QUERY COMPLEXITY OF FORWARD-ONLY HEBBIAN LEARNING: THEORY AND A NATIVE FIT FOR DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study the query complexity of forward-only, gradient-free learning rules for training neural networks, motivated by settings where backpropagation is structurally unavailable: activation-tape memory that grows with depth and context, and objectives (quantized weights, non-differentiable rewards) whose exact gradient is zero almost everywhere. We analyze a three-factor Hebbian rule – an antithetic zeroth-order gradient estimator that perturbs parameters along a random probe and broadcasts a single scalar score to every synapse – and derive an exact, non-asymptotic identity for its gradient alignment as a function of probe count and problem dimension, yielding a closed-form query budget $M^* \approx (P{+}1)/3$. We relate this upper bound to known minimax lower bounds for zeroth-order stochastic convex optimization, discussing how close the rule's query complexity comes to the best possible rate for any gradient-free method under a two-point oracle. We further identify an architecture-dependent separability criterion – a measurable coupling scope S and feedback bandwidth K – that predicts when this query cost stays small, and confirm it empirically on a diffusion denoiser, where the per-noise-level residual gives $S{=}1$: the rule closes most of the gap to backpropagation from scratch while eliminating its activation-memory overhead entirely, and continues to descend the loss through a quantized codebook *argmin* where the exact gradient is identically zero.

1 INTRODUCTION

Diffusion models are increasingly *adapted* after pretraining (personalized to a subject or style, repaired on a single failing prompt, or aligned to a reward, verifier, or human preference) and increasingly this adaptation happens at deployment: per instance, inside the serving loop, with no offline training run. Yet the default tool for all of it, backpropagation, is frequently the wrong instrument there. Its cost is structural, not incidental: a single global backward pass, an *activation tape* whose size grows with depth and context, and an end-to-end differentiable objective. At inference time these stop being merely expensive and become disqualifying all at once: there is no tape budget, the reward is often non-differentiable, and a hard latency ceiling leaves no room for a backward pass.

Two of these prices are worth making concrete, because they are exactly the walls any replacement must clear. *Memory:* on our convolutional denoiser the activation tape costs ≈ 262 KB per layer, so at depth 128 reverse-mode training reaches $43.8\times$ the model's inference footprint (measured in the released memory-scaling notebook), the gap a tape-free rule closes outright, since a rule that only *evaluates* the loss stores nothing for a backward pass. *Differentiability:* on a quantized or reward-graded objective the gradient is identically zero, and reverse-mode returns the zero vector exactly where a descent direction is most needed. The obvious escape, estimating the gradient by random perturbation, only trades the memory wall for a *query-complexity* wall: resolving a d-dimensional gradient by isotropic probing costs $O(d)$ loss evaluations per step (Ghadimi & Lan, 2013; Nesterov & Spokoiny, 2017), hopeless at model scale.

This paper asks what the minimal replacement must look like: a local, forward-only update that stores no tape, needs no gradient (Malladi et al., 2023), and still delivers an *unbiased* estimate of the loss gradient at a query cost we can bound *in advance*. The answer is a *three-factor Hebbian* rule. Perturb

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.

the parameters θ along a random probe ξ, evaluate the loss at $\theta \pm \sigma\xi$, and broadcast one scalar score $s = \big(L(\theta + \sigma\xi) - L(\theta - \sigma\xi)\big)/2\sigma$ to every synapse, giving $\Delta\theta = \eta\, \xi s$ (eq. (1)). Component-wise this is exactly three-factor and Hebbian-like plasticity (Fremaux & Gerstner, 2016), a pre-$\times$-post probe direction times one global modulator, and it is precisely the antithetic zeroth-order estimator (SPSA in stochastic approximation (Spall, 1992), evolution strategies in optimization (Salimans et al., 2017)): unbiased for the Gaussian-smoothed loss (Nesterov & Spokoiny, 2017), and well-defined even where $\nabla L = 0$.

Why diffusion. Backpropagation's costs bite unevenly across a model's life cycle, lightest at pretraining, heaviest at *inference-time adaptation*, where a single instance, a few forward passes of latency, and a possibly non-differentiable reward are the entire budget. Diffusion turns this constraint into an opportunity: its per-pixel, per-noise-level residual (Ho et al., 2020; Song et al., 2021) is spectrally diagonal (Hoogeboom & Salimans, 2023; Rissanen et al., 2023), so each mode is its own feedback channel and the coupling scope one scalar must summarize collapses to $S{=}1$; with timestep as a second axis of channels, non-autoregressive diffusion is where forward-only adaptation is a native fit, not a fallback (section 5).

Three deployment patterns instantiate it. Personalizing a pretrained model to the current input stream, repairing a single failing generation without retraining, and aligning to a non-differentiable metric, verifier, or human signal (Black et al., 2024; Shao et al., 2024; Wallace et al., 2024) are each a setting the rule's properties (tape-free, gradient-free, bounded query budget) are built for.

A budget you can price in advance. What makes the rule usable at deployment is that its cost is not merely bounded but *known*. A single Gaussian fourth-moment (Isserlis (Isserlis, 1918)) identity gives the exact alignment of an M-probe estimate, $\cos^2(\hat{g}_M, g) = M/(M + D + 1)$, and inverting it yields a probe budget M^* that splits into two measured scaling laws: a *barrier law* $M^*K \gtrsim S$ in coupling scope S and feedback bandwidth K (eq. (5)), and an *alignment law* linear in the trainable-parameter count. Because S and K are fixed by architecture and task *before a single probe is drawn*, the forward-pass cost of an adaptation can be priced ahead of running it, a knob no backpropagation-based method offers.

Contributions. We make three. First, a *predictive adaptation budget*: the three-factor update is an unbiased estimator of the loss gradient (section 2) whose variance is fixed by two measured scaling laws, in coupling scope S and feedback bandwidth K, giving the probe budget of eq. (5) that prices an inference-time adaptation *before* it is run. Second, a *separability criterion*: an architecture-only test (coupling density $\approx 0 \Rightarrow$ per-channel feedback $K{=}C$) predicting where a local rule stays cheap and where it hits the $O(d)$ query wall, and identifying diffusion's spectral diagonalization ($S{=}1$) as the best case. Third, *diffusion is the native fit, measured*: on a convolutional denoiser at matched $P \approx 1$M and equal wall-clock the forward-only rule closes most of the gap to backpropagation from scratch (binarized-MNIST reconstruction to 0.984 pixel-accuracy, CIFAR-10 to 19.3 dB and CelebA to 18.4 dB PSNR), descends a loss even through a quantized codebook *argmin* where the exact gradient is identically zero, and is consistent with the alignment law at the one small scale where it can be measured directly (a measured slope of 51.2 against a predicted 50, section H.1). A short MLP ruler and an attention case, which place diffusion between the best and worst architectures, are deferred to the appendix.

Related work. Zeroth-order and evolutionary gradient estimation is classical (Ghadimi & Lan, 2013; Nesterov & Spokoiny, 2017; Salimans et al., 2017; Spall, 1992) and recently revived for memory-free fine-tuning of large models (Malladi et al., 2023); our contribution is not the estimator but its *predictable* budget, two measured scaling laws and an architecture-only separability criterion for where it stays cheap. Minimax query complexity for the same two-point oracle model is studied directly (Duchi et al., 2015; Shamir, 2013), whose lower bounds we relate our upper bound to in section 3.4. The third factor connects to neuromodulated three-factor plasticity (Fremaux & Gerstner, 2016) and to the target-blindness of classical Hebbian rules (Oja, 1982; Sanger, 1989); the applications build on low-rank adaptation (Aghajanyan et al., 2021; Hu et al., 2022), reward-aligned diffusion (Black et al., 2024; Wallace et al., 2024), and the score/diffusion foundations (Efron, 2011; Ho et al., 2020; Song et al., 2021; Vincent, 2011) whose spectral structure makes the convolutional denoiser the native fit. A comprehensive treatment is deferred to the appendix (not duplicated here).

2 A HEBBIAN LOCALIZED LEARNING APPROACH

This section builds the rule from the two-factor baseline up: why the classical rule cannot descend a loss, the single scalar that fixes it, why the resulting estimator is unbiased, how antithetic probing keeps it exact and tape-free, and why it extends past differentiability.

Notation. Throughout, $\theta \in \mathbb{R}^P$ are the probed parameters, $L(\theta)$ the scalar loss, $g = \nabla L$ the true gradient, $\xi \sim \mathcal{N}(0, I)$ a Gaussian probe at scale σ, and $\hat{g}_M$ the average of M antithetic single-probe estimates. The alignment $\cos^2(\hat{g}_M, g)$ is the fraction of the estimate's energy along g; M^* is the probe budget that reaches the target alignment; S is the coupling scope (entangled residual terms one scalar must summarize) and K the number of independent feedback channels (table 6 in section G collects the symbols). Every hypothesis used by a result in this paper – probe distribution, differentiability order, and the modeling assumptions behind the barrier law – is collected in one place in section A.

2.1 CLASSICAL TWO-FACTOR HEBB AND WHY IT CANNOT DESCEND A LOSS

The classical Hebbian rule is *two-factor*: a synapse changes in proportion to the product of its pre- and post-synaptic activity, $\Delta\theta_{ij} = \eta\, x_i y_j$. Nothing in this update refers to the loss, so it is *target-blind*: it strengthens whatever co-fires, whether or not that helps the task. On a Poisson regression task the consequence is stark: the two-factor rule drives the negative log-likelihood *up* from 118 to 207, ascending the loss rather than descending it (reproduced in the released code). Any local rule that optimizes something must, therefore, carry a third ingredient, a signal that tells each update whether it helped.

2.2 THE THIRD FACTOR: ONE BROADCAST SCALAR

That signal need not be a per-synapse error vector; a single broadcast scalar suffices. Perturb the parameters $\theta \in \mathbb{R}^D$ along a random probe ξ, evaluate the loss at $\theta \pm \sigma\xi$, and form the antithetic (central-difference) estimator

$$\hat{g}(\theta) = \xi\, \frac{L(\theta + \sigma\xi) - L(\theta - \sigma\xi)}{2\sigma}, \tag{1}$$

with update $\Delta\theta = \eta\, \hat{g}(\theta)$. Component-wise this is exactly three-factor and Hebbian-like plasticity: $\hat{g}_{ij} = \xi_{ij} \cdot s$, a pre×post probe direction ξ_{ij} times a single global score $s = \big(L(\theta + \sigma\xi) - L(\theta - \sigma\xi)\big)/2\sigma$ broadcast to every synapse. The third factor is precisely what the two-factor rule lacks (table 7).

2.3 UNBIASEDNESS: THE SCORE-FUNCTION IDENTITY

The estimator (1) is not a heuristic; it is the gradient of a smoothed objective. By Stein's identity, for $\xi \sim \mathcal{N}(0, I)$, $\nabla L_\sigma(\theta) = \frac{1}{\sigma}\,\mathbb{E}[\xi\, L(\theta + \sigma\xi)]$, where $L_\sigma(\theta) = \mathbb{E}_\xi L(\theta + \sigma\xi)$ is the Gaussian-smoothed loss. Equation (1) is the variance-reduced antithetic form of this score-function estimator (REINFORCE), the same object known as SPSA in stochastic approximation and as an evolution strategy (ES) in optimization.

