ for the resource calculus (§4) and reduced to $n = 6$ on hardware; images are bilinearly resampled to a $\sqrt{D} \times \sqrt{D}$ grid. Quality is measured by total variation, $\mathrm{TV}(a, b) = \frac{1}{2}\sum_i |a_i - b_i|$, the natural metric on Δ^{D-1} and directly comparable across n.

Forward and reverse. The forward process is Eq. (1). It models *ideal* finite-shot computational-basis readout, and we use that phrase rather than "the physical forward process": a real device also has readout error, SPAM and gate noise, none of which Eq. (1) represents, and Appendix C measures the resulting discrepancy on `ibm_kawasaki`. Within that scope it has the three properties a cold-diffusion forward process needs: it is monotone in m (Proposition 1 makes this quantitative), it is supplied by the measurement rather than injected by the modeller, and it is indexed by a scalar time that is also the resource being spent. The reverse process is a restoration $R_\theta(\hat{p}_m, m)$ trained across the whole shot schedule and conditioned on $\log m$ (Algorithm 1).

Two roles for the restorer. Sections 4–6 use a classical convolutional restorer, because there the question is a property of the *forward* process and the restorer is only an instrument. §7 then asks whether that instrument can itself be a circuit, which is what decides whether the pipeline can be pure-quantum.

4 WHAT THE CONTROL PARAMETER COSTS

Cold diffusion offers a menu of degradations. They are usually treated as interchangeable design choices. Under a resource account they are not, and the difference is not subtle: one of them is metered and the others are free. This section makes that precise. Proofs are in Appendix F.

For $p \in \Delta^{D-1}$ define the *effective support*

$$D_{\text{eff}}(p) \;:=\; \Big(\sum_{i=1}^{D} \sqrt{p_i}\Big)^2 \;\in\; [1, D], \tag{2}$$

which equals D for the uniform distribution and 1 for a point mass.

The regime that matters. The order of limits decides whether the statement is true. Writing $p_i = c_i/D$, the budgets this paper is about satisfy $m = \Theta(D)$ at fixed error, hence $mp_i = \Theta(1)$; a per-coordinate central limit theorem requires $mp_i \to \infty$ and is *not* available here. We state the result in the Poisson regime, where it is, and recover the CLT expression as a corollary with its hypothesis explicit.

Proposition 1 (The shot budget is metered, and the meter reads 2^n)**.** *Let $p_i^{(D)} = c_i^{(D)}/D$ with $\frac{1}{D}\sum_i c_i^{(D)} = 1$, let the empirical measure of $\{c_i^{(D)}\}_{i \le D}$ converge weakly to a law ν on $[0, \infty)$, and assume the array has uniformly bounded empirical second moments,*

$$\sup_D \; \frac{1}{D}\sum_{i=1}^{D} \big(c_i^{(D)}\big)^2 \;<\; \infty. \tag{3}$$

Let $m = \alpha D$ for fixed $\alpha > 0$. Then

$$\begin{aligned} &\mathbb{E}\big[\mathrm{TV}(\hat{p}_m, p^{(D)})\big] \;\xrightarrow[D\to\infty]{}\; \Phi(\alpha) \;:=\; \frac{1}{2\alpha}\int g(\alpha c)\, d\nu(c), \\ &\textit{where} \quad g(\lambda) := \mathbb{E}\big|\mathrm{Pois}(\lambda) - \lambda\big| = \frac{2\,\lambda^{\lfloor\lambda\rfloor+1} e^{-\lambda}}{\lfloor\lambda\rfloor!}. \end{aligned} \tag{4}$$

Φ is continuous and strictly decreasing with $\Phi(0^+) = 1$ and $\Phi(\infty) = 0$, so for every $\varepsilon \in (0, 1)$ there is a unique $\alpha(\varepsilon)$ with $\Phi(\alpha(\varepsilon)) = \varepsilon$, and the budget required to hold the error at ε is

$$m(\varepsilon) \;=\; \alpha(\varepsilon)\, D \;=\; \Theta(2^n). \tag{5}$$

In particular the error depends on m and D only through the ratio $\alpha = m/D$.

Corollary 2 (The CLT form, and when it may be used)**.** *If in addition $\min_i mp_i \to \infty$, then $g(\lambda) = (1 + o(1))\sqrt{2\lambda/\pi}$ uniformly on the relevant range and Eq. (4) reduces to $\mathbb{E}[\mathrm{TV}(\hat{p}_m, p)] = (1 + o(1))\sqrt{D_{\mathrm{eff}}(p)/(2\pi m)}$ with $D_{\mathrm{eff}}(p) = (\sum_i \sqrt{p_i})^2$, whence $m(\varepsilon) = (1 + o(1))D_{\mathrm{eff}}(p)/(2\pi\varepsilon^2)$.*

Remark (The corollary does not apply here, and we can say by how much)**.** The hypothesis $\min_i mp_i \to \infty$ fails in our setting, where $mp_i = \Theta(1)$. Applying Corollary 2 regardless overestimates the required budget by $+5.7\%, +4.0\%, +3.5\%$ at $n = 8, 10, 12$, whereas Proposition 1 is accurate to $-0.3\%, -1.3\%, -2.1\%$. We report both because the discrepancy is a quantitative confirmation that the Poisson regime, not the Gaussian one, is the right description of finite-shot readout at realistic budgets. Note also that the scaling conclusion $m = \Theta(D)$ survives either way; it is the constant that the CLT form gets wrong.

Proposition 3 (The mask rate is free, and is exactly the error)**.** *Let S retain each coordinate independently with probability $1 - r$, conditioned on $S \neq \emptyset$; the empty mask has probability $r^D > 0$ and leaves the renormalization undefined, and an empty draw is resampled in the implementation. Let q be p restricted to S and renormalized. Then $\mathrm{TV}(q, p) = 1 - \sum_{i \in S} p_i$ and $\mathbb{E}[\mathrm{TV}(q, p)] = r + O(r^D)$, independent of p and, up to the $O(r^D)$ correction, of D. At the values used here ($r \leq 0.95$, $D \geq 256$) the correction is below 10^{-5}.*

Proposition 4 (The Gaussian scale is free, and shrinks as $1/D$)**.** *Let $q \propto (p + \sigma\varepsilon)_+$ with $\varepsilon \sim \mathcal{N}(0, I_D)$, and assume the dense-profile regime in which we use it: $p_i = \Theta(1/D)$ uniformly in i and $\sigma = \Theta(1/D)$. Then $\mathbb{E}[\mathrm{TV}(q, p)] = \Theta(\sigma D)$, so holding the error at ε requires $\sigma(\varepsilon) = \Theta(\varepsilon/D)$: the control parameter must be made* smaller *as the outcome space grows. Outside this regime—for instance for p supported on $O(1)$ coordinates—the conclusion does not apply.*

Proposition 5 (The blur width is free, and grows as $\sqrt{D}$)**.** *Let the data be a fixed bounded profile on $[0, 1]^2$ resampled to a $\sqrt{D} \times \sqrt{D}$ grid and q be p blurred by a Gaussian kernel of width w* pixels*, renormalized. Blurring the grid by w pixels is blurring the profile by $w/\sqrt{D}$ in normalized coordinates, so in the continuum limit $\mathrm{TV}(q, p)$ depends on (w, D) only through $w/\sqrt{D}$; holding the error at ε thus requires $w(\varepsilon) = \Theta(\sqrt{D})$, set by the grid rather than paid for.*

Corollary 6 (Only the measurement control parameter is a resource)**.** *The control parameter needed to hold a fixed error scales as $\Theta(D)$ for multinomial and Poisson readout, $\Theta(1)$ for masking, $\Theta(1/D)$ for additive Gaussian noise and $\Theta(\sqrt{D})$ for blur (Propositions 1–5). Only in the first case is turning the knob something one must pay for: σ, r and w may be dialled to any value at no charge, whereas m cannot.*

Remark (Why this is the argument against a Gaussian forward process)**.** The usual objection to Gaussian corruption on a device is that it is unphysical. Corollary 6 gives a sharper one: it is *incapable of expressing the cost*. A model whose forward knob is σ can be made arbitrarily gentle at no charge, so no resource frontier can be read off it—precisely the accounting pure-quantum generative work has been faulted for omitting.

Empirical check. Proposition 1 makes two falsifiable claims. The *structural* one is that the error depends on (m, D) only through $\alpha = m/D$: Table 1 shows α constant to 6.1–6.8% as D ranges over $256\times$, and the finer ten-rung ladder at $n \leq 12$ over three seeds (Table 7) gives 0.3–2.7%, its two worst rungs being the two most degraded ones we exclude elsewhere, where $\alpha \approx 0.006$ leaves most bins empty. The $\Theta(2^n)$ law is therefore observed over five octaves, not extrapolated from three points. The *quantitative* one is the constant: solving $\Phi(\alpha) = 0.17$ from Eq. (4) predicts $262, 1058, 4197$ shots against measured $263, 1072, 4288$, errors of $-0.3\%, -1.3\%, -2.1\%$, where Corollary 2 would give $+5.7\%, +4.0\%, +3.5\%$ (Remark 4). Proposition 3 predicts a mask rate of exactly 0.170 at $\varepsilon = 0.17$; bisection returns $0.171, 0.170, 0.170$.

At matched difficulty the prior gets stronger, on this data. Fixing $\alpha = 1$ holds the naive error at 0.17 for every n, yet restored error falls $0.100 \to 0.026$ as D grows $256\times$ (Table 1, bottom): the prior does not merely keep up with the outcome space, it exploits it. The §5.1 qualification applies unchanged—finer MNIST is the same digit with more pixels—and the Ising family shows this does not extend to data whose complexity grows with D.

Table 1: **The $\alpha = m/D$ collapse over a $256\times$ range in D.** Budget at which each target error is reached, expressed as $\alpha = m/D$. Constancy of a row is Proposition 1's structural claim; it survives two further octaves. Single seed at each n; the three-seed version of the low-n columns is Table 7. Bottom block: at matched α, and therefore matched naive difficulty, the learned restoration improves steadily with n.

	n=8 (D=256)	n=10 (1024)	n=12 (4096)	n=14 (16384)	n=16 (65536)	spread
α at TV $= 0.33$	0.2676	0.2725	0.2721	0.2681	0.2850	6.5%
α at TV $= 0.24$	0.5106	0.5229	0.5218	0.5146	0.5453	6.8%
α at TV $= 0.17$	1.0365	1.0484	1.0486	1.0329	1.0958	6.1%
naive TV at α=1	0.172	0.174	0.174	0.173	0.178	—
restored TV at α=1	0.100	0.070	0.051	0.042	**0.026**	—

Table 2: **The shot frontier.** Left: error at a fixed budget of $m = 64$. Right: budget required to reach TV $= 0.17$. Naive readout pays the $\Theta(2^n)$ law of Proposition 1; the learned restoration does not, so the saving compounds. Naive budgets by bisection on m, restoration budgets by interpolation on the shot grid.

	TV at $m = 64$ shots			shots to reach TV $= 0.17$		
	n=8	n=10	n=12	n=8	n=10	n=12
naive readout	0.341	0.668	0.900	263	1072	4288
MCD restoration (ours)	0.169	0.171	0.186	64	64	≈ 90
shot saving	—	—	—	$4.1\times$	$16.9\times$	$47.6\times$

5 The Shot Frontier

Question. Prior pure-quantum generative models have been criticized on the grounds that readout shots scale exponentially in qubit count, making small-n image quality uninformative. Proposition 1 confirms the barrier is real for prior-free readout. Does it survive a learned manifold prior?

Result. Table 2 shows it does not. The saving grows as $4.1\times$, $16.9\times$, $47.6\times$.

5.1 How far does the flat frontier extend?

On MNIST, adding qubits makes the same digit higher-resolution; it does not make the data harder. A prior that flattens the barrier there may be exploiting fixed intrinsic complexity rather than saying anything general. That objection is correct, and Table 3 settles it by exhibiting cases on both sides. We hold the restorer, training budget and shot grid fixed and vary only the data.

Two synthetic families isolate the variable: images built from K Gaussian blobs with $K = 8$ fixed, and with $K = D/32$ so that degrees of freedom scale with the outcome space. Two real image families, MNIST and Fashion-MNIST, have fixed complexity by construction. The fifth family is not images at all but genuine Born distributions: ground states of a disordered transverse-field Ising chain,

$$H = -\sum_i J_i Z_i Z_{i+1} - \sum_i h_i X_i, \qquad J_i \sim \mathrm{Unif}[0.5, 1.5],\ h_i \sim \mathrm{Unif}[0.2, 1.8], \tag{6}$$

with $p_x = |\langle x|\psi_0\rangle|^2$. Their 2^n outcomes carry no spatial structure, so for that family we also report a permutation-agnostic MLP rather than assume a convolutional prior is appropriate for quantum data.

Result. The three fixed-complexity families are flat (1.05–$1.20\times$); the growing-complexity family degrades $1.64\times$; the Ising family degrades $2.12\times$ with a convolutional restorer and $2.85\times$ with the MLP. On genuine Born distributions a learned prior bends the frontier but does not flatten it. We

Table 3: **The flattening is a property of fixed-complexity data.** Restored error at a fixed budget $m = 64$. Only families whose degrees of freedom scale with D degrade. Last column is a property of the data, not of any model: Proposition 1 assumes D_{eff}/D tends to a constant, which the image families satisfy and the Ising ground states do not.

family	complexity	n=8	n=10	n=12	growth	D_{eff}/D
mnist	fixed	0.172	0.178	0.181	1.05×	0.197 / 0.198 / 0.197
fashion	fixed	0.148	0.148	0.177	1.20×	0.500 / 0.491 / 0.495
blobs-fixed	fixed (K=8)	0.136	0.134	0.140	1.03×	0.836 / 0.844 / 0.849
blobs-grow	grows (K=$D/32$)	0.141	0.189	0.231	1.64×	0.836 / 0.855 / 0.871
ising, conv	Born distributions	0.151	0.242	0.320	**2.12**×	0.356 / 0.286 / 0.218
ising, MLP		0.130	0.195	0.369	**2.85**×	

take this as the correct statement of what §5 shows: the flattening is a statement about data priors on low-intrinsic-dimension distributions, not a general improvement in the measurement complexity of Born distributions.

What survives, and where the hypothesis bites. The naive side is untouched: Proposition 1 depends on D_{eff}/D, not on intrinsic complexity, and at $\text{TV} = 0.25$ the naive budget grows $66\times$ over this range while the restored budget grows $8\times$—the prior bends the frontier where it stops flattening it. Two honest notes: the blob families independently confirm Proposition 1 on data it was never fitted to ($D_{\text{eff}}/D \approx 0.84$ against MNIST's 0.197 predicts $4.3\times$ more shots, measured 4.0), but the Ising family violates its hypothesis, D_{eff}/D falling $0.356 \to 0.218$ as amplitude mass concentrates, so the conclusion is not directly available there.

What the saving does and does not count. The numbers in Table 2 are *inference-time* readout costs. They exclude the one-time cost of obtaining and training the prior, which in our experiments is free only because the clean targets are free: MNIST and Fashion-MNIST are given, and the Ising Born distributions come from exact diagonalization. If the clean distribution were itself unknown—the situation the pipeline is ultimately for—acquiring training labels would carry its own, possibly large, measurement cost. What §5 reports is therefore an *amortized inference measurement cost*, not end-to-end resource accounting, and the saving is realized only when the prior is reused across many readouts. We state this explicitly because the $47.6\times$ figure is otherwise easy to over-read.

Reading. The claim is deliberately narrow. It is not that restoration is free—it is a learned classical model, and §7 shows it cannot presently be moved onto the device. It is that the exponential barrier is a property of *prior-free* readout rather than an information-theoretic obstruction: the data manifold has n-independent intrinsic complexity, and a model that knows the manifold need not resolve all 2^n bins to locate a point on it. Any resource accounting for quantum generative models should therefore quote the frontier, not the naive shot count. The flatness is a property of data whose intrinsic complexity does not grow with the grid, which §5.1 establishes by exhibiting a family where it does and the flatness fails.

6 Is the Measurement Forward Process Replaceable?

If the mechanism of §5 is merely "a learned prior beats no prior", the measurement forward process is a narrative device and any cold-diffusion corruption would serve. We test this directly and did not assume the answer. Holding data, architecture and training budget fixed, we vary only the corruption family—*shot* $\text{Multinomial}(m, p)/m$ (ours); *poisson* $\text{Poisson}(mp)/m$, the "it is only counting noise" null; *gauss* $p + \sigma\varepsilon$, the corruption used by prior quantum diffusion work; and *blur* and *mask*, the two canonical cold-diffusion degradations (Bansal et al., 2023). Families are matched by bisection (Algorithm 2) onto a common ladder of naive errors, so only structure differs.

Result 1: the cost calculus is confirmed. The counting families pay 2^n and the others pay nothing, exactly as Corollary 6 requires: at fixed naive error the shot budget scales $\times 4.08, \times 4.00$ per two added qubits, the Gaussian scale $\times 0.25$, the blur width $\times 1.9$ and the mask rate $\times 1.0$ (Table 4).

Table 4: **Only the counting families have a control parameter that is a resource.** Control parameter required to produce naive TV $= 0.17$, mean$\pm$s.d. over 3 seeds; ratios are per two added qubits. Predicted column from Propositions 1–4. This is the empirical face of Corollary 6.

family	control parameter	n=8	n=10	n=12	measured ratio	predicted
shot	shots m	263 ± 1	1072 ± 4	4288 ± 6	$\times4.08, \times4.00$	$\times4$ ($\propto D$)
poisson	rate m	263 ± 1	1072 ± 3	4289 ± 2	$\times4.08, \times4.00$	$\times4$ ($\propto D$)
gauss	scale σ	$2.37{\cdot}10^{-3}$	$5.93{\cdot}10^{-4}$	$1.48{\cdot}10^{-4}$	$\times0.25, \times0.25$	$\times\frac{1}{4}$ ($\propto 1/D$)
blur	width w	0.563	1.011	2.027	$\times1.79, \times2.01$	$\times2$ ($\propto \sqrt{D}$)
mask	rate r	0.171	0.170	0.170	$\times0.99, \times1.00$	$\times1$ (exactly ε)

Result 2: counting corruption is its own equivalence class. Training a restorer on one family and testing on another at matched degradation (Table 8), the shot and Poisson rows form a flat 2×2 block: substituting Poisson for multinomial readout costs $+0\%, +2\%, +3\%$ at $n = 8, 10, 12$, within seed noise. We therefore claim nothing more exotic for the measurement forward process than that it is a counting process—an honest reading that also sharpens the contrast. Restorers trained on Gaussian and blur degrade sharply on real readout: $57\% \pm 10$ and $152\% \pm 20$ at $n = 8$, $69\% \pm 14$ and $206\% \pm 23$ at $n = 10$, $62\% \pm 19$ and $229\% \pm 50$ at $n = 12$ (extremes over all seeds and n: $+43$ to $+89\%$ Gaussian, $+124$ to $+284\%$ blur). The penalty is large at every n; it is not monotone, and we report it as large rather than growing.

An honest limitation of the matching. At the two most severe levels the Gaussian and blur families cannot reach the target error at all—bisection saturates, with residual mismatch up to 0.16—because very low shot counts produce degradations additive Gaussian noise and blur cannot reproduce. Those levels are excluded; transfer is evaluated at a mid level where all five families match to better than 0.01.

7 Can the Restorer Be Quantum?

Sections 4–6 establish that the measurement forward process is the right one and that a learned prior flattens its frontier. If the restorer could be a PQC the pipeline would be pure-quantum. In every configuration we tested it is not, and this section tries to say what restricts the class it lives in. We state the shape of the argument at the outset, because it is weaker than a ceiling: a function-class characterization, plus a range of empirical failures, supporting a plausible hypothesis. Remark 7.1 says exactly where the gap is.