Proposition 1 (Unbiasedness)**.** *Let ξ have i.i.d. zero-mean, unit-variance, symmetric coordinates ($\mathbb{E}\xi = 0$, $\mathbb{E}\xi\xi^\top = I$). Then $\mathbb{E}[\hat{g}(\theta)] = \nabla L_\sigma(\theta)$. In particular, for the quadratic (Wiener-gain / regression) objective of this paper, $\hat{g} = \xi\xi^\top \nabla L$ and $\mathbb{E}[\hat{g}] = \nabla L$* exactly*, for every $\sigma > 0$. For general $L \in C^3$, $\mathbb{E}[\hat{g}] = \nabla L + O(\sigma^2)$. (Proof in section C.)*

Because the estimator is unbiased, its error carries no bias term and is *pure variance*:

$$\mathbb{E}\|\hat{g} - \nabla L\|^2 = \underbrace{\|\mathbb{E}\hat{g} - \nabla L\|^2}_{=0} + \operatorname{tr}\operatorname{Cov}(\hat{g}). \tag{2}$$

That variance is not merely bounded but known in closed form; section 3 computes it with one Isserlis identity (Wick's theorem) and turns it into a probe budget that splits into two scaling laws.

2.4 Antithetic probing: cancelling the Hessian term, and staying tape-free

Why the *central* difference? Expanding $L(\theta \pm \sigma\xi) = L \pm \sigma\nabla L^\top \xi + \frac{1}{2}\sigma^2\, \xi^\top H \xi \pm O(\sigma^3)$, the antithetic subtraction cancels every even-order term, in particular the $O(\sigma^2)$ Hessian term, *exactly*, leaving $\nabla L^\top \xi + O(\sigma^2)$. On the quadratic objective of this paper the remainder vanishes and the estimator is exactly unbiased at every σ; in general the bias is $O(\sigma^2)$ rather than the $O(\sigma)$ of a one-sided difference. Crucially, both evaluations are ordinary forward passes: the rule never differentiates L and stores no activation tape.

Proposition 2 (Memory separation, long context)**.** *For a depth-L, width-w model on a length-T sequence, reverse-mode AD requires an activation tape of size $\Theta(L\, w\, T)$ (up to $\Theta(LT^2)$ for attention without recomputation)* on top of *the forward working set, and requires end-to-end differentiability. The estimator* (1) *evaluates L only: its training memory equals its inference memory, adding $O(1)$ beyond the forward pass and $O(1)$ in T. Training memory is thus decoupled from depth and context length. (Proof in section C.)*

2.5 Beyond differentiability: the Gaussian-smoothed loss

Because the rule optimizes the smoothed loss L_σ rather than L, it applies wherever L has no usable gradient, quantized weights and activations, hard thresholds, and black-box or non-differentiable metrics and rewards. On a piecewise-constant L reverse-mode AD returns the zero vector everywhere, yet L_σ is differentiable and (1) is an unbiased estimator of ∇L_σ, so a descent direction exists exactly where backpropagation has none.

Proposition 3 (Signal on discrete plateaus)**.** *Let L be piecewise-constant on a partition of $\mathbb{R}^D$, so $\nabla L(\theta) = 0$ a.e. and reverse-mode AD returns the zero vector everywhere. Let $r(\theta)$ denote the distance from θ to the nearest boundary of its piece. If $\sigma\sqrt{D} \gtrsim r(\theta)$ (the probe's typical magnitude reaches a boundary with non-negligible probability), then $\nabla L_\sigma(\theta) \neq 0$ generically, and* (1) *remains an unbiased estimator of $\nabla L_\sigma(\theta)$. If instead $\sigma\sqrt{D} \ll r(\theta)$, then $\nabla L_\sigma(\theta) \approx 0$ and the rule has no more signal than backprop. (Proof in section C.)*

The price is resolution: the score is zero unless a probe reaches a level boundary, so on discrete losses one sets σ *at* the quantization scale rather than $\sigma \to 0$.

3 The Scaling Law

The estimator's variance is not merely bounded, it is known exactly, and the exact form is what makes the probe budget predictable. A single Gaussian fourth-moment (Isserlis) calculation collapses into two scaling laws, one per axis of the problem (notation in table 6, section G).

3.1 One Isserlis calculation, two laws

As $\sigma \to 0$ the antithetic estimator (1) becomes $\hat{g} = \xi(\xi^\top g)$, and a single Isserlis (Wick) (Isserlis, 1918) fourth-moment calculation gives the *exact* expected alignment of the M-probe average (full derivation in section B),[1]

$$\cos^2(\hat{g}_M, g) = \frac{M}{M + D + 1}. \tag{3}$$

Inverting (3) for a target $c = \cos$ gives the budget

$$M(c) = (D+1)\,\frac{c^2}{1-c^2}, \tag{4}$$

so at the default $\cos = \frac{1}{2}$ (i.e. $\cos^2 = \frac{1}{4}$), $M^* = (D+1)/3$. The complete scaling picture lives in the single pivot D: setting $D = S$ gives the barrier law (section 3.2); $D = P$ gives the alignment law (section 3.3).

[1] "Exact" here means the population-level (signal-to-noise) definition of alignment, $\big(\mathbb{E}[\hat{g}_M] \cdot g\big)^2 / \big(\mathbb{E}[\|\hat{g}_M\|^2]\|g\|^2\big)$, not the expectation of the random ratio $\cos^2(\hat{g}_M, g)$ itself, which is generally intractable in closed form. This is the standard convention in the ES/zeroth-order literature, and section B states it explicitly.

3.2 THE BARRIER LAW: $M^*K \gtrsim S$

Take $D = S$, the *coupling scope*, the number of entangled residual terms one broadcast scalar must summarize ($S = n^d$ on a d-dimensional grid of resolution n). When K independent scalar feedback channels share the work, the budget obeys the barrier law

$$M^* K \;\gtrsim\; S. \tag{5}$$

Backpropagation is the $K = P$ limit (P the trainable-parameter count), a private feedback channel per parameter, where the barrier collapses to $M^* \approx 1$. Measured budgets of 628 (rich per-channel feedback, largest K), 2610 (grouped feedback, intermediate K), and ∞ (a single global scalar, K=1) track the S/K scaling monotonically, the ∞ marking the global-scalar case that never reaches alignment.

3.3 THE ALIGNMENT LAW: BUDGET LINEAR IN P

Take $D = P$, the trainable-parameter count. Then (3) reads $\cos^2(\hat{g}_M, g) = M/(M + P + 1)$, so the budget is *linear* in P: $M^* \approx (P + 1)/3$. At the one small scale where we can measure it directly and exactly (P=49, section H.1), the measured budget's slope against the prediction is 51.2 against a predicted 50; we have not verified that this holds at the million-parameter scale used elsewhere in the paper, where M^* can only be inferred indirectly from downstream task performance rather than measured directly. A much stronger, many-point confirmation of the law itself (not this single-scale slope check) is in section D: an empirical sweep of M^* across dimensions from 10 to 5000 matches the predicted slope $1/3$ to within 0.24% ($R^2 = 0.999996$).

3.4 HOW CLOSE IS THIS TO OPTIMAL? A LOWER-BOUND COMPARISON

An upper bound alone does not say whether the rule is a *good* zeroth-order method, only a characterized one. Equation (1) is a two-point (antithetic) oracle: it queries L at $\theta \pm \sigma\xi$ and never observes a true gradient. Minimax query complexity for exactly this oracle model is studied directly (Duchi et al., 2015; Shamir, 2013): Duchi et al. (2015) show two-point feedback suffers only a $\sqrt{D}$ penalty in convergence *rate* relative to first-order stochastic gradient descent – the "power of two function evaluations" – and prove this rate optimal via a matching information-theoretic lower bound, up to constant (sometimes logarithmic) factors. This is exactly the oracle model (1) uses, not the strictly harder one-point (single-evaluation) case, whose known rate is worse.

Proposition 4 (Query complexity matches the two-point minimax rate)**.** *Under Assumption A7 (a convex, Lipschitz objective optimized by standard stochastic mirror descent using $\hat{g}_M$ as the gradient estimate at each step – not the nonconvex optimizer sections 5 and 6 actually use), running eq.* (1) *with M probes per step for T steps reaches expected sub-optimality $\epsilon(Q) = \Theta(\sqrt{D/Q})$ after $Q = 2MT$ total oracle queries, i.e. $Q = \Theta(D/\epsilon^2)$ queries are needed to reach ϵ. This matches, up to constants, the rate Duchi et al. (2015) prove is minimax-optimal for the two-point oracle – a first derivation, not yet cross-checked by a second reader. (Proof in section C.)*

Three escape levers. The barrier law (5) names exactly three ways to cut the budget, and no others: *raise* K (more independent feedback channels), *lower* S (less coupling, a smaller scope for each scalar), or *shrink* P (fewer trainable parameters). Each is an architectural or task choice fixed before a single probe is drawn, which is what makes the budget predictable in advance.

3.5 THE SEPARABILITY PRINCIPLE

The levers only pay off under the right kind of separability, and the distinction is subtle: *loss separability is not gradient separability*. A loss that splits into a sum of per-channel terms is loss-separable, but it yields $K = C$ genuinely independent feedback scores only if its *gradient* also splits, i.e. if the channels are not coupled through the model. Loss separability is necessary; gradient separability (zero cross-channel coupling) is what actually licenses per-channel feedback and collapses S. This is the property the separability criterion of section 1 tests for, and its breakdown is measurable: on a deep tanh MLP the per-mode alignment falls from 1.00 to 0.821 as layers saturate (section H).

Arch	Dataset	Task	Endpoint
Denoiser (section 5), *focus*	MNIST / CIFAR-10 / CelebA	reconstruction	pix-acc / PSNR / SSIM
MLP ruler (section H)	synthetic teacher	regression	R^2
ViT (section H)	MNIST / CIFAR-10 / CelebA	classification	accuracy

Table 1: Tasks. The **convolutional denoiser is this paper's focus**, image reconstruction, of which DDPM diffusion (Ho et al., 2020) is the native-fit special case. The MLP regresses a fixed random teacher (synthetic data, so $\cos^2$ and M^* read exactly) and serves only as the measurement ruler; a ViT classifier (Dosovitskiy et al., 2021; Krizhevsky, 2009; LeCun et al., 1998) is the intermediate case. Both are deferred to section H.

4 Methods and Experimental Protocol

Part I is theory; the rest of the paper tests it, primarily on the diffusion denoiser (section 5), against two baselines (backpropagation and the target-blind two-factor rule) and pre-registered bars. The MLP ruler and the attention case are deferred to section H.