7.1 What class does a circuit restorer realize?

The architecture we test does not merely trace out its ancillas: it *measures* them, keeps the surviving block and renormalizes. That map is not linear in ρ, so the textbook "a circuit is linear" argument does not, as stated, apply to it. We therefore characterize the class the tested restorer actually realizes.

Proposition 7 (Projective-linear confinement)**.** *Let the restorer be handed $\hat{\rho}_m = |\psi_m\rangle\langle\psi_m|$ with $|\psi_m\rangle \propto \sum_i \sqrt{\hat{p}_{m,i}}\,|i\rangle$ embedded by an isometry ι into $n{+}n_a$ qubits, let it apply a parameterized unitary U_θ, measure the n_a ancillas obtaining outcome r, and renormalize the surviving n-qubit block. Write $K_r^\theta := (\langle r| \otimes I)\, U_\theta\, \iota$. Then the realized map is*

$$\hat{\rho}_m \;\longmapsto\; \Phi_r(\hat{\rho}_m) \;=\; \frac{K_r^\theta\, \hat{\rho}_m\, K_r^{\theta\dagger}}{\mathrm{Tr}\big(K_r^\theta\, \hat{\rho}_m\, K_r^{\theta\dagger}\big)}, \qquad \Pr[r] = \mathrm{Tr}\big(K_r^\theta \hat{\rho}_m K_r^{\theta\dagger}\big), \tag{7}$$

and consequently, for every θ, every circuit depth and every n_a: (i) *the unnormalized output $K_r^\theta \hat{\rho}_m K_r^{\theta\dagger}$ is a degree-1 functional of $\hat{\rho}_m$, so the realized map is the projectivization of a linear map;* (ii) *a single forward pass selects among at most 2^{n_a} such branches; and* (iii) *if K_r^θ is invertible then Φ_r is injective on pure states. Part (iii) does* not *transfer to what we measure: the readout that follows can merge distinct states into one distribution. With $D = 2$, positive amplitude encoding and $U_\theta = H$, the inputs $q = (0.2, 0.8)$ and $(0.8, 0.2)$ give distinct states $(0.949, \mp 0.316)$ but the same Born distribution $(0.9, 0.1)$. We therefore do not use injectivity to argue the induced map on distributions is not many-to-one.*

Part (i) is the honest replacement for the linearity slogan: renormalization does buy nonlinearity, but only the nonlinearity of a projective quotient, and the numerator stays degree-1 no matter how deep the circuit is. Related quantum implicit-representation work likewise identifies the restriction to pre-measurement unitary transformations as an important expressivity limitation (Wang et al., 2025), without giving the projectivization characterization used here. Part (iii) says the class cannot implement an exactly many-to-one map, which restoration is.

A natural reading of Proposition 7 is that adding measurement rounds escapes it. It does not. Appendix A shows that T coherent rounds compose to $K_{\mathbf{r}} = K_{r_T} \cdots K_{r_1}$, again the projectivization of a single linear map (Proposition 8): depth, rounds and ancillas buy $2^{n_a T}$ branches, not degree. What *is* outside the class is *repeated* re-encoding, and Proposition 9 sharpens this against us: a single re-uploading round does not leave the class at all—it is Eq. (7) with $K = U_\theta$—so the escape begins only at $T \geq 2$, where the phase discarded between rounds makes the composite piecewise rather than globally analytic. We retract two earlier phrasings: "a measurement raises the degree", and "re-encoding leaves the class for any $T \geq 1$" (Remark A).

Remark (What this does and does not establish)**.** This is the weakest joint in the paper and we mark it explicitly. Proposition 7 characterizes the realizable class exactly; it does *not* show restoration is unrealizable within it. A projective-linear map can contract volumes sharply through small singular values, and—as Proposition 7(iii) now records—the computational-basis readout can merge distinct states into one distribution outright, so nothing here forbids the many-to-one behaviour restoration needs. "Restoration requires high-degree functionals" is a conjecture we find plausible and do not prove. The copy-complexity bound of Zhang et al. (2026) prices a *different* escape route from the one our probe takes (spending copies of ρ), does not apply to classical re-encoding, and is not a lower bound on arbitrary restoration. What we have is a characterization, a hypothesis it suggests, and a floor that seven probes agree on.

A hypothesis, and its test. Propositions 7–9 sort architectural changes into two groups: more parameters, ancillas, depth, coherent rounds or a different objective all leave the realizable class unchanged, and only re-encoding repeated over $T \geq 2$ rounds leaves it. That suggests—no more strongly than that, since we have not shown the class cannot restore—the hypothesis that *among circuit-based variants, only those that re-encode classical statistics should improve materially*. It is suggested by the characterization, not implied by it. It survives the test in §7.2: of five architectural variants, re-uploading is both the only one that re-encodes and the only one that improves materially.

Remark (Scope)**.** This argument applies because the restorer's input is a quantum state. It does *not* say PQCs are linear models in general: with classical inputs supplied by data re-uploading the output is a Fourier series in the classical input and is highly nonlinear in it, which is exactly how re-uploading models obtain their expressivity (Pérez-Salinas et al., 2020; Schuld et al., 2021). The projective-linear restriction binds in the quantum-data regime—inputs presented as ρ and carried coherently—which is the regime a pure-quantum generative pipeline is committed to, because its whole point is that the data never leaves the device as a classical vector.

7.2 PROBES

We use the QDDPM backward circuit (Zhang et al., 2024) as the restorer: a hardware-efficient ansatz on $n+n_a$ qubits with $2(n+n_a)L$ parameters, ancilla measurement and renormalization. Each probe removes one escape route, and each leaves the circuit at a floor (Table 9; details in Appendix D). Two are decisive because they hand the circuit the answer. *Memorization*: asked to fit eight fixed pairs and evaluated on those same pairs, the circuit reaches a *training* TV of 0.287 against a naive 0.489—it cannot memorize eight images, and Figure 6 shows what it produces instead. *Distillation*: handed a trained classical teacher's outputs as targets ($\mathrm{TV} = 0.218$), the student reaches only 0.384. Neither failure can be attributed to generalization, and the optimization traces below make simple undertraining unlikely; we do not claim the circuit is globally optimized.

Measurement re-uploading is the exception that tests the hypothesis. It is the one architecture that performs the classical re-encoding of Proposition 9: each round discards the state and prepares a new one, $|\psi_q\rangle = \sum_i \sqrt{q_i}|i\rangle$, from the current classical estimate q. By Proposition 8 this is the only one of our variants that leaves the projective-linear class at all. It is also the only circuit-based variant that improves materially, from 0.489 to 0.337—and with four rounds it still loses to a classical control of comparable size at 0.241, which is what a linear-in-k price for degree predicts.

Table 5: **Ideal forward model against `ibm_kawasaki`**, $n = 6$, six amplitude-encoded states, one job. "ratio" is device naive error divided by ideal naive error; "cut" is the reduction the ideal-trained restorer achieves on device data. The device floor near TV ≈ 0.21 is invisible to Eq. (1).

	naive		restored (ideal-trained)			
m	ideal	device	ideal	device	ratio	cut on device
32	0.274	0.345	0.230	0.304	1.26×	12%
128	0.129	0.244	0.112	0.187	1.90×	23%
512	0.069	0.222	0.081	0.126	3.24×	43%
2048	0.032	0.211	0.043	0.106	6.67×	50%
8192	0.016	0.213	0.034	0.104	13.13×	51%

7.3 Scale and budget controls

Reading the table. *(i)* The classical control uses *fewer* parameters at every row and wins every time; against the *worst* of ten classical restarts the circuit still loses everywhere (0.382 vs 0.319, 0.469 vs 0.306, 0.486 vs 0.323, 0.373 vs 0.279), so the deficit is neither a parameter-count artifact nor a lucky draw. *(ii)* It does *not* simply grow with n, and we correct an earlier reading of our own data: the gap runs $+0.118, +0.175, +0.168$, larger at $n \geq 10$ but with $n = 10$ and $n = 12$ within a s.d., so we claim no monotone trend. What is robust is the deficit itself, 0.12–0.18, an order of magnitude above the 0.003–0.021 restart spread. *(iii)* The memorization probe *does* degrade monotonically, in a comparison rigged in the circuit's favour—its column is a *training* error, the teacher's a *test* error—the losing margin running $0.073, 0.116, 0.181$. A circuit that cannot memorize eight examples, by a margin growing with the outcome space, is not short of training signal. *(iv)* Doubling the circuit buys almost nothing: $216 \to 432$ parameters moves distillation only $0.382 \to 0.373$, while the classical control gains more ($0.257 \to 0.244$), so the gap widens to $+0.130$.

The probes are converged, not truncated. The natural objection to a negative result is undertraining, so we log the optimization. At $n = 10$ memorization runs $0.695 \to 0.413 \to \cdots \to 0.350$ over 1400 steps, with 82% of the descent in the first 200 and the remaining 1200 buying 0.063; distillation is starker, $0.645 \to 0.403 \to 0.407 \to 0.404$, flatlining after step 200 with the teacher's own outputs as targets. The circuit is not still climbing when we stop it.

A hardware check on the forward model itself. Amplitude-encoding six $n = 6$ distributions on `ibm_kawasaki`, a 156-qubit Heron device, gives Table 5 (Appendix C develops it). The ideal model is adequate only while shots bind: device error saturates near TV ≈ 0.21 by $m \approx 512$ while the ideal model keeps falling as $m^{-1/2}$, so the discrepancy grows to 13.1×—Eq. (1) cannot express a state-preparation floor. Yet the restorer, trained only on the ideal model and never shown device data, halves device error at that floor, $0.213 \to 0.104$: a floor a manifold prior can halve is structured error, not an information-theoretic limit.

8 Conclusion

Taking finite-shot readout to be the forward process makes the dominant cost of a Born-distribution pipeline expressible, which importing Gaussian diffusion onto a device does not: it is the only standard cold-diffusion corruption whose control parameter is a resource, costing $\Theta(2^n)$ shots to hold error fixed—proved in the Poisson regime where realistic budgets sit and verified over a 256× range in D—while the Gaussian, blur and mask knobs are free. A learned prior bends that frontier and on fixed-complexity data flattens it, though not on Ising ground-state Born distributions. It cannot presently be moved onto the device: a postselecting circuit restorer realizes the projectivization of a linear map, closed under coherent composition, and every variant inside that class sits at a floor while only classical re-encoding repeated over rounds escapes—a characterization plus converging empirical failures, not an impossibility proof. Whether the degree restoration needs can be bought at a price still beating classical readout—most plausibly under purified access, where degree k costs $\Theta(\sqrt{k})$ copies rather than $\Omega(k)$ (Zhang et al., 2026)—is the open question.

Reproducibility statement

Every number in this paper was produced by code we ran; none is quoted from another source. Appendix H gives the data preparation, all three restorer architectures with their parameter counts, optimizers, batch sizes, step counts, shot grids and seed counts, together with the settings for each probe (Appendix D), the synthetic and Ising families, and the extended-range runs. Proofs of all propositions are in Appendix F, each stating the hypotheses it uses. The hardware experiment (Appendix C) reports the backend, the transpiled depth and two-qubit gate counts, and the single-job subsampling protocol; it is one job on one backend and we say so. Where a result rests on a single run rather than multiple seeds—the $n = 12$ and $L = 24$ circuit configurations—the table marks it with a dagger, and where run-to-run spread matters we report it as mean$\pm$s.d. over restarts rather than a single draw.

References

Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie S. Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. arXiv:2208.09392.

Alice Barthe, Michele Grossi, Sofia Vallecorsa, Jordi Tura, and Vedran Dunjko. Parameterized quantum circuits as universal generative models for continuous multivariate distributions. *npj Quantum Information*, 2025. doi:10.1038/s41534-025-01064-3.

Marcello Benedetti, Delfina Garcia-Pintos, Oscar Perdomo, Vicente Leyton-Ortega, Yunseong Nam, and Alejandro Perdomo-Ortiz. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Information*, 5(45), 2019.

M. Cerezo, Martin Larocca, Diego García-Martín, N. L. Diaz, Paolo Braccia, Enrico Fontana, Manuel S. Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, Eric R. Anschuetz, and Zoë Holmes. Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing. *Nature Communications*, 16:7907, 2025. doi:10.1038/s41467-025-63099-6.

Chuangtao Chen, Qinglin Zhao, MengChu Zhou, Zhimin He, Zhili Sun, and Haozhen Situ. Quantum generative diffusion model: A fully quantum-mechanical model for generating quantum state ensemble. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026. arXiv:2401.07039, doi:10.1109/TPAMI.2026.3718311.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Emiel Hoogeboom and Tim Salimans. Blurring diffusion models. In *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2209.05557.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16:1050–1057, 2020.

Yihua Li, Jiayi Chen, Tamanna S. Kumavat, and Kyriakos Flouris. Unitary scrambling and collapse: A quantum diffusion framework for generative modeling. *arXiv preprint arXiv:2506.10571*, 2025.

Jin-Guo Liu and Lei Wang. Differentiable learning of quantum circuit born machines. *Physical Review A*, 98:062324, 2018. arXiv:1804.04168.

Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(4812), 2018.

Adrián Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I. Latorre. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020. arXiv:1907.02085.

Arthur G. Rattew, Po-Wei Huang, Naixu Guo, Lirandë Pira, and Patrick Rebentrost. Accelerating inference for multilayer neural networks with quantum computers. In *International Conference on Learning Representations (ICLR)*, 2026.

Manuel S. Rudolph, Sacha Lerch, Supanut Thanasilp, Oriel Kiss, Oxana Shaya, Sofia Vallecorsa, Michele Grossi, and Zoë Holmes. Trainability barriers and opportunities in quantum generative modeling. *npj Quantum Information*, 10(116), 2024.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430, 2021.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Shuteng Wang, Christian Theobalt, and Vladislav Golyanik. Quantum visual fields with neural amplitude encoding. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Jian Xu, Wei Chen, Shigui Li, Chao Li, Delu Zeng, John Paisley, and Qibin Zhao. Local-time riemannian score matching on the quantum pure-state manifold. *arXiv preprint arXiv:2605.03573*, 2026.

Bingzhi Zhang, Peng Xu, Xiaohui Chen, and Quntao Zhuang. Generative quantum machine learning via denoising diffusion probabilistic models. *Physical Review Letters*, 132:100602, 2024. doi:10.1103/PhysRevLett.132.100602.

Yukun Zhang, Yusen Wu, You Zhou, and Xiao Yuan. Measuring less to learn more: Quadratic speedup in learning nonlinear properties of quantum states. *Physical Review Letters*, 136:130602, 2026. doi:10.1103/cllg-15kd.

A Deferred Statements

A natural reading of Proposition 7 is that adding measurement rounds escapes it. It does not, and saying so precisely is what makes the characterization useful.

Proposition 8 (Coherent composition stays in the class)**.** *Suppose the restorer performs T rounds of "unitary, then measure n_a ancillas, then renormalize the surviving block", with the post-measurement state carried forward coherently. Then, conditional on the measurement trajectory $\mathbf{r} = (r_1, \ldots, r_T)$, the realized map is*

$$\hat{\rho} \;\longmapsto\; \frac{K_{\mathbf{r}}\,\hat{\rho}\,K_{\mathbf{r}}^{\dagger}}{\mathrm{Tr}\big(K_{\mathbf{r}}\,\hat{\rho}\,K_{\mathbf{r}}^{\dagger}\big)}, \qquad K_{\mathbf{r}} \;=\; K_{r_T}^{\theta_T}\cdots K_{r_1}^{\theta_1}, \tag{8}$$

which is again the projectivization of a single linear map. Depth, rounds and ancilla count enlarge the number of branches to $2^{n_a T}$; they do not take the map outside the class of Eq. (7).

So repeated conditional measurement is, by itself, not a route out. What *is* outside the class is re-encoding repeated over $T \geq 2$ rounds: taking the classical statistics produced by a measurement and using them to prepare a *fresh* state.

Proposition 9 (Re-encoding leaves the class, but only from the second round)**.** *Fix the amplitude encoding $E(q) = |\psi_q\rangle\langle\psi_q|$ with $|\psi_q\rangle = \sum_i \sqrt{q_i}|i\rangle$ and the computational-basis readout $R(\rho) = \mathrm{diag}(\rho)$. A re-uploading round maps a classical estimate q to $R\big(U_\theta E(q) U_\theta^{\dagger}\big)$, so, writing $u = \sqrt{q}$ componentwise, T rounds realize*

$$u \;\longmapsto\; \Big|\,U_{\theta_T}\big|\cdots|U_{\theta_1}u|\cdots\big|\,\Big|^2, \tag{9}$$

with $T-1$ componentwise absolute values, one for each intermediate re-preparation from probabilities. Then: (i) *for $T = 1$ the map is exactly of the form Eq. (7) with $K = U_{\theta_1}$, whose denominator is 1—a single round does* not *leave the class;* (ii) *for $T \geq 2$ and generic θ it does. In u the single-encoding maps of Eq. (7) are quadratic forms, hence real-analytic, while Eq. (9) is only piecewise*

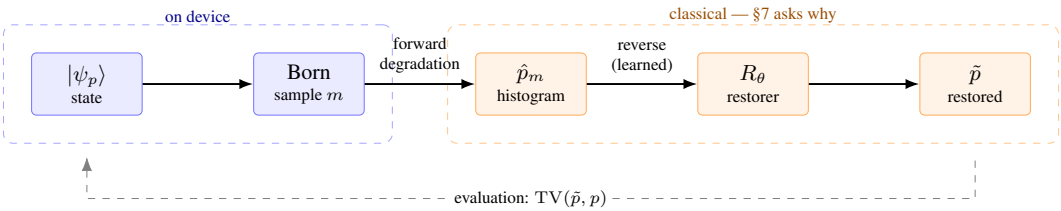


Figure 1: **Measurement as the forward process.** The corruption is not injected; it is what leaving the device costs. The shot budget m is simultaneously the diffusion time *and* the resource being spent—which is exactly what makes the resulting frontier meaningful, and what a Gaussian forward process cannot express (Corollary 6). §4 prices this knob against the standard cold-diffusion alternatives; §7 asks whether the orange block can be moved back inside the blue one, and finds that it presently cannot.

Algorithm 1: Training the reverse process. The shot budget is the diffusion time: one model serves the whole schedule, conditioned on $\log m$.

Input: data $\{p^{(j)}\}$; shot grid $\mathcal{M}$; restorer R_θ; steps T
for $t \leftarrow 1$ **to** T **do**
  sample a minibatch B and a shot budget $m \sim \mathrm{Unif}(\mathcal{M})$
  foreach $p \in B$ **do**
    $\hat{p}_m \leftarrow \frac{1}{m}\mathrm{Multinomial}(m, p)$ `// forward: Born readout`
  end
  $\mathcal{L} \leftarrow \frac{1}{|B|}\sum_{p\in B} \left\| D\left(R_\theta(\hat{p}_m, \log m) - p\right)\right\|_2^2$ `// D-scaled so pixels are` $O(1)$
  $\theta \leftarrow \mathrm{Adam}(\theta, \nabla_\theta \mathcal{L})$
end
Output: R_θ, usable at any $m \in \mathcal{M}$

so: the componentwise absolute value is non-differentiable on the hyperplanes where a component of an intermediate amplitude vector vanishes, and for generic θ these meet the interior of the simplex.

Remark (A correction to an earlier draft)**.** Part (i) corrects a claim we previously made. An earlier version asserted that a re-uploading round leaves the class "for any $T \geq 1$" on the grounds that $q \mapsto E(q)$ has entries $\sqrt{q_i q_{i'}}$ and is not affine in q. That argument compares a map on q against a class of maps on ρ and so changes the standard: Proposition 7 is stated for an input that is *already* amplitude-encoded, and the paper's own pipeline encodes $\hat{p}_m$ the same way. The nonlinearity of E is therefore common to both sides and cannot separate them. What does separate them is the phase discarded *between* rounds, which is why the statement now begins at $T = 2$.