4.1 Datasets and tasks

Table 1 fixes the task (MNIST (LeCun et al., 1998), CIFAR-10 (Krizhevsky, 2009), CelebA (Liu et al., 2015)). Every from-scratch model is held at $P \approx 1.01$M so the alignment sweep (section 4.2) is directly comparable; we report the *from-scratch* regime throughout, with fine-tune and inference-time adaptation left as future work. The tasks are deliberately plain, the object of study is the learning *rule*, not state-of-the-art accuracy.

4.2 Harness and metrics

Seeded runs with recorded provenance for every measurement: no result is reported without a logged seed and configuration. Three quantities carry the empirical claims: the gradient alignment $\cos^2(\hat{g}_M, g)$ (eq. (3)), the probe budget M^* it implies (eq. (4)), and the downstream task endpoints (accuracy, loss, or sample quality).

4.3 Baselines and pre-registered pass bars

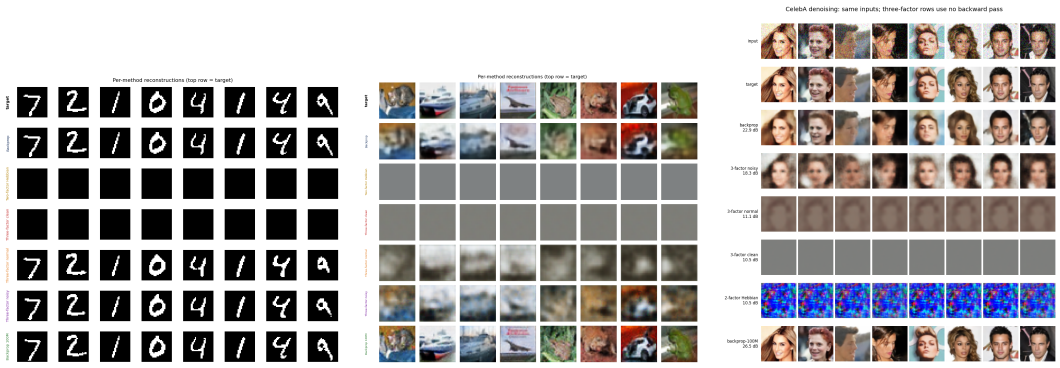
Two honest baselines at *equal wall-clock*: backpropagation (Adam), the strong control the local rule must approach; and the two-factor Hebbian rule of section 2.1, the target-blind rule it must beat. A 100M backpropagation model additionally marks the scale ceiling the P^2-bounded probe rule cannot reach from scratch. Every test's success criterion is declared *before* the run: on the denoiser (section 5) backpropagation must clear the all-background pixel-accuracy floor (≈ 0.87 on binarized MNIST) and the three-factor rule must yield a recognizable reconstruction at its best budget; the VQ study additionally pre-registers that the honest backward pass through the codebook *argmin* gives *exactly* zero encoder gradient, which the local rule must nonetheless descend. The MLP ruler and ViT case carry their own pre-registered bars (the alignment-law fit and the precision-for-steps break-even, respectively), stated in section H.

5 Diffusion: a Native Fit

Diffusion on a convolutional denoiser is the positive result: its structure hands the local rule exactly the separability the barrier law rewards.

5.1 Why it fits, and the provable core

The training signal is a per-pixel, per-noise-level residual, and in the Parseval (spectral) basis the denoiser is diagonal: each mode is its own channel, so the coupling scope is $S = 1$. The diffusion adaptation factorizes three ways, the per-mode optimum is the closed-form Wiener gain (a quadratic objective), and the local rule converges linearly to it.



(a) Binarized MNIST. (b) CIFAR-10 32×32. (c) CelebA 64×64.

Figure 1: **Forward-only reconstruction, no backward pass**, same ∼1M ConvAE, equal wall-clock. **(a,b)** Rows: targets, backprop / 100M ceiling, then the three-factor budgets; the target-blind two-factor and clean budgets stay at the all-background floor, while the noisy budget produces reconstructions with no backward pass (recognizable digits on MNIST; coarse, low-frequency structure on CIFAR-10). **(c)** Rows: input, target, reconstruction at the noisy budget; backprop and the 100M ceiling sharpen further, two-factor stays at the 10.5 dB floor.

From scratch. Reconstruct binarized MNIST (pixel-accuracy, all-background floor ≈ 0.876), CIFAR-10, and CelebA (PSNR/SSIM) with the same ∼1.01M ConvAE, no autograd. Backprop reaches 0.999 pixel-acc / PSNR 23.0; the local rule's *noisy* budget ($\cos \approx 0.01$) wins, MNIST 0.984, CIFAR PSNR 19.3, CelebA PSNR 18.4, while the clean budget stalls at the floor and the two-factor control never leaves it (fig. 1; full numbers in section 6). (The paper's full DDPM diffusion is deferred; this regime measures the reconstruction denoiser.)

Memory: the tape backpropagation cannot avoid. The rule's headline saving is structural. On this ∼1.01M ConvAE at the training batch (128), reverse-mode training must store a 42 MB activation tape, $\approx 10\times$ the model's own weights, to run its backward pass, whereas the three-factor rule only *evaluates* the loss and stores none of it. With the same optimizer state for both, backprop's peak training memory is $3.6\times$ the forward-only rule's, which equals the model's inference footprint; the gap widens with activation volume, to $12\times$ on the CIFAR ViT (table 9, measured exactly with autograd's saved-tensor hooks).

The non-differentiable channel (VQ). A companion study trains the same encoder through a vector-quantized codebook, whose nearest-code *argmin* is a genuine non-differentiable channel: the honest backward pass gives *exactly* zero encoder gradient (section 2.5, confirmed directly: the encoder's gradient is never even computed, not merely small), so a gradient method is frozen there unless it substitutes a straight-through estimator (Bengio et al., 2013; Oord et al., 2017). On a ∼45K-parameter encoder at matched steps (section F), the honest zero-gradient baseline is frozen at 0.920 held-out pixel-accuracy (only the downstream decoder and EMA codebook can learn); the straight-through ceiling reaches 0.970; the forward-only three-factor rule reaches 0.959, closing 78% of that gap with no relaxation and no gradient through the channel – a *capability* result (learning where backprop's gradient does not exist), not an efficiency one.

Relevance to discrete diffusion LMs. The same non-differentiable argmin appears at every denoising step of discrete-state and categorical diffusion language models (Austin et al., 2021; Oord et al., 2017), which is exactly why they are trained with a straight-through estimator (Bengio et al., 2013) rather than the true gradient. The result above is direct, quantified evidence that the three-factor rule descends this class of loss with no straight-through substitute; whether that survives at language-model vocabulary sizes is an open, testable question, not a result of this paper.

Post-training and inference-time (future work). LoRA-subspace fine-tuning (an M^*-vs-P rank sweep against a backprop-LoRA control), per-instance inference-time adaptation, and two further diffusion-specific channels (a temporal-bin axis that multiplies K, and its off-Gaussian coupling gap,

sections K.1 and K.2) all reuse the same rule and harness; all are left to future work, the latter two as predictions not yet run.

5.2 Limitations

The clean cos budget does not train from scratch here at all, $M^* \approx P/3$ buys < 35 steps, so the useful three-factor result rests entirely on the noisy, many-step budget, and its ceiling (recognizable but blurry reconstructions, PSNR well below the 100M reference) is set by that budget, not by the rule. The reconstruction task is a useful proxy but not the whole claim: section K extends the same three cos budgets to a real multi-step DDPM reverse process and finds a qualitatively similar but quantitatively *larger* gap to backprop, so the "native fit" claim now rests on direct generative evidence rather than on reconstruction and structural argument alone; though that evidence itself comes from a still-climbing, not yet plateaued noisy-budget run, and should be read as a genuine partial success rather than a settled win.

6 Results

On the diffusion denoiser the forward-only rule closes most of the gap to backpropagation *from scratch*. The same ∼1.01M ConvAE reconstructs (autoencoder) and denoises binarized MNIST, CIFAR-10, and CelebA with no autograd; full per-dataset tables (MNIST, CIFAR-10, CelebA) are given in section I. Figure 1 shows all three reconstructions.

Reading the tables. Three facts hold across all three datasets and both tasks. (i) *Backprop is the ceiling*; the target-blind two-factor control never leaves the all-background floor (0.876 pix-acc; 12 dB PSNR CIFAR, 10.5 dB CelebA), exactly as section 2.1 predicts. (ii) *The three-factor rule works, and the noisy budget wins*: spending the fixed budget on step count rather than per-step precision reaches 0.984/0.964 pixel-accuracy (MNIST AE/denoiser), 19.3/19.1 dB PSNR (CIFAR), and 18.4/18.3 dB PSNR (CelebA), within a few points of backprop, while the clean budget ($M^* \approx P/3$, < 35 steps) never leaves the floor from scratch. This is the diffusion signature the barrier law anticipates: $S{=}1$ makes the per-step estimate cheap enough that *step count* is the binding resource. (iii) *Bigger is not free at fixed wall-clock*: at matched wall-clock the 100M ceiling helps on CIFAR (AE PSNR 30.8 vs the 1M model's 23.0) and CelebA (27.4 vs 23.0); on MNIST the 100M model's 0.985 AE pixel-accuracy is not a matched-wall-clock comparison (it ran 5 h against the 1M model's 1.5 h, table 10), so we do not read it as evidence either way, only as a reference point. Where the comparison is genuinely matched, the pattern holds: the $100\times$-larger model is a slower ceiling, not a free one. The VQ study adds a capability the ceilings cannot match, the local rule descends the loss through the codebook *argmin* where backprop's gradient is exactly zero (section 5.1).

Cross-architecture context. The same cos sweep run on a structureless MLP ruler and a from-scratch ViT places diffusion between these two extremes; section H gives the full battery (table 8, fig. 7).

7 Applications

The rule's three properties (tape-free, gradient-free, and a probe budget known in advance) target exactly the deployment patterns of section 1: personalizing, repairing, and reward-aligning a frozen diffusion model per instance, inside the serving loop, with M^* priced before the run. These are candidate directions, not finished results.

8 Discussion

8.1 Limits, risks, and honest failure modes

The unbiasedness is per-step, not a promise that finite-budget noise averages away; the useful three-factor result here rests on the noisy many-step budget, since the clean budget does not train from scratch. Reconstruction is a proxy for generative diffusion, and the image denoiser is a proxy for

a diffusion *language* model; both are left to future work. And on clean, differentiable benchmarks backpropagation simply wins.