This is what the architecture we label *measurement re-uploading* actually does: between rounds the quantum state is discarded and a new one is prepared from the current classical estimate. Our probe uses four rounds, so it is in the $T \geq 2$ regime where Proposition 9(ii) applies. The escape is therefore effected by the phase discarded between successive re-encodings, not by "a measurement raising the degree". We flag this because the latter phrasing, which an earlier draft of this paper used, is false: by Proposition 8 interleaving Born measurements coherently composes to another projective-linear map.

Positioning. This is not a state-of-the-art quantum image generator: the restoration that bends the frontier is classical, and §7 characterizes what restricts the circuit alternative rather than proving no circuit can succeed. It is a map of the terrain—the forward process, a cost calculus for it, the frontier it induces, and where the wall sits—which we take to be the analysis prior pure-quantum generative work was missing.

Table 6: **The barrier is not a scale or budget artifact.** $m = 32$ shots. "cls-matched" is a classical restorer held to *strictly fewer* parameters than the PQC in the same row (counts in parentheses); being small it is the noisiest quantity here, and an unlucky draw would flatter our conclusion, so we report mean±s.d. over 10 restarts. PQC columns are 2-seed means, or single runs marked † (6–21 hours each); the circuit's spread is far smaller, distillation repeating to within 0.003. "PQC-ovf8" is a *training* TV, every other entry a test TV; "gap" is PQC-distill minus the classical mean.

n	L	PQC par.	naive	teacher	cls-matched (par.)	PQC-ovf8 (train)	PQC-distill	gap
scaling in n at fixed depth								
8	12	216	0.486	0.214	0.266±0.021 (142)	0.287±0.008	0.384±0.002	+0.118
10	12	264	0.819	0.240	0.293±0.011 (225)	0.356±0.006	0.468±0.001	+0.175
12	12	312	0.950	0.259	0.318±0.003 (225)	0.440†	0.486†	+0.168
budget control: double the circuit at fixed n								
8	24	432	0.489	0.218	0.243±0.012 (326)	0.252†	0.373†	+0.130

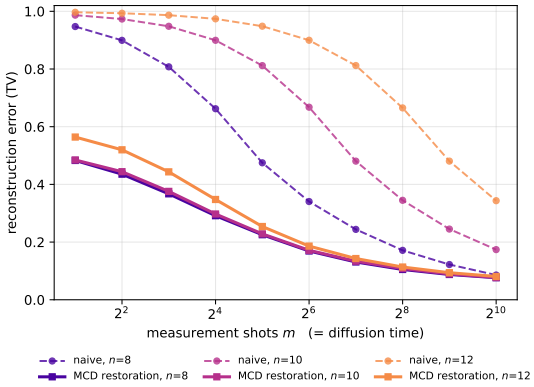


Figure 2: **The shot frontier.** Dashed: naive readout, whose curve shifts right by one octave per added qubit pair, exactly as Proposition 1 requires. Solid: MCD restoration, essentially independent of n. The horizontal gap between a dashed curve and its solid partner is the shot saving, and it compounds with n.

B Deferred Algorithms, Tables and Figures

C How Far Does the Ideal Forward Model Describe a Real Device?

Everything above treats readout as ideal Born sampling. A device does more than sample: it has state-preparation error, gate error, and readout error, none of which Eq. (1) represents. We ran the experiment that measures the gap.

Setup. We amplitude-encode six $n = 6$ distributions on `ibm_kawasaki`, a 156-qubit Heron device, and read out in the computational basis; transpiled depth is 311–335 with 83–108 two-qubit gates. Circuits run once at 8192 shots and smaller budgets come from subsampling that record, so the curve costs one submission. The restorer is trained *only* on the ideal model and never sees device data, making this the §6 transfer experiment with the device as a sixth corruption family.

Result 1: the ideal model is adequate only while shots are the binding constraint. Table 5 and Figure 4 show device error tracking the ideal model within $1.26\times$ at $m = 32$, then diverging: $1.9\times$ at $m = 128$, $3.2\times$ at $m = 512$, $13.1\times$ at $m = 8192$. The reason is visible in the shape of the curve. Ideal error falls as $m^{-1/2}$ without limit, while device error saturates near TV ≈ 0.21 by $m \approx 512$ and does not improve thereafter. That floor is state preparation and gate error, and Eq. (1) cannot express it, because in the ideal model $m \to \infty$ recovers p exactly. We therefore restrict the claims

Algorithm 2: Severity matching. Every family is placed on the *same* ladder of naive errors, so a comparison across families holds degradation level fixed and varies only degradation structure.

Input: family F with control range $[s_{\mathrm{lo}}, s_{\mathrm{hi}}]$; targets $\{\varepsilon_t\}$; held-out p
foreach *target* ε_t **do**
  $(a, b) \leftarrow (s_{\mathrm{lo}}, s_{\mathrm{hi}})$
  for 1 **to** 34 **do**
    $s \leftarrow$ midpoint of (a, b) `// geometric for` m, w`; arithmetic for` σ, r
    $e \leftarrow \mathbb{E}\,\mathrm{TV}\big(F_s(p),\, p\big)$ estimated on p
    if *e exceeds ε_t in the direction the knob corrects* **then** $a \leftarrow s$
    else $b \leftarrow s$
  end
  $s_F(\varepsilon_t) \leftarrow$ midpoint of (a, b)
end
Output: control ladder $s_F(\cdot)$ reproducing the common error ladder

Table 7: **The shot budget collapses onto** $\alpha = m/D$, as Proposition 1 requires. Each row is a rung of the degradation ladder; entries are the α at which that rung is reached, mean over 3 seeds. The last column is the spread across n. Rows marked ‡ are the two most degraded rungs, excluded from the analyses in §6; they are also where the Poisson limit is entered most slowly.

target TV	$0.95^{\ddagger}$	$0.88^{\ddagger}$	0.78	0.66	0.54	0.43	0.33	0.24	0.17
α at n=8	0.0059	0.0176	0.0371	0.0642	0.0972	0.1525	0.2680	0.5140	1.026
α at n=10	0.0073	0.0190	0.0376	0.0640	0.0994	0.1563	0.2732	0.5215	1.047
α at n=12	0.0076	0.0189	0.0375	0.0640	0.0998	0.1560	0.2723	0.5209	1.047
spread	29.6%	8.3%	1.3%	0.3%	2.7%	2.5%	1.9%	1.5%	2.1%

of §§4–6 to the shot-limited regime and say so: the resource calculus prices the sampling cost, not the total error of a device.

Result 2: the learned prior transfers to hardware anyway. Trained on simulated ideal readout and never shown this device, the restorer still reduces device error at every budget—12% at $m = 32$, 43% at $m = 512$, 51% at the floor, where it takes 0.213 to 0.104. This is the one part of the pipeline that validates on hardware without qualification, and it sharpens what the floor is: a floor a manifold prior can halve is not an information-theoretic limit on the device but structured error a prior partially removes.

D Probe Details

Settings. Unless stated, $n = 8$, $m = 32$, $n_a = 1$, $L = 12$, Adam at $4 \cdot 10^{-3}$. *Ancillas* sweeps $n_a \in \{1, 2, 3\}$ at $n = 6$, $L = 12$, where the naive error is 0.248; restored error is 0.239 at every n_a. *Parameters* enlarges the ansatz to n_a=2, L=16 with per-m output heads and a TV loss, which moves m=16 from 0.339 to 0.329 and m=64 from 0.178 to 0.173. *Topology* replaces the ansatz with a quantum U-Net whose pooling stage discards the detail restoration needs. *Flow matching* replaces diffusion with a measurement-feedback PQC velocity field: the regression loss falls $1.25 \to 0.84$ but generation is essentially Haar (MMD = 0.25 against a Haar baseline of 0.145), while an identical setup with a classical velocity field reaches 0.010.

Measurement re-uploading. Four rounds, $L = 10$, 652 circuit parameters. Each round forms a classical mixture of the running estimate and the original degraded input, re-encodes it as $|\psi_q\rangle = \sum_i \sqrt{q_i}|i\rangle$, applies a unitary and reads out; the classical control substitutes a convolutional unit of 700 parameters for the circuit at each round. This is the construction of Proposition 9.

Generation, and a faster scheme that does not rescue it. Because Eq. (1) has a closed form at every m, restoration targets can be matched *slice-wise* at independently sampled budgets rather than

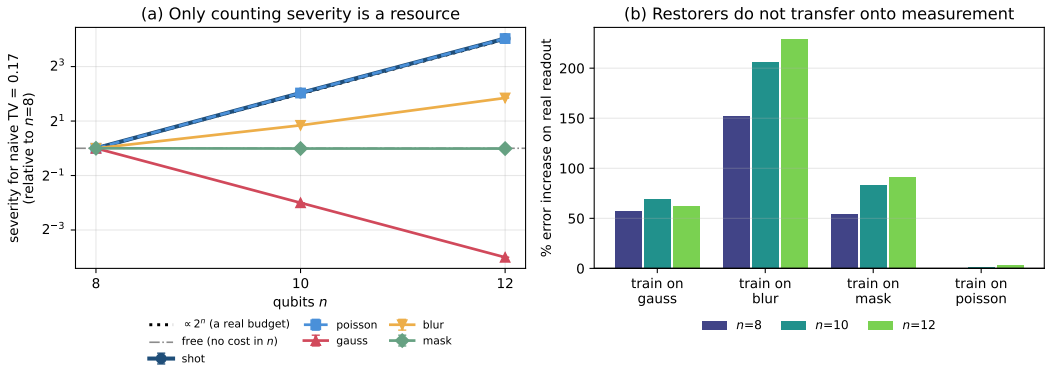


Figure 3: **(a)** Severity required to hold the naive error at TV $= 0.17$, normalized to its n=8 value. Counting families track 2^n exactly, the mask rate is exactly flat, and the Gaussian scale *falls* as 2^{-n}—the three behaviours predicted by Propositions 1–4. **(b)** Error increase when a restorer trained on one corruption is applied to real finite-shot readout, relative to one trained on readout itself. Poisson is free; Gaussian and blur are not.