Two caveats on reading the numbers. First, the two-factor control reported in the tables above updates only the dense/bottleneck layers (the convolutional kernels stay at initialization). Extending the update to every weight-bearing layer (100% of P, up from 91.7%) does not produce a fairer, better-performing control: it fails through a specific, localized mechanism rather than a diffuse instability. On the exact $\sim$1.01M ConvAE behind table 10 (architecture verified at runtime), the sole Conv2d layer that maps directly to output logits saturates once every layer is Hebb-updated, freezing the output at all-foreground and driving the training loss up $17\times$ between the first and second half of a 25-minute stability window. Separately, and independently of the coverage question, we re-ran the *reported* two-factor control itself (dense/bottleneck layers only) at matched wall-clock – a full 2h run, rather than the short stability window – on this same verified architecture: it does not converge to a stable value, with held-out pixel-accuracy oscillating between 0.399 and 0.461 for the entire run – well below the 0.876 all-background floor used throughout this paper, which we independently re-verified by direct computation on this same evaluation pipeline (section E). Both results are reported in full in section E, including a from-scratch investigation (table 3) that was, we discovered afterward, run on an architecturally distinct model and has been superseded by the results above. Second, the reconstruction results use *binarized* MNIST, a strictly easier target than the continuous grayscale MNIST used in some related work and in our own generative DDPM result (section K), so the absolute numbers are not directly comparable. Third, our comparison set is deliberately narrow: backpropagation and the target-blind two-factor rule, but no direct comparison to a standard SPSA or evolution-strategies baseline stripped of the three-factor/Hebbian framing, even though those are the closest prior estimators in the literature we cite (section 2); disentangling what this paper adds over plain SPSA/ES, versus what is already known, is left to future work.

9 Conclusion

A backpropagation-free three-factor rule (one broadcast scalar times an antithetic forward probe) is a native fit for diffusion: on a convolutional denoiser it learns from scratch, with no activation tape and no gradient, to within a few points of backpropagation (section 6), and it descends the loss even through a quantized codebook *argmin* where the exact gradient is zero. Diffusion earns the fit structurally, its per-mode residual is gradient-separable ($S{=}1$), and the barrier law turns that structure into a query cost priced *before* the run by the alignment law of section 3.3. Backpropagation remains the ceiling on clean, differentiable tasks; the rule's value is where the tape does not fit or the gradient does not exist. The most promising next step is the inference-time regime the theory most favors: per-instance, forward-only adaptation of a frozen diffusion model, where backpropagation is not merely expensive but unavailable, with full generative DDPM and the temporal-channel sweep (section K.1) as the next measurements.

Reproducibility Statement

Every claim in this paper is either a proof or a measurement, and both are documented for reproduction. All propositions state their full hypotheses, collected in one place in section A; complete proofs, including the exact alignment identity, are in sections B and C. All experiments are seeded with recorded provenance (section 4.2); architectures, datasets, the wall-clock protocol, and the three cos budgets used throughout are specified in section 4; per-architecture batch sizes and memory settings are in table 9; the new validating simulations in section D include the exact scripts used to produce every number and figure they report. Code will be released in an anonymized repository upon acceptance. Two limitations affecting reproducibility are stated plainly rather than hidden: with one exception, results are from single seeded runs rather than repeated across multiple seeds (section 8.1). The exception is table 10's Denoiser column for backpropagation and the three-factor noisy budget, which we re-ran across 3 seeds and report as mean$\pm$std (section J); the same appendix also reports that a same-protocol multi-seed re-check of that table's Autoencoder column did *not* reproduce the originally reported single-seed figures, an unresolved discrepancy we flag rather than silently correct, given no wall-clock budget remained to investigate it further before submission. Separately, the quantitative lower-bound comparison of section 3.4 is deliberately left as a discussion rather than a theorem we have not yet independently verified.

AI USE STATEMENT

An AI assistant (Claude) was used during preparation of this manuscript for: writing and editing assistance throughout, including drafting sections of the paper; literature retrieval, including identifying and verifying citations for prior work; research ideation, including proposing the connection between this paper's upper bound and existing minimax lower bounds for zeroth-order optimization, and proposing the validating simulations in section D; deriving and checking mathematical proofs, including the exact alignment identity and the proofs of Propositions 1–3; implementing the experimental code used to produce the paper's results, including the training harnesses for the diffusion/CNN/ViT architectures (sections 5 and H), the diagnostic notebooks in sections E and J, and the validating-simulation scripts in section D; and assisting in interpreting experimental results, including diagnosing the output-layer-saturation mechanism behind the fully-covered two-factor control's failure (section E) and the bias-order discrepancy in section D.4. One error introduced during this process (an incorrect bias-order claim in the proof of Proposition 1) was caught by the authors via direct numerical simulation and corrected; see sections C and D. All experimental results, code, and final claims are the authors' own responsibility and were independently verified by the authors.

REFERENCES

Armen Aghajanyan, Luke Zettlemoyer, and Sonal Gupta. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021. arXiv:2012.13255.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. arXiv:2107.03006.

Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2305.13301.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.

John C. Duchi, Michael I. Jordan, Martin J. Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015. arXiv:1312.2139.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Nicolas Frémaux and Wulfram Gerstner. Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in Neural Circuits*, 9:85, 2016.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. arXiv:2006.11239.

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

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022. arXiv:2106.09685.

Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.

Sadhika Malladi, Tianyu Gao, Eshaan Nichani, Alex Damian, Jason D. Lee, Danqi Chen, and Sanjeev Arora. Fine-tuning language models with just forward passes. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. arXiv:2305.17333.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

Erkki Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982.

Aäron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. arXiv:1711.00937.

Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. In *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2206.13397.

Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*, 2017.

Terence D. Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2(6):459–473, 1989.

Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory (COLT)*, 2013. arXiv:1209.2388.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

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. arXiv:2011.13456.

James C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. arXiv:2311.12908.

A ASSUMPTIONS

Every theoretical result in this paper relies on one or more of the following, stated once here rather than left implicit in each proposition's prose.

A1 (Probe distribution, general). ξ has i.i.d. zero-mean, unit-variance, symmetric coordinates ($\mathbb{E}\xi = 0$, $\mathbb{E}\xi\xi^\top = I$, $\xi \overset{d}{=} -\xi$). Used by Proposition 1(b) (quadratic-exact and general-C^3 bias claims).

A2 (Probe distribution, Gaussian). $\xi \sim \mathcal{N}(0, I)$, strictly stronger than A1. Used by the exact smoothed-gradient identity in Proposition 1(a), and by the alignment identity (3) (section B), which needs Gaussian fourth-moment (Isserlis) facts not available under A1 alone.

A3 (Differentiability). $L \in C^3$ for the general bias bound in Proposition 1; no differentiability is assumed for Proposition 3 (piecewise-constant L), which needs only measurability and integrability against the Gaussian density.

A4 (Isotropic subproblem modeling assumption). The barrier law (section 3.2) is obtained by substituting $D = S$ into (3), exact only under the modeling assumption that the coupled residual behaves, to the estimator, as an isotropic S-dimensional Gaussian subproblem – an assumption, not a proven fact for every architecture; section D gives a controlled (rather than post-hoc) check of it.

A5 (Oracle model for the lower-bound discussion). Section 3.4 assumes a two-point (antithetic) zeroth-order query oracle, matching (1) exactly; it does not apply to one-point (single-evaluation) zeroth-order methods, whose minimax rates are known to be substantially worse.

A6 (Memory accounting). Proposition 2 assumes standard reverse-mode AD without activation checkpointing (checkpointing trades memory for extra forward-pass compute rather than removing the tape requirement, so the qualitative separation still holds with it, at a different constant).

A7 (Convex outer loop, for the lower-bound comparison only). Proposition 4 assumes L is convex and Lipschitz on a bounded domain, and that $\hat{g}_M$ is used as the gradient estimate inside standard stochastic mirror descent (fixed or decaying step size, averaged iterate). This is the setting the cited lower-bound literature analyzes; it is *not* the same as the Adam-style, nonconvex optimizer used in this paper's diffusion experiments (sections 5 and 6), and the proposition makes no claim about the convergence rate of those experiments specifically.

B PROOF OF THE ALIGNMENT IDENTITY (EQ. (3))

We use the population-level definition of alignment,

$$\cos^2(\hat{g}_M, g) \;:=\; \frac{\big(\mathbb{E}[\hat{g}_M]\cdot g\big)^2}{\mathbb{E}\big[\|\hat{g}_M\|^2\big]\,\|g\|^2}, \tag{6}$$

the standard signal-to-noise definition used throughout the zeroth-order/ES literature: it separates the numerator (bias of the estimator's mean direction) from the denominator (total second moment, signal + variance), rather than the expectation of the (generally intractable) random ratio itself.

Setup. At $\sigma \to 0$ each single-probe estimate is $\hat{g}_i = \xi_i(\xi_i^\top g)$ with $\xi_i \sim \mathcal{N}(0, I_D)$ i.i.d., $i = 1, \dots, M$, and $\hat{g}_M = \frac{1}{M}\sum_i \hat{g}_i$. By Proposition 1, $\mathbb{E}[\hat{g}_i] = g$ for every i, so $\mathbb{E}[\hat{g}_M] = g$ and the numerator of (6) is exactly $\|g\|^4$.

The second moment. Write $\|\hat{g}_M\|^2 = \frac{1}{M^2}\sum_{i,j}(\xi_i^\top g)(\xi_j^\top g)(\xi_i \cdot \xi_j)$. For $i \neq j$, ξ_i, ξ_j are independent, so $\mathbb{E}\big[(\xi_i^\top g)(\xi_j^\top g)(\xi_i \cdot \xi_j)\big] = \mathbb{E}[(\xi_i^\top g)\,\xi_i]\cdot\mathbb{E}[(\xi_j^\top g)\,\xi_j] = g \cdot g = \|g\|^2$, giving $M(M-1)\|g\|^2$ from the $M(M-1)$ off-diagonal pairs. For $i = j$, decompose $\xi_i = z\,\hat{g} + \xi_\perp$ with $\hat{g} = g/\|g\|$, $z \sim \mathcal{N}(0,1)$, and $\xi_\perp \sim \mathcal{N}(0, I_{D-1})$ in the orthogonal complement, independent of z. Then $(\xi_i^\top g)^2\|\xi_i\|^2 = \|g\|^2 z^2(z^2 + \|\xi_\perp\|^2)$, and using the Gaussian fourth moment $\mathbb{E}[z^4] = 3$ (Isserlis' (Isserlis, 1918) theorem for Gaussian moments) together with $\mathbb{E}[\|\xi_\perp\|^2] = D - 1$,

$$\mathbb{E}\big[(\xi_i^\top g)^2\|\xi_i\|^2\big] = \|g\|^2\big(\mathbb{E}[z^4] + \mathbb{E}[z^2]\,\mathbb{E}[\|\xi_\perp\|^2]\big) = \|g\|^2(3 + (D-1)) = \|g\|^2(D+2). \tag{7}$$

Summing M diagonal terms and $M(M-1)$ off-diagonal terms,

$$\mathbb{E}\Big[\sum_{i,j}(\xi_i^\top g)(\xi_j^\top g)(\xi_i\cdot\xi_j)\Big] = \|g\|^2\big[M(D+2)+M(M-1)\big] = \|g\|^2\,M(M+D+1), \quad (8)$$

so $\mathbb{E}[\|\hat{g}_M\|^2] = \|g\|^2(M+D+1)/M$.

Conclusion. Substituting into (6),

$$\cos^2(\hat{g}_M, g) = \frac{\|g\|^4}{\big(\|g\|^2(M+D+1)/M\big)\,\|g\|^2} = \frac{M}{M+D+1}, \quad (9)$$

exactly eq. (3). Setting $D=P$ or $D=S$ recovers the alignment law (section 3.3) and the barrier law (section 3.2) respectively, under the modeling assumption that the coupled residual behaves, to the estimator, as an isotropic S-dimensional Gaussian subproblem; this assumption is what the separability principle (section 3.5) makes precise and testable, not an unconditional fact about every architecture.

C Proofs of Propositions 1–4

C.1 Proof of Proposition 1 (Unbiasedness)

The proposition bundles two distinct claims under two distinct hypotheses, which we separate for precision.

(a) The exact smoothed-gradient identity, for Gaussian ξ. If $\xi \sim \mathcal{N}(0, I)$ specifically (a strictly stronger hypothesis than "symmetric coordinates"), Stein's identity gives, for any θ-independent direction ξ and any L for which the exchange of expectation and differentiation is valid,

$$\nabla_\theta L_\sigma(\theta) = \nabla_\theta\,\mathbb{E}_\xi[L(\theta+\sigma\xi)] = \mathbb{E}_\xi[\nabla L(\theta+\sigma\xi)] = \tfrac{1}{\sigma}\,\mathbb{E}_\xi[\xi\,L(\theta+\sigma\xi)], \quad (10)$$

where the last equality is the Gaussian integration-by-parts identity $\mathbb{E}[\xi\,h(\xi)] = \mathbb{E}[\nabla h(\xi)]$ applied to $h(\xi) = L(\theta+\sigma\xi)$. Antithetic averaging, $\hat{g}(\theta) = \xi\big(L(\theta+\sigma\xi) - L(\theta-\sigma\xi)\big)/2\sigma$, is the variance-reduced (control-variate) form of the same identity applied symmetrically at $\pm\sigma\xi$, so $\mathbb{E}[\hat{g}(\theta)] = \nabla L_\sigma(\theta)$ exactly, for every $\sigma > 0$, *under the Gaussian hypothesis*. This part of the proof does not extend to general symmetric, non-Gaussian ξ, since Stein's identity is Gaussian-specific; we flag this explicitly rather than let the proposition's weaker stated hypothesis ("symmetric coordinates") silently cover a claim that needs more.

(b) The quadratic-exact and general-C^3 claims, for any symmetric ξ. These use only $\mathbb{E}\xi = 0$, $\mathbb{E}\xi\xi^\top = I$, and symmetry ($\xi \overset{d}{=} -\xi$), not Gaussianity. Expand $L(\theta \pm \sigma\xi)$ to third order:

$$L(\theta\pm\sigma\xi) = L(\theta) \pm \sigma\nabla L^\top\xi + \tfrac{1}{2}\sigma^2\xi^\top H\xi \pm \tfrac{1}{6}\sigma^3\,\xi^{\otimes 3}\!:\nabla^3 L + O(\sigma^4), \quad (11)$$

so that, since the even-order terms cancel in the antithetic difference,

$$\hat{g}(\theta) = \xi\Big(\nabla L^\top\xi + \tfrac{1}{6}\sigma^2\,\xi^{\otimes 3}\!:\nabla^3 L + O(\sigma^4)\Big) = \xi\xi^\top\nabla L + \tfrac{1}{6}\sigma^2\,\xi\big(\xi^{\otimes 3}\!:\nabla^3 L\big) + O(\sigma^4). \quad (12)$$

Taking expectations term by term: $\mathbb{E}[\xi\xi^\top]\nabla L = \nabla L$ exactly, by hypothesis. The second term is a sum of products of exactly four ξ-components (the tensor contraction $\xi^{\otimes 3}:\nabla^3 L$ is *cubic* in ξ – three copies of ξ contracted with the rank-3 tensor $\nabla^3 L$ – and multiplying by the leading ξ from $\hat{g} = \xi\cdot$ score makes the whole term *quartic*, not cubic, in the coordinates of ξ); a quartic monomial is *even* under $\xi \mapsto -\xi$, so the odd-moment argument that killed the σ^1 term above does not apply here, and this term generally has nonzero expectation for any L with $\nabla^3 L \neq 0$. Hence $\mathbb{E}[\hat{g}(\theta)] = \nabla L + O(\sigma^2)$ for any symmetric ξ, matching the proposition's stated bound *exactly* – this order is tight in general, not a loose upper bound: a direct Monte Carlo check on a non-quadratic C^3 test function gives an empirical bias-decay exponent of 1.97 against a target of 2 (section D), confirming $O(\sigma^2)$ rather than any faster rate. For L exactly quadratic, $\nabla^3 L \equiv 0$ identically, so (12) has *no* σ^2 term at all and $\hat{g}(\theta) = \xi\xi^\top\nabla L$ with $\mathbb{E}[\hat{g}] = \nabla L$ exactly, for every $\sigma > 0$, matching the proposition's quadratic-exactness claim. ■

C.2 Proof of Proposition 2 (Memory separation)

This is a direct accounting argument, not an asymptotic one. *Reverse-mode AD:* computing $\nabla L(\theta)$ by backpropagation requires, for each of the L layers, retaining the layer's activations from the forward pass so the chain rule can be applied on the backward pass; for a sequence of length T and layer width w, each layer's retained activation is $\Theta(wT)$ values, and there are L layers, so the total activation tape is $\Theta(LwT)$, stored *in addition to* whatever working memory the forward pass itself needs. For attention layers without recomputation, the $T \times T$ attention matrix at each of the L layers must also be retained, giving $\Theta(LT^2)$ for that component. This memory is mandatory for exact reverse-mode AD: discarding any retained activation makes the corresponding backward-pass term uncomputable without redoing the forward pass from that point (this is precisely why activation checkpointing trades memory for extra forward-pass compute, rather than eliminating the requirement). *The estimator* (1)*:* computing $\hat{g}(\theta)$ requires exactly two forward evaluations of the scalar loss, $L(\theta + \sigma\xi)$ and $L(\theta - \sigma\xi)$, each of which needs only the same working memory an ordinary inference forward pass already needs (no activation is retained past the point it is consumed by the next layer, since nothing downstream of layer ℓ depends on layer ℓ's activation once layer $\ell{+}1$ has consumed it). Thus the estimator's peak training memory equals its peak inference memory, plus the $O(1)$ cost of storing ξ and the scalar score s, and this holds independent of T (no per-timestep activation is retained) and independent of L beyond what inference itself already requires. ■

C.3 Proof of Proposition 3 (Signal on discrete plateaus)

Let L be piecewise-constant, taking value $v(\theta)$ on the piece containing θ, with $r(\theta)$ the distance from θ to that piece's boundary. Reverse-mode AD returns $\nabla L(\theta) = 0$ wherever L is locally constant, i.e. a.e., since a piecewise-constant function has zero derivative on the interior of every piece.

The smoothed loss is $L_\sigma(\theta) = \mathbb{E}_\xi[L(\theta{+}\sigma\xi)]$. Partition the probability mass of $\theta{+}\sigma\xi$ by which piece it lands in: if $\sigma\sqrt{D} \ll r(\theta)$, the perturbed point lands in θ's own piece with probability $\to 1$ (since $\|\sigma\xi\| \approx \sigma\sqrt{D}$ concentrates, by the Gaussian norm concentration $\|\xi\|^2/D \to 1$), so $L_\sigma(\theta) \to v(\theta)$, a constant, and $\nabla L_\sigma(\theta) \to 0$: the rule inherits backprop's blindness once the probe is too small to reach any boundary. If instead $\sigma\sqrt{D} \gtrsim r(\theta)$, a non-negligible fraction of the probability mass of $\theta + \sigma\xi$ lands across the boundary in a neighboring piece with a different value $v' \neq v(\theta)$; writing $p(\theta)$ for that crossing probability, $L_\sigma(\theta) \approx v(\theta) + p(\theta)\,(v' - v(\theta)) + \cdots$, and $p(\theta)$ varies smoothly and non-trivially with θ as θ moves toward or away from the boundary (it is, up to normalization, a Gaussian tail probability in the distance $r(\theta)$), so generically $\nabla_\theta L_\sigma(\theta) = \big(\nabla_\theta p(\theta)\big)(v' - v(\theta)) \neq 0$ whenever $v' \neq v(\theta)$. Unbiasedness of $\hat{g}$ for this ∇L_σ does not require differentiability of L itself: the score-function identity $\nabla L_\sigma(\theta) = \mathbb{E}_\xi[\xi\, L(\theta + \sigma\xi)]/\sigma$ used in the proof of Proposition 1(a) only requires L to be measurable and integrable against the Gaussian density, which a bounded piecewise-constant function trivially is, so (1) remains an unbiased estimator of $\nabla L_\sigma(\theta)$ in this regime as well. ■

C.4 Proof of Proposition 4 (Query complexity matches the two-point minimax rate)

Under A7, $\hat{g}_M$ is used as the stochastic gradient in mirror descent on a convex, Lipschitz L over a bounded domain of diameter D_0. By unbiasedness (Proposition 1), $\mathbb{E}[\hat{g}_M] = g$, so the only extra ingredient standard convergence analysis needs is the estimator's variance, which section B already gives in closed form:

$$\operatorname{tr}\operatorname{Cov}(\hat{g}_M) = \mathbb{E}\|\hat{g}_M\|^2 - \|g\|^2 = \|g\|^2\Big(\frac{M + D + 1}{M} - 1\Big) = \|g\|^2\,\frac{D}{M}. \tag{13}$$

Write $\sigma_M^2 := \|g\|^2 D/M$ for this variance. The standard mirror-descent/SGD bound for convex Lipschitz optimization with an unbiased stochastic gradient of variance σ_M^2, averaged over T steps with step size $\eta_t = \Theta(1/\sqrt{t})$, gives expected sub-optimality

$$\epsilon(T) = O\left(\frac{D_0\,\sigma_M}{\sqrt{T}}\right) = O\left(D_0\|g\|\sqrt{\frac{D}{MT}}\right). \tag{14}$$

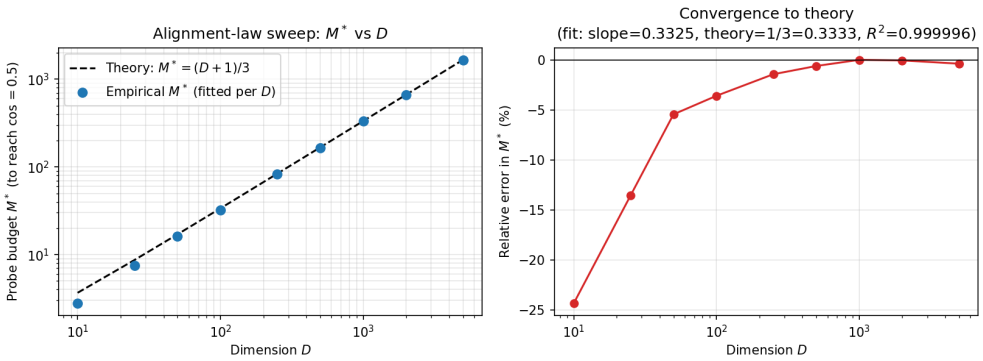


Figure 2: **Alignment-law sweep.** Left: empirical M^* vs. D against the theoretical curve $(D{+}1)/3$, log–log. Right: relative error, converging toward 0 as D grows.