Table 8: **Cross-family transfer at** $n = 12$, matched to naive TV $= 0.33$, mean over 3 seeds. Rows: training corruption. Columns: test corruption. Diagonals are strong for every family, consistent with Bansal et al. (2023); the off-diagonals show the families are nonetheless not interchangeable.

train \ test	shot	poisson	gauss	blur	mask
shot	**0.069**	0.069	0.095	0.354	0.059
poisson	0.072	**0.071**	0.087	0.351	0.061
gauss	0.112	0.112	**0.052**	0.287	0.106
blur	0.228	0.228	0.149	**0.112**	0.208
mask	0.132	0.132	0.208	0.330	**0.031**

by rolling the backward circuit through all T steps. This replaces an $O(T^2)$ rollout with an $O(T)$ one and runs $3.5\times$ faster (73s versus 257s at matched T) at a small cost in natural distance (0.150 versus 0.131). It is a real property of this forward process, but it does not change the conclusion: with either scheme the pure-quantum generator produces sparse blobs rather than digits. A cheaper route through the same function class reaches the same ceiling.

E Discussion and Limitations

What we claim. For pipelines that reconstruct a full 2^n-outcome Born distribution from finite-shot readout, measurement is the right cold-diffusion forward process: it is the only corruption family we tested whose control parameter is a resource, and that resource provably costs $\Theta(2^n)$ (Proposition 1, verified over a $256\times$ range in D). A learned prior bends that frontier, and on fixed-complexity data flattens it. The prior is not realized by any shallow circuit we tested, and we characterize the class such a circuit lives in (Proposition 7).

What we do not claim. We do not demonstrate a quantum advantage: the model that bends the frontier is classical, and in this pipeline the device supplies the forward process while the classical model supplies the prior. We do not claim measurement is the right forward process for quantum generative modeling in general, since many models emit samples and never reconstruct a distribution. And we do not prove that a circuit cannot restore—Remark 7.1 states exactly what is characterization and what is conjecture.

Limitations. Each is measured rather than asserted; §E.1 elaborates.

Table 9: **Six probes, one floor.** Each removes a different escape route. Only the last leaves the projective-linear class of Proposition 7, and only the last improves materially. Details and settings in Appendix D.

probe	what it varies	result	baseline
ancillas	$n_a \in \{1, 2, 3\}$	flat at 0.239	naive 0.248
parameters	n_a=2, L=16, per-m heads	$\approx 3\%$ gain	—
topology	quantum U-Net	worse (pooling loses detail)	—
flow matching	PQC velocity field	MMD $= 0.25$	Haar 0.145, classical 0.010
memorization	fit 8 pairs, test on them	train TV 0.287	naive 0.489
distillation	teacher outputs as targets	0.384	teacher 0.218
re-uploading	*re-encodes* each round	**0.337**	naive 0.489, classical 0.241

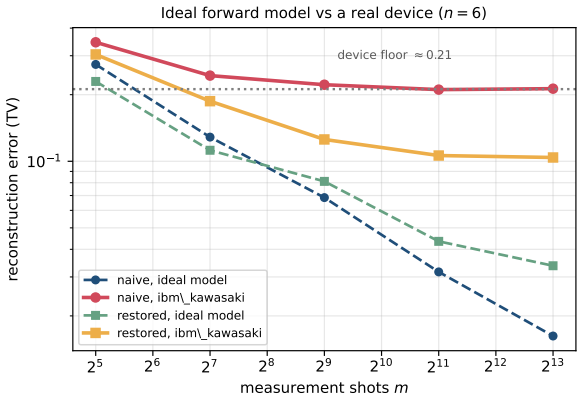


Figure 4: **Where the ideal model stops describing the device.** Ideal error (dashed) falls without limit; device error (solid) saturates at a floor that Eq. (1) cannot represent. The restorer, trained only on the ideal model, nonetheless halves device error at the floor.

- **The forward model is idealized.** It omits device error, and §C shows it stops describing `ibm_kawasaki` once shots stop binding, diverging to 13×. The calculus prices sampling cost alone.
- **The flattening is a data-prior statement.** It holds for fixed-complexity data and fails for a synthetic family whose degrees of freedom scale with D (1.64×, §5.1) and for Ising ground-state Born distributions (2.12×, §5.1). The $\Theta(2^n)$ naive law is unaffected.
- **The savings are amortized.** They are inference-time readout costs and exclude the one-time cost of acquiring the prior (§5).
- **Scale.** The calculus reaches $n = 16$; the ablations and all circuit experiments are at $n \leq 12$, and the device experiment is one $n = 6$ job on one backend.
- **The circuit result is empirical.** It concerns shallow ansatze inside the projective-linear class; a construction that leaves it is not excluded.

E.1 Extended discussion

Scope of the forward model. Eq. (1) is ideal finite-shot computational-basis readout. §C measures the gap directly: device error saturates at TV ≈ 0.21 while the ideal model continues to fall as $m^{-1/2}$, so the discrepancy grows from 1.26× at $m = 32$ to 13.1× at $m = 8192$. The resource calculus of §4 should therefore be read as pricing the sampling cost of readout, not the total error of a device. The one part that does validate on hardware without qualification is the prior itself, which halves device error at the floor despite never having seen device data.

What the flat frontier does and does not say. The flatness in n is a property of data whose intrinsic complexity does not grow with grid resolution. We establish this by exhibiting cases on

both sides: it holds for MNIST ($1.05\times$), Fashion-MNIST ($1.20\times$) and a fixed-K synthetic family ($1.03\times$), and fails for a synthetic family whose degrees of freedom scale with D ($1.64\times$) and for ground-state Born distributions of a disordered Ising chain ($2.12\times$ with a convolutional restorer, $2.85\times$ with a permutation-agnostic MLP). We do not extrapolate the flat result to data classes that grow harder with n. The $\Theta(2^n)$ law for naive readout is untouched by this, since Proposition 1 depends only on D_{eff}/D; but note that even that hypothesis is a property of the data family, and the Ising ground states do not satisfy it (D_{eff}/D falls from 0.356 to 0.218), so the proposition's conclusion is not directly available there.

What the open problem is. The characterization in §7 makes the open problem concrete rather than open-ended. It is not "find a better ansatz". It is: restoration appears to require functionals of the input state of degree $k \gg 1$; degree k costs $\Omega(k)$ copies under sample access and $\Theta(\sqrt{k})$ under purified access (Zhang et al., 2026); so the question is whether the degree restoration actually needs can be bought at a price that still leaves the pipeline cheaper than reading the state out and restoring it classically. Our re-uploading probe is one point on that curve and it is not encouraging, but it is a single point, and the purified-access branch—where the price is quadratically lower—is untested here.

F PROOFS

Proof of Proposition 1. Let $N_i = m\hat{p}_{m,i} \sim \mathrm{Bin}(m, p_i)$ marginally, so that

$$\mathbb{E}\big[\mathrm{TV}(\hat{p}_m, p)\big] \;=\; \frac{1}{2m}\sum_{i=1}^{D}\mathbb{E}\big|N_i - mp_i\big|. \tag{10}$$

Put $\lambda_i = mp_i = \alpha c_i$. We compare each summand with its Poisson counterpart. The test function $f(x) = |x - \lambda_i|$ is 1-Lipschitz, so

$$\big|\mathbb{E}|N_i - \lambda_i| - g(\lambda_i)\big| \;\le\; W_1\big(\mathrm{Bin}(m, p_i),\, \mathrm{Pois}(\lambda_i)\big), \tag{11}$$

which avoids having to upgrade a total-variation bound to an unbounded test function. Wasserstein-1 is subadditive over independent sums, and for a single trial $W_1(\mathrm{Bern}(p), \mathrm{Pois}(p)) \le Cp^2$; writing $\mathrm{Bin}(m, p_i)$ and $\mathrm{Pois}(\lambda_i)$ as sums of m i.i.d. terms gives $W_1\big(\mathrm{Bin}(m, p_i), \mathrm{Pois}(\lambda_i)\big) \le Cmp_i^2 = C\alpha^2 c_i^2/m$. Substituting into Eq. (10), the total discrepancy is

$$\frac{1}{2m}\sum_{i=1}^{D}\frac{C\alpha^2 c_i^2}{m} \;=\; \frac{C\alpha^2}{2m^2}\sum_{i=1}^{D}c_i^2 \;=\; \frac{C}{2D}\cdot\frac{1}{D}\sum_{i=1}^{D}c_i^2 \;=\; O(1/D)$$

using $m = \alpha D$ and the uniform second-moment bound Eq. (3). No moment assumption beyond Eq. (3) is used, and in particular no third moments of either the limit law or the triangular array are required. Therefore

$$\mathbb{E}\big[\mathrm{TV}(\hat{p}_m, p)\big] \;=\; \frac{1}{2\alpha D}\sum_{i=1}^{D} g(\alpha c_i) + O(1/D) \;\xrightarrow[D\to\infty]{}\; \frac{1}{2\alpha}\int g(\alpha c)\, d\nu(c),$$

the limit by weak convergence of the empirical measure of $\{c_i\}$ together with continuity of g and the bound $g(\lambda) \le \sqrt{\lambda}$, which follows from $\mathbb{E}|X - \lambda| \le \sqrt{\mathrm{Var}\, X}$ for $X \sim \mathrm{Pois}(\lambda)$; uniform integrability of $\sqrt{\alpha c_i}$ is then immediate from Eq. (3). The closed form for g is de Moivre's expression for the mean absolute deviation of a Poisson law. Monotonicity and the endpoint values of Φ follow from $g(\lambda)/\lambda$ being strictly decreasing with $g(\lambda)/\lambda \to 2$ as $\lambda \to 0$ and $g(\lambda) \sim \sqrt{2\lambda/\pi}$ as $\lambda \to \infty$; uniqueness of $\alpha(\varepsilon)$ and hence $m(\varepsilon) = \alpha(\varepsilon)D = \Theta(2^n)$ follow. Note that no per-coordinate central limit theorem was used, so the argument is valid precisely in the regime $mp_i = \Theta(1)$ that the paper operates in. □