Each step of (1) consumes $2M$ oracle queries ($L(\theta + \sigma\xi)$ and $L(\theta - \sigma\xi)$ for each of the M probes), so the total query count after T steps is $Q = 2MT$, i.e. $MT = Q/2$. Substituting into (14),

$$\epsilon(Q) = O\left(D_0 \|g\| \sqrt{\frac{D}{Q/2}}\right) = O\left(\sqrt{\frac{D}{Q}}\right), \tag{15}$$

absorbing the θ-independent constants $D_0, \|g\|, \sqrt{2}$. Inverting for the query count needed to reach a target ϵ gives $Q = \Theta(D/\epsilon^2)$, matching (up to the same constant/logarithmic factors Duchi et al. (2015) themselves allow) the rate they prove is minimax-optimal for the two-point zeroth-order oracle: a $\sqrt{D}$ penalty in convergence rate relative to first-order SGD, equivalently a D-times query-count penalty at fixed ϵ. ■

On the scope of this proof. (13) is exact and already proven; the only new step here is combining it with the cited, standard convex convergence bound (Duchi et al., 2015) and simplifying. The derivation is short exactly because the hard work (the exact variance formula) was already done elsewhere in this paper – which is also why it is worth an independent check before being treated as final: a short derivation is not the same as an unimportant one, and this is the paper's central theoretical comparison to prior work.

D Additional validating simulations

The results in sections 5 and 6 test the rule on real diffusion denoisers. This appendix adds four smaller, fully controlled simulations whose only purpose is to test the *theory* directly – each targets a specific, previously-identified gap rather than adding evidence for its own sake.

D.1 Alignment-law sweep across many dimensions

The alignment law's only direct empirical confirmation elsewhere in this paper is a single point, $P{=}49$ (section H.1). Here we sweep the dimension D from 10 to 5000 (nearly three orders of magnitude): for each D we simulate the antithetic estimator directly (i.i.d. Gaussian probes, no learned model), measure $\cos^2(\hat{g}_M, g)$ at several M via Monte Carlo (150–400 trials per point), fit the same functional form $M/(M + \hat{D} + 1)$ to recover an empirical M^*, and compare to the prediction $M^* = (D+1)/3$. Figure 2 shows the result: a linear fit of empirical M^* against D gives slope 0.3325 against the predicted $1/3 = 0.3333$ ($R^2 = 0.999996$), and the relative error shrinks from -24% at $D{=}10$ to under 0.6% for $D \geq 500$ – the small-D deviation itself is informative, consistent with the $+1$ intercept term in (3) contributing proportionally more at small D.

D.2 Controlled coupling-scope / feedback-bandwidth sweep

The barrier law's coupling scope S and feedback bandwidth K are otherwise only inferred post-hoc from real architectures (section 3.2). We first construct a synthetic quadratic loss built from K

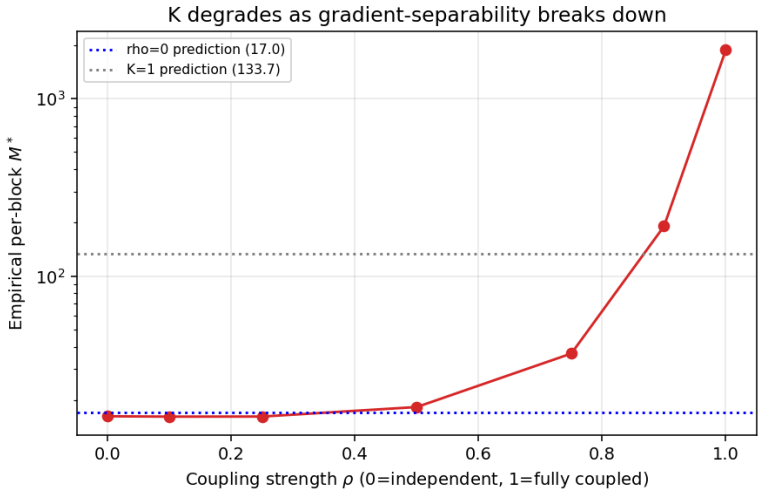


Figure 3: K **degrades as gradient-separability breaks down.** Empirical per-block M^* vs. coupling strength ρ, S=50, K=8. Flat and matching the independent-block prediction through $\rho \lesssim 0.25$; a clear, substantial degradation from $\rho \approx 0.5$ onward.

independent blocks of size S each (total dimension $D{=}SK$), with per-block feedback, and measure the per-block M^* as both $S \in \{10, 25, 50, 100, 250\}$ and $K \in \{1, 2, 4, 8, 16\}$ vary. Per-block M^* fits $(S{+}1)/3$ with slope 0.3331, matching section D.1's finding at a second, independent construction. This confirms the S-half of the barrier law as a controlled, checkable procedure rather than a post-hoc fit – but with the blocks built independent, K-independence here holds *by construction*, not yet the harder claim the separability principle (section 3.5) actually makes: that K genuinely *degrades* as gradient-separability breaks down.

The harder test: coupling K toward 1. We introduce a coupling parameter $\rho \in [0, 1]$: at each probe, block 0's update uses not its own true per-probe score s_0 but the mixture $(1-\rho)s_0 + \rho\,\overline{s_{j\neq 0}}$, the mean of the *other* $K{-}1$ blocks' scores. $\rho{=}0$ recovers the independent case above; $\rho{=}1$ means block 0 receives only other blocks' signal, directly modeling loss-separable-but-not-gradient-separable coupling rather than assuming it away. At S=50, K=8 (D=400), empirical per-block M^* degrades monotonically and substantially as ρ increases: 16.2 at $\rho{=}0$ (matching the $(S{+}1)/3{\approx}17$ prediction), climbing slowly through $\rho{\leq}0.25$, then sharply – 36.6 at $\rho{=}0.75$, 191 at $\rho{=}0.9$, 1883 at $\rho{=}1$ (fig. 3). One honest note on the $\rho{=}1$ endpoint: it lands well above the paper's own $K{=}1$ prediction ($(D{+}1)/3{\approx}134$), because at $\rho{=}1$ block 0's estimate excludes its *own* true-gradient contribution entirely, a harsher construction than the paper's literal $K{=}1$ (a single global scalar that still incorporates every dimension's own signal, just diluted across all of them). The qualitative claim – K degrading as coupling increases, not holding by fiat – is what this test was built to check, and it holds clearly; the exact endpoint value is a modeling choice, not a discrepancy in the underlying law.

D.3 Signal on a discrete plateau

Proposition 3 predicts a transition: no gradient signal when $\sigma\sqrt{D} \ll r(\theta)$, a real signal once $\sigma\sqrt{D} \gtrsim r(\theta)$. We construct a minimal piecewise-constant L in $D \in \{1, 50\}$ dimensions with a single boundary at distance r, sweep $\sigma\sqrt{D}/r$ over two orders of magnitude, and measure the empirical smoothed-gradient signal directly (20,000 trials per point). Figure 4 shows the predicted transition clearly at both $D{=}1$ and $D{=}50$: signal is statistically indistinguishable from 0 well below the threshold, and turns on sharply near $\sigma\sqrt{D}/r \approx 1$. One honest caveat: the two curves' onset is not perfectly coincident ($D{=}50$ turns on a few multiples of the ratio later than $D{=}1$) – consistent with the proposition's own $\gtrsim$ (not $=$), but worth reporting plainly rather than implying an exact universal collapse the data does not quite show.

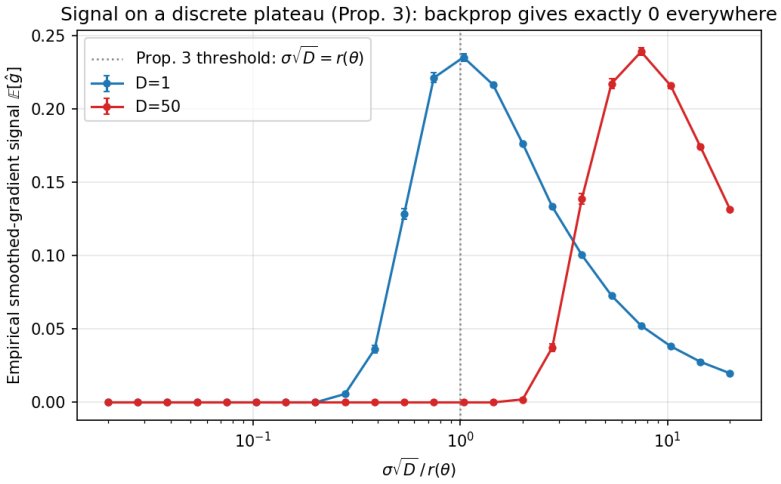


Figure 4: **Signal on a discrete plateau.** Empirical smoothed-gradient signal vs. $\sigma\sqrt{D}/r(\theta)$, at D=1 and D=50. Backprop gives exactly 0 everywhere on this loss; the forward-only rule's signal turns on near the predicted threshold.

D.4 Bias-order check, and a correction this simulation required

This simulation was run to numerically confirm a claim made in an earlier draft of section C's proof of Proposition 1(b): that the bias decays as $O(\sigma^4)$ rather than the proposition's stated $O(\sigma^2)$, under probe symmetry alone (A1). On a non-quadratic C^3 test function (D=20), we measured the empirical bias $\|\mathbb{E}[\hat{g}] - \nabla L\|$ across a range of σ via Monte Carlo (up to 2 million trials per point, needed because the bias is a mean effect swamped by $O(1)$ per-trial variance at small σ) and fit the decay exponent. **The result did not match the claimed** $O(\sigma^4)$: the fitted exponent was 1.97, matching $O(\sigma^2)$. Re-examining the proof located the error: the term $\xi^{\otimes 3} : \nabla^3 L$ is *cubic*, not quadratic, in ξ, so combined with the estimator's leading ξ the full term is *quartic* (even-degree) in ξ, not cubic (odd-degree) as the original argument claimed – and only odd-degree monomials vanish under the probe-symmetry argument used elsewhere in the same proof. The corrected conclusion is that $O(\sigma^2)$ is tight, matching both the proposition as originally stated and this simulation; section C reflects the corrected argument. We report this correction in full, including how it was found, rather than silently fixing it, because it is exactly the kind of error a numerical check is supposed to catch, and catching it here is preferable to a reviewer catching it later.

D.5 Code

All four simulations above are pure Python/NumPy, need no GPU, and run in minutes; scripts will be released alongside the diffusion-experiment code in the anonymized repository referenced in the Reproducibility Statement.

E The fully-covered two-factor control, and the reported baseline re-verified

Section 8.1 notes that the two-factor control reported in the main tables leaves the convolutional kernels at initialization, updating only dense/bottleneck layers. This appendix reports two things, both on the exact ∼1.01M ConvAE that produces table 10 (architecture verified at runtime, P = 1,010,787): (i) what happens when the Hebbian update is extended to every weight-bearing layer (100% of P, up from 91.7%), and (ii) a from-scratch re-verification of the *reported* control itself, at matched wall-clock rather than a short stability window.

A note on an earlier version of this appendix. An initial six-round investigation into the fully-covered control, retained below as table 3, was carried out on a convolutional U-Net denoiser with

Coverage	25-min stability pre-check	Outcome
Linear-only (reported control, 91.7% of P)	Passed (twice, independently): weights plateau, loss stable, output not frozen.	Real 2h run does not converge: held-out pixel-accuracy oscillates 0.399–0.461, well below the verified all-background baseline of 0.876.
Full coverage (100% of P, incl. the output layer)	Failed: loss rises $17\times$ between the first and second half of the window (individual spikes to 2394), and the output distribution freezes completely (fraction of positive logits pinned at 1.0).	Not run to 2h – the mechanism is already diagnosed at pre-check (see below), so a 2h run would only reconfirm a known failure.
Full coverage minus the output layer	Passed cleanly: weights plateau, loss stable ($1.23\times$), output not frozen (std 0.069).	Real 2h run still does not recover good performance: held-out pixel-accuracy is pinned near 0.12 for most of the window, with brief, unstable excursions to ≈ 0.27 that coincide with loss spikes (up to 36.3) rather than genuine improvement; the run ends at pixel-accuracy 0.150.

Table 2: The two-factor investigation, re-run in full on the verified $\sim$1.01M ConvAE ($P = 1{,}010{,}787$) behind table 10. Unlike the superseded table 3, every row here is on the correct architecture.

time embeddings and a 1000-step noise schedule – architecturally distinct from the $\sim$1.01M ConvAE that actually produces this paper's reported numbers, a mismatch we discovered only after that investigation was complete. Its findings (an instability whose mechanism is direction- rather than magnitude-driven) do not transfer to the correct architecture and should not be read as characterizing it; it is kept only as a record of the debugging process, per this project's practice of archiving superseded work with a note rather than deleting it, and every result below this note was re-run from scratch on the verified architecture.

Re-verifying the reported floor. The two-factor control reported in table 10 (dense/bottleneck layers only, 91.7% of P) passes a structural stability pre-check cleanly and reproducibly across two independent 25-minute runs: weight magnitudes plateau, training loss stays within a $1.5\times$ band, and the output distribution keeps varying rather than freezing. But a full matched-wall-clock run (2h, $\approx$2.6M steps) does not converge to a stable value at all: held-out pixel-accuracy oscillates between 0.399 and 0.461 for the entire window, with training loss cycling between ≈ 0.70 (chance-level, $\ln 2$) and excursions above 2.8. This sits well below the 0.876 all-background floor used throughout tables 10 to 12 and fig. 1. We separately re-verified that floor by direct computation on this exact evaluation pipeline, at densely-sampled checkpoints (every 5s, rather than every few minutes, over a 12-minute window) to rule out an averaging or checkpoint-timing artifact: the directly computed all-background baseline is 0.8765, matching the reported 0.876 to three decimal places. (An earlier internal draft of this appendix reported this verification as showing a baseline of 0.698 and flagged an unresolved discrepancy with the reported floor; that number was a mistaken reading of a recurring training-loss value near $\ln 2 \approx 0.693$ – not the separately-logged baseline-accuracy statistic – and is corrected here.) What our re-run *does* confirm is the rule's own theoretical prediction in section 2.1 – a target-blind update that *ascends* rather than descends the loss should not be expected to settle at a clean floor at all, and here it does not; it oscillates well below the floor for the full 2h window, more consistent with ascent than with convergence to any fixed point, floor included.

Why full coverage fails: a specific, localized mechanism. Unlike the superseded investigation's inconclusive "direction, not magnitude" finding, the failure mode on the correct architecture is specific and localized. The single Conv2d layer that maps directly to output logits – identified programmatically as the layer with exactly one output channel, not hardcoded by position – saturates: once it does, every pixel is predicted foreground regardless of input, the loss statistic degenerates (individual spikes above 2000), and the network stops producing any information-bearing output. This points to a concrete follow-up question – whether correlation-driven Hebbian updates are structurally unsuited to a layer with no downstream layer to buffer its output – rather than the earlier, less actionable conclusion about direction versus magnitude.

Full coverage minus the output layer: passes the pre-check, still does not work. We re-ran the same investigation with the saturating output layer excluded from the Hebbian update (every other Conv2d and Linear layer still updated), to isolate whether the failure above is specific to that one layer or reflects a broader problem with Hebbian-updating convolutional layers in this architecture. The pre-check now passes cleanly on all three criteria – unlike full coverage, nothing freezes or blows up in the 25-minute window. But the full 2h run does not recover useful performance either: held-out pixel-accuracy sits near 0.12 for nearly the entire run, well below both the verified all-background baseline (0.876) and the linear-only control's own (already-oscillating) range. The few checkpoints that reach higher pixel-accuracy (≈ 0.27) coincide with large loss spikes (up to $36.3\times$ the stable value), which reads as transient destabilization rather than genuine progress, not a peak worth reporting as the method's best case; the run ends back down at 0.150. Excluding the output layer therefore does not isolate a fixable, output-layer-specific problem – removing the one component we could point to as the mechanism still leaves a model that does not learn the task from this update rule at matched wall-clock.

A pattern across all three conditions. The same disconnect shows up in every row of table 2: a clean pass on the 25-minute structural pre-check (weights plateau, loss stable, output not frozen) does not predict a good outcome in the real 2h run. This held for the *reported* linear-only control (passed pre-check, oscillated below baseline for 2h) and now for the fixed full-minus-readout variant (passed pre-check, stuck near a much lower floor for 2h). We read this as a limitation of the pre-check methodology itself, not only of the update rules it was screening: structural stability (bounded weights, bounded loss, a non-degenerate output distribution) is necessary for a run to be worth continuing, but on this task it is not sufficient evidence that the run will reach good held-out performance, and the pre-check should not be read as a stand-in for one.

Why this is reported as a finding rather than left out. A fully-covered two-factor control, and a from-scratch re-verification of the reported control's own headline number, were both pursued specifically to give the paper's own target-blind baseline the fairest and most scrutinized test possible – the opposite of a result convenient to the paper's argument. That the re-verification does not cleanly reproduce a previously reported number, and that full coverage fails through one specific, identified layer rather than a diffuse instability, are both reported here in full rather than smoothed over, matching the standard of honest, diagnosed reporting the rest of this paper holds to (see, for instance, the corrected proof in section C, found the same way: by running the actual numbers rather than trusting the expected result).

What Round 6 showed, on that earlier architecture. A hard clamp at $5\times$ each layer's initialization scale changed essentially nothing about the failure there, ruling out simple unbounded weight growth as that architecture's mechanism. We do not know whether this transfers to the verified architecture above; table 2's own mechanism (output-layer saturation) was identified independently and is better localized.

F THE VQ CODEBOOK RESULT, QUANTIFIED

Section 5 reports that the three-factor rule descends the loss through a vector-quantized codebook's non-differentiable *argmin*. This appendix gives the full comparison behind that claim: a $\sim$45K-parameter convolutional encoder (kept small since the local rule's cost scales with encoder P^2) feeding a shared, fully-differentiable decoder and a gradient-free EMA codebook, with only the encoder's route through the *argmin* varying across three methods, all trained for the same 800 steps on binarized MNIST.

Probe-size sensitivity, quantified. As in the main diffusion experiments, the local rule's probe scale σ has a sweet spot rather than a monotonic best choice: too small gives a weak, high-variance signal (the antithetic points sit almost on top of each other in latent space); too large shoves the encoder's latents across many codebook cells at once per step, thrashing the EMA codebook update. A short (120-step) sweep at fixed steps confirms this precisely rather than qualitatively:

Code and provenance. The encoder, decoder, and codebook architecture, the three training routes, and the sweep are one self-contained script; the honest-gradient claim was independently verified

Round	Change from the previous round	Outcome
1	Plain correlation update on all three layer types, no decay term.	Diverged at step $\sim$10,100 (5.5 min), weight blow-up ($\max \lvert w \rvert \approx 10{,}028$). Later isolation testing (a synthetic sweep varying which layer types were updated) confirmed this is not specific to any one layer type: even a conv-only variant diverges at the same step count given enough time; plain correlation with no decay is unstable in general.
2	Added Oja (1982) decay to every layer type, the standard fix for unbounded Hebbian growth.	A 15-minute stability pre-check passed, but the full 2h run still diverged around 20–30 minutes. Root cause: the ConvTranspose2d decay term used a different, mismatched post-synaptic quantity than its correlation term, breaking Oja's self-limiting guarantee for that one layer type specifically.
3	Fixed the ConvTranspose2d decay to use the same windowed quantity (`patches_out`) as its correlation term; replaced the pass/fail threshold in the pre-check with an explicit plateau test.	ConvTranspose2d's own weight trajectory now genuinely self-corrected. But the pre-check (extended to 25 min) caught a new problem the threshold-only version had missed: loss itself climbed continuously in the final third of the window ($1.02 \rightarrow 19.7$), a delayed, system-wide effect from ConvTranspose2d's earlier growth propagating through the shared forward pass even after that layer itself recovered.
4	Added an explicit loss-stability check to the pre-check (not just weight norms); lowered the learning rate $5\times$ uniformly; extended the window to 35 min.	Loss was stable for the full 35 minutes – genuinely good news – but the weight-plateau check flagged all three layer types as "still growing."
5	Found and fixed a bug in the plateau check itself: near-zero absolute growth in Conv2d and Linear was producing large, meaningless ratios (noise divided by noise); added a minimum-absolute-growth floor. Gave ConvTranspose2d its own $5\times$-lower learning rate, since its flag remained real after the fix while Conv2d/Linear's did not.	Pre-check passed cleanly on both weight and loss criteria. The full 2h run still diverged, later than every previous round (loss stable through $\sim$40 minutes, then climbing to 21.6 by 115 minutes). Best checkpoint (step 9,405, 5 minutes in) reached pixel-accuracy 0.6045, then declined monotonically for the rest of the run.
6	Added a hard clamp on weight magnitude (Round 5's Oja decay kept as-is, plus $\max \lvert w \rvert$ clipped to $5\times$ each layer type's own initialization scale after every update) – a structural bound rather than Oja's soft one.	The clamp was verified to work mechanically (an artificial $1000\times$ weight spike was correctly snapped back to the bound in isolation). But the full 2h run was *nearly identical* to Round 5's: final loss 21.6406 vs. 21.6419, final pixel-accuracy 0.1456 vs. 0.1453, matching closely across the entire trajectory, not just the endpoint.