Proof of Corollary 2. If $\min_i \lambda_i \to \infty$ then $g(\lambda_i) = (1 + o(1))\sqrt{2\lambda_i/\pi}$ uniformly, so Eq. (10) becomes $\frac{1+o(1)}{2m}\sum_i \sqrt{2mp_i/\pi} = \frac{1+o(1)}{\sqrt{2\pi m}}\sum_i \sqrt{p_i}$. Squaring the sum gives $D_{\text{eff}}(p)$ and solving at $\mathbb{E}[\mathrm{TV}] = \varepsilon$ gives the stated budget. The same conclusion follows from a per-coordinate central limit theorem, which is legitimate here precisely because $\min_i mp_i \to \infty$ has been assumed. □

Proof of Proposition 3. Condition on the surviving set S and write $\sigma_S = \sum_{i \in S} p_i \le 1$. The renormalized vector is $q_i = p_i/\sigma_S$ for $i \in S$ and 0 otherwise. Then

$$\mathrm{TV}(q,p) = \tfrac{1}{2}\Big[\sum_{i \notin S} p_i + \sum_{i \in S} p_i\big(\tfrac{1}{\sigma_S} - 1\big)\Big] = \tfrac{1}{2}\Big[(1-\sigma_S) + \sigma_S \cdot \tfrac{1-\sigma_S}{\sigma_S}\Big] = 1 - \sigma_S,$$

where the second term is non-negative because $\sigma_S \le 1$. Under *unconditioned* deletions, $\mathbb{E}[\sigma_S] = (1-r)\sum_i p_i = 1 - r$ and therefore $\mathbb{E}[\mathrm{TV}(q,p)] = r$ exactly, with no dependence on D or on p. The proposition conditions on $S \neq \emptyset$, since the empty mask leaves q undefined, and conditioning shifts this by the mass of that event: writing $\pi = r^D$ for its probability and noting $\mathrm{TV} = 1$ there,

$$\mathbb{E}\big[\mathrm{TV} \mid S \neq \emptyset\big] = \frac{\mathbb{E}[\mathrm{TV}] - \pi \cdot 1}{1 - \pi} = \frac{r - r^D}{1 - r^D} \;=\; r + O(r^D),$$

which is the form stated. For $r \le 0.95$ and $D \ge 256$ the correction is below 10^{-5}, so the two statements are numerically indistinguishable here; we keep the conditioned one because it is the quantity the implementation actually samples. □

Proof sketch of Proposition 4. Before clipping, $\|(p+\sigma\varepsilon)-p\|_1 = \sigma\|\varepsilon\|_1$ with $\mathbb{E}\|\varepsilon\|_1 = D\sqrt{2/\pi}$, so the unclipped total variation is $\frac{1}{2}\sigma D\sqrt{2/\pi} = \Theta(\sigma D)$. Clipping at zero and renormalizing removes a constant fraction of the perturbation when $\sigma = \Theta(1/D)$ and $p_i = \Theta(1/D)$, changing the constant but not the order. Solving $\Theta(\sigma D) = \varepsilon$ gives $\sigma = \Theta(\varepsilon/D)$. Empirically the measured σ exceeds the unclipped prediction by a factor ≈ 1.4, constant in n, consistent with clipping affecting only the constant. □

Proof of Proposition 7. The embedding ι, the conjugation $\rho \mapsto U_\theta \rho U_\theta^\dagger$ and the projection onto ancilla outcome r are each linear in ρ, and their composition is the map $\rho \mapsto K_r^\theta \rho K_r^{\theta\dagger}$ with $K_r^\theta = (\langle r| \otimes I)U_\theta \iota$; the Born rule gives $\Pr[r] = \mathrm{Tr}(K_r^\theta \rho K_r^{\theta\dagger})$, and renormalization divides by exactly that scalar, yielding Eq. (7). Claim (i) is the observation that the numerator is this linear map and the denominator is its trace, so Φ_r is the projectivization of a linear map; the numerator's degree in ρ is 1 for any θ and any depth, since a product of unitaries is a unitary. Claim (ii) is immediate as r ranges over 2^{n_a} outcomes. For (iii), suppose $\Phi_r(|\psi\rangle\langle\psi|) = \Phi_r(|\varphi\rangle\langle\varphi|)$. Then $K_r^\theta|\psi\rangle$ and $K_r^\theta|\varphi\rangle$ span the same ray, so $K_r^\theta(|\psi\rangle - c|\varphi\rangle) = 0$ for some scalar $c \neq 0$; invertibility of K_r^θ forces $|\psi\rangle = c|\varphi\rangle$, i.e. the inputs coincide as states. Genericity of invertibility: K_r^θ is a $2^n \times 2^n$ block of a Haar- generic unitary, hence invertible outside a measure-zero set of θ. □

Proof of Proposition 8. By Proposition 7 one round conditional on outcome r_1 sends $\hat\rho$ to $K_{r_1}^{\theta_1}\hat\rho K_{r_1}^{\theta_1\dagger}$ up to normalization. Feeding this forward coherently, round t applies $K_{r_t}^{\theta_t}$ to the current unnormalized operator. After T rounds the unnormalized operator is $K_{\mathbf{r}}\hat\rho K_{\mathbf{r}}^\dagger$ with $K_{\mathbf{r}} = K_{r_T}^{\theta_T}\cdots K_{r_1}^{\theta_1}$, and the intermediate normalizations cancel because each divides by the trace of the operator it is applied to; the final normalization is by $\mathrm{Tr}(K_{\mathbf{r}}\hat\rho K_{\mathbf{r}}^\dagger)$, which is also the probability of the trajectory $\mathbf{r}$. Hence Eq. (8), which has the form Eq. (7) with K replaced by $K_{\mathbf{r}}$. The trajectory ranges over $2^{n_a T}$ values. □

Proof of Proposition 9. One round sends $u \mapsto |U_\theta u|$ in amplitudes and reports $|U_\theta u|^2$; iterating and discarding phase between rounds gives Eq. (9). For (i), take $K = U_{\theta_1}$ in Eq. (7): since U_{θ_1} is unitary, $\mathrm{Tr}(KE(q)K^\dagger) = 1$ and $R(KE(q)K^\dagger) = |U_{\theta_1}u|^2$, which is the $T = 1$ case verbatim. For (ii), a map of the form Eq. (7) reads, in u, as $u \mapsto |Ku|^2/\|Ku\|^2$, a ratio of quadratic forms and hence real-analytic on the interior of the positive orthant. Now let $T = 2$ and consider u crossing a hyperplane $\mathcal{H}_j = \{(U_{\theta_1}u)_j = 0\}$. On $\mathcal{H}_j$ the map $u \mapsto |(U_{\theta_1}u)_j|$ has a kink, and it propagates to the output coordinate k whenever $(U_{\theta_2})_{kj} \neq 0$, because the remaining terms of $\sum_{j'}(U_{\theta_2})_{kj'}|(U_{\theta_1}u)_{j'}|$ do not vanish there in general. The set of θ for which every such coefficient vanishes is a proper algebraic subvariety, so for generic θ the composite is not real-analytic on $\mathcal{H}_j \cap \Delta^{D-1}$ and therefore is not of the form Eq. (7). The same obstruction applies for any $T \ge 2$. Numerically, along a line crossing $\mathcal{H}_j$ at $D = 2$ with $U_{\theta_1} = U_{\theta_2} = H$, the discrete second difference of the composite is $9.0 \cdot 10^{-4}$ against $4.4 \cdot 10^{-16}$ for the single-encoding map, a ratio of $2 \cdot 10^{12}$. □

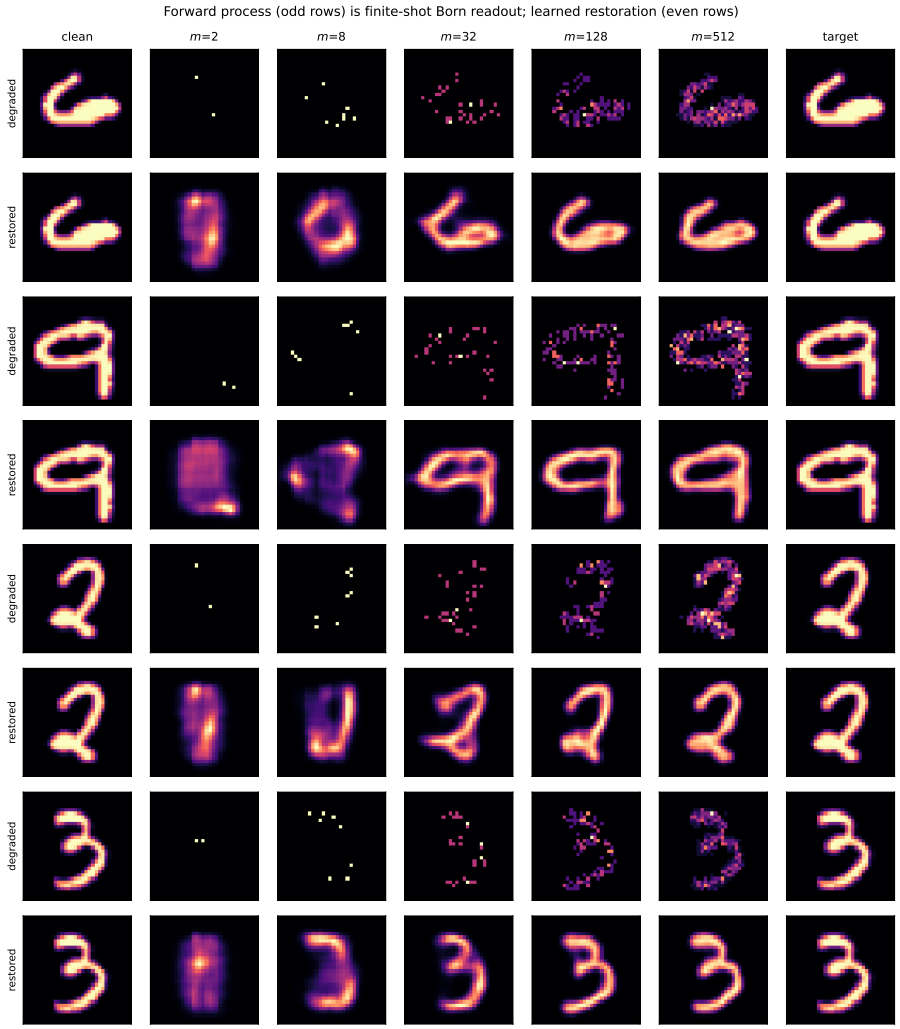


Figure 5: **Forward process and learned restoration,** $n = 10$**.** Odd rows: finite-shot Born readout at increasing budget—this is the entire forward corruption, with nothing injected. Even rows: the learned restoration conditioned on $\log m$. Leftmost and rightmost columns repeat the clean target for reference.

G What the Pipeline Actually Produces

The main text reports scalars. Because our central claims about the circuit restorer are negative, it matters that a reader can see what is and is not being produced.

Figure 5 shows the forward process and its inverse at $n = 10$. At $m = 2$ the readout is two lit pixels; the trajectory fills in as the budget grows, and the learned restoration recovers a recognizable digit from $m = 32$ onward. This is the qualitative content of Table 2.

Figure 6 is the one we consider most informative, and it is the qualitative content of the memorization probe. The circuit is trained to memorize these eight specific images—it is shown nothing else, and is evaluated on its training set. Its output is visibly a partially denoised copy of the degraded input rather than a digit: strokes remain fragmented and the background speckle is only attenuated. The classical teacher in the last row, which is solving the harder problem of generalizing, returns clean digits. Total variations for this panel are 0.42 (degraded), 0.27 (circuit) and 0.14 (classical).

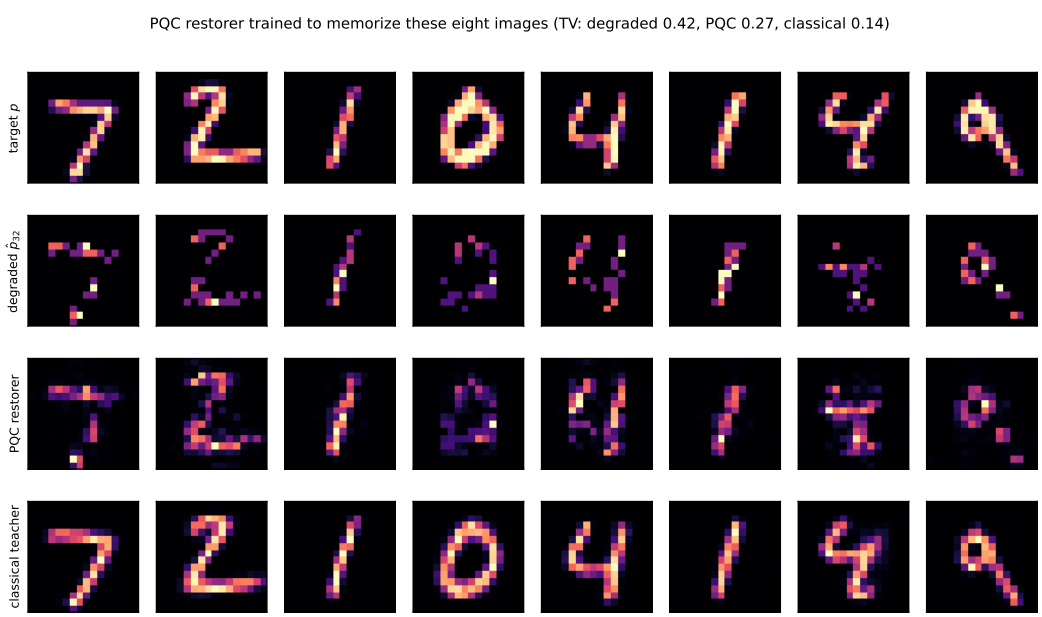


Figure 6: **What the circuit restorer outputs when asked to memorize eight images** ($n = 8$, $m = 32$, $L = 12$, $n_a = 1$). Row 3 is the circuit's output on its own training set after 1400 steps. It attenuates the readout speckle but does not reconstruct the digits; the classical restorer in row 4 does. This is the same failure that Table 6 reports as a number.

H Experimental Details

Data. MNIST test-split images, bilinearly resampled to $\sqrt{D} \times \sqrt{D}$ with $D = 2^n$ (16×16, 32×32, 64×64 for $n = 8, 10, 12$), clipped at zero, offset by 10^{-6} and normalized to lie in Δ^{D-1}. §5 and §6 use 1800 train and 300 test images; §7 uses 512 train and 256 test, the smaller split being forced by circuit simulation cost.

Classical restorer. A small U-Net: two downsampling stages with 64 and 128 channels, GroupNorm and GELU, transposed-convolution upsampling with skip connections, and a softplus output renormalized to the simplex. Inputs are scaled by D so pixel values are $O(1)$, which matters because raw probabilities are $\sim 10^{-4}$ and otherwise vanish in the gradient. Conditioning on $\log_{10} m$ enters as a constant channel. Adam, learning rate 10^{-3}, batch 128, 1500 steps, 182,273 parameters.

Parameter-matched classical control. A three-layer convolutional stack with $9c^2 + 20c + 1$ parameters, c chosen as the largest integer keeping the count *below* the PQC budget of the same row; Adam, learning rate $3 \cdot 10^{-3}$, batch 128, 800 steps, TV loss. Reported as mean$\pm$s.d. over 10 restarts with distinct seeds.

PQC restorer. The QDDPM backward circuit (Zhang et al., 2024): a hardware-efficient ansatz on $n_{\text{tot}} = n + n_a$ qubits with L layers of per-qubit R_x, R_y and a brickwork of CZ gates, $2n_{\text{tot}}L$ parameters, followed by measurement of the n_a ancillas and renormalization of the surviving block. The degraded histogram enters as amplitudes $\sqrt{\hat{p}_m}$ padded into the $2^{n_{\text{tot}}}$-dimensional register. Unless stated, $n_a = 1$. Adam, learning rate $4 \cdot 10^{-3}$; memorization uses batch 8 for 1400 steps, distillation batch 16–48 for 800 steps. Test errors average three measurement realizations.

Severity matching. Algorithm 2 with 34 bisection iterations, geometric midpoints for m and w and arithmetic for σ and r, evaluated on the held-out split with two independent corruption draws per evaluation. Target ladder $\{0.95, 0.88, 0.78, 0.66, 0.54, 0.43, 0.33, 0.24, 0.17, 0.12\}$; the top two levels are excluded for the Gaussian and blur families as described in §6. Transfer is evaluated at level index 6 (TV $= 0.33$), where all families match to better than 0.01.

Intrinsic-complexity families (§5.1). *blobs-fixed* and *blobs-grow* images are $\sum_{k\le K} a_k \exp(-\|x-\mu_k\|^2/2w_k^2)$ on the $\sqrt{D}\times\sqrt{D}$ grid with μ_k uniform on the square, $a_k \sim \mathrm{Unif}[0.4, 1.4]$ and $w_k = (0.16/\sqrt{K/8})\cdot \mathrm{Unif}[0.6, 1.4]$, the width shrinking with K so that components stay resolvable and total ink is comparable across families. $K = 8$ for *blobs-fixed* and $K = \max(8, D/32)$ for *blobs-grow*, giving $K = 8, 32, 128$ at $n = 8, 10, 12$. Restorer, optimizer, shot grid and train/test split are identical to §5; the only difference across the three families is the data.

Additional families (§5.1). Fashion-MNIST is resampled exactly as MNIST is. Ising ground states are obtained by sparse Lanczos (`scipy.sparse.linalg.eigsh`, $k = 1$, smallest algebraic, tolerance 10^{-7}) on the Hamiltonian equation 6 with periodic boundary conditions and a fresh disorder realization per sample; $p_x = |\langle x|\psi_0\rangle|^2$ with ψ_0 real. We use 600 train and 200 test states, fewer than for images because each requires a diagonalization. The MLP restorer is $D{+}1 \to 512 \to 512 \to D$ with GELU and a softplus output renormalized to the simplex. Restorer, optimizer, shot grid and training budget are otherwise identical across all families.

Extended range (Table 1). At $n = 14$ and 16 the training set is reduced to 900 and 400 images and the batch to 48 and 16 to fit activations for a 128×128 and 256×256 restorer; the channel width is 48 throughout this experiment. The shot grid is defined by $\alpha = m/D \in \{2^{-6}, \ldots, 2^2\}$ rather than by absolute shot counts, so every n sees the same range of α; multinomial draws are chunked, since $\alpha = 4$ at $n = 16$ requires 262144 shots per image. Budgets are located by 22 bisection steps on $\log\alpha$ with two corruption draws per evaluation.

Compute. Corruption sampling and all classical training run on a single RTX A6000; circuit simulation is CPU-bound through a `tensorcircuit` PyTorch backend and does not benefit from the GPU (measured speedup $1.3\times$, the cost being per-gate kernel launches rather than arithmetic). Individual PQC probes take between 8 minutes and 21 hours depending on n and machine load; wall-clock numbers are therefore not comparable across runs and are not used in any claim.

Reproducibility. §6 runs 3 seeds throughout. §7 runs 10 restarts for the classical control and 2 seeds for the PQC at $n = 8, 10$ with $L = 12$; the $n = 12$ and $L = 24$ configurations are single runs and are marked † in Table 6.

AI USE STATEMENT

The authors take full responsibility for the final content of this paper, including any text produced with the assistance of generative AI.