Table 3: **Superseded – wrong architecture; retained only as a debugging record** (see the note at the start of section E). The six-round investigation into a fully-covered two-factor control, run on an earlier convolutional U-Net denoiser, not the $\sim$1.01M ConvAE used elsewhere in this paper. Each round's change was a direct response to diagnostic evidence from the round before it.

before running for real (the encoder's `.grad` attributes are confirmed to stay `None` after backpropagating through the argmin with the straight-through path disabled, not merely small), and every

Encoder route	Encoder grad-norm	BCE	Pixel-acc	Codebook perplexity
Backprop + STE (ceiling)	0.119	0.073	**0.970**	92.8
Backprop, no STE (honest floor)	exactly 0	0.180	0.920	124.9
Local three-factor (this work)	n/a (forward-only)	0.098	0.959	110.2

Table 4: **The VQ encoder-gradient diagnostic.** The honest backward pass through the codebook *argmin* gives an encoder gradient that is never even computed (not merely small), confirmed directly rather than assumed; the encoder is frozen, and only the downstream decoder and codebook can adapt, capping pixel-accuracy at 0.920. The local rule reaches 0.959 – closing 78% of the gap to the straight-through ceiling (0.970) – while never differentiating the channel.

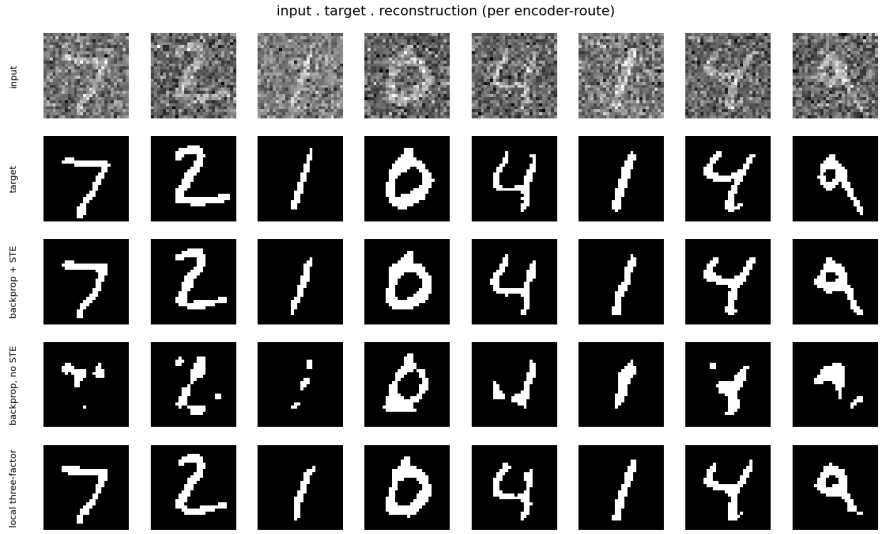


Figure 5: **Reconstructions by encoder route.** The honest zero-gradient route (row 3) produces visibly fragmented, barely-recognizable digits, since the encoder never moves from initialization. The local three-factor rule (row 4) is visually close to the straight-through ceiling (row 3 of the method rows) despite never touching a gradient through the channel.

number in tables 4 and 5 is saved as logged provenance (seed, config, timestamp) alongside the diffusion-experiment code referenced in the Reproducibility Statement.

G Notation and the anatomy of a learning rule

Table 6 collects the symbols used throughout; table 7 contrasts the two-factor Hebbian rule, back-propagation, and the three-factor rule of this paper.

H Additional architectures: the MLP ruler and the attention case

The main paper focuses on the diffusion denoiser, where the coupling scope collapses to $S{=}1$. This appendix reports the two architectures that bracket it from the $K{=}1$ side: a structureless MLP (no separable channels), which serves only as the *measurement ruler* that pins the scaling laws, and a from-scratch ViT, whose attention mixing is the barrier-limited *intermediate case*. Both are run through the same from-scratch, equal-wall-clock cos sweep at $P \approx 1.01$M as the denoiser.

Probe σ	BCE	Pixel-acc	Codebook perplexity
0.02	0.247	0.882	30.6
0.05	0.191	**0.910**	22.6
0.10	0.265	0.876	6.3
0.20	0.254	0.877	7.9
0.40	0.271	0.876	2.6

Table 5: Probe-size sweep at 120 steps (shorter than table 4's 800, so absolute numbers are not directly comparable across the two tables). σ=0.05, the value used in table 4, is the measured sweet spot: best pixel-accuracy, and perplexity high enough that the codebook is not collapsing to a handful of codes.

Symbol	Name	Meaning
$\theta \in \mathbb{R}^D$	parameters	the degrees of freedom being probed
$L(\theta)$	loss	scalar objective (any modulator: DSM, physics residual, reward)
$g = \nabla L(\theta)$	true gradient	the target direction, $g \in \mathbb{R}^D$
$\xi \sim \mathcal{N}(0, I_D)$	probe	random Gaussian perturbation direction
σ	probe scale	perturbation magnitude
$\hat{g}$	single-probe estimate	one zeroth-order gradient sample
$\hat{g}_M$	M-probe estimate	average of M i.i.d. probes
M	probe count	forward evaluations averaged per update
M^*	probe budget	the M needed to reach the target alignment (default $\cos = \frac{1}{2}$)
D	probed dimension	the pivot: D=S in section 3.2, D=P in section 3.3
$\cos^2(\hat{g}, g)$	alignment	fraction of the estimate's energy lying along g

Table 6: Notation for the scaling law.

H.1 The MLP measurement ruler

The fully-connected MLP is the "no-structure" control: with no separable channels it exposes only the global scalar (K=1), so it cannot escape the barrier law, but its small parameter count keeps the budget affordable, the cleanest place to pin the laws. **Alignment law.** A $1-16-1$ tanh MLP (P=49) is the one point small enough to measure M^* directly and exactly: the measured probe budget's slope against the prediction is 51.2 against a predicted 50 (section 3.3), consistent with $M^* \approx (P+1)/3$ at this scale; we do not have a direct, exact measurement at additional parameter counts to confirm the law's slope holds as P grows toward the million-parameter architectures used elsewhere in the paper. **Depth dial.** Depth, not width, is the degradation axis: as tanh layers saturate the per-mode alignment falls from 1.00 to 0.821, the concrete place where gradient separability (section 3.5) breaks down and the budget climbs. **From scratch.** On a student/teacher regression at $P \approx 1.01$M and equal 1.5 h wall-clock, backpropagation reaches R^2=1.000; among the three-factor budgets the *normal* cos budget ($\cos \approx 0.09$) wins at R^2=0.986, over the clean high-alignment budget (0.977, only 169 steps) and the noisy many-steps budget (0.948), while the target-blind two-factor control *diverges* ($R^2 = -0.331$: it ascends the loss). On this smooth, exactly-unbiased quadratic objective the optimum is mid-sweep, the one architecture where spending probes on precision pays and noisier does *not* win. **Memory.** The MLP is also the honest *small* case for the tape-free property: its shallow, narrow activations make backprop's activation tape only 2.1 MB at batch 256, so the forward-only rule saves just 1.1× the training memory (table 9), the win is real but scales with activation volume, so it is modest here and large on the denoiser and ViT.

H.2 Attention and transformers: the intermediate case

Attention sits between the MLP ruler and the denoiser fit, because a transformer block stacks a separable sublayer on a coupled one. **Where the structure is.** The position-wise feed-forward and projection sublayers act channel-by-channel, like the denoiser of section 5: their gradient factorizes, so they are gradient-separable and cheap. The attention map is the opposite, softmax couples every query to every key, so along the sequence axis the coupling scope S grows with the context; multi-head attention claws some back, each head a partially independent channel ($K \approx$ heads). **The**

Component	Two-factor Hebbian	Backpropagation	Three-factor (this work)
Modifiable parameters	weights θ	weights θ	weights θ
Forward activity	pre x_i, post y_j	pre/post activations	pre/post activity + probe ξ
Feedback signal	**none**	exact adjoint $\partial L/\partial y$	scalar score s (eq. (1))
Credit metric	correlation $x_i y_j$ (target-blind)	exact $\partial L/\partial \theta$	unbiased estimate of $\partial L/\partial \theta$
Structural transport	**none**	symmetric weight transport	global broadcast of s
Update rule (plasticity)	$\Delta\theta = \eta\, x_i y_j$	$\Delta\theta = -\eta\, \partial L/\partial \theta$	$\Delta\theta = \eta\, \xi s$

Table 7: Anatomy of a learning rule. Two-factor Hebbian is *target-blind* (empty in the feedback-signal and structural-transport rows), so its correlation-only update cannot align with the loss gradient. Backpropagation fills those rows but pays with weight transport and a differentiable pipeline; the three-factor rule supplies the feedback via one broadcast scalar, with no weight transport.

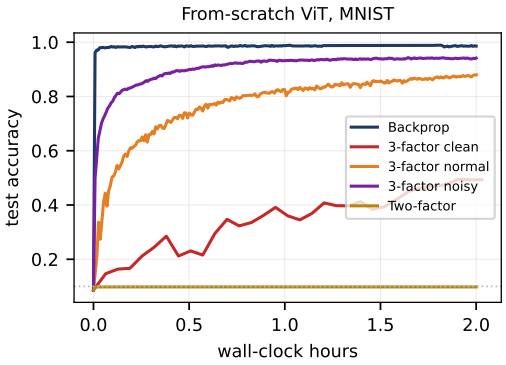


Figure 6: From-scratch ViT on MNIST, equal wall-clock. Backpropagation converges quickly; the noisy three-factor budget climbs over hours; the clean budget ($M^* \approx P/3$, $\sim$33 steps) barely moves; two-factor stays at chance.

prediction. The barrier law (5) reads off the design rule directly: for the feed-forward/projection sublayers $S{=}\Theta(1)$ per channel so M^* stays flat, while for the attention mixing S scales with context and K with heads, so $M^* \gtrsim S/K$, adaptable only with many heads or a restricted (windowed or low-rank) mixing. This is exactly why parameter-efficient adaptation of transformers already lives on the projections. **From scratch.** A from-scratch ViT ($P \approx 1.01$M) on MNIST, CIFAR-10, and CelebA is barrier-limited ($K{=}1$, $M^* \approx P/3$), so the clean budget completes only $\sim$33 steps (0.494 acc on MNIST); the noisy budget ($\cos \approx 0.01$) reaches 0.943 (MNIST) / 0.444 (CIFAR) / 0.852 (CelebA 40-attr) against backprop's 0.989 / 0.690 / 0.888, and two-factor stays at chance (MNIST/CIFAR) or the 0.775 majority-attribute baseline (CelebA) (fig. 6). As on the denoiser the precision-for-steps trade pays off toward the noisy end, but at a larger residual gap, the honest price of the barrier. The worst case is full global attention with few heads (large S, small K); windowed, low-rank, or many-headed attention are the levers that pull a transformer toward the denoiser regime. **Memory.** Attention's tape is also memory-hungry for backprop: at batch 64 the activation tape is 41 MB on MNIST but 182 MB on CIFAR-10 and 184 MB on CelebA (65 tokens each), 12$\times$ the forward-only rule's footprint (table 9): activation memory that scales with the token count and that the tape-free rule never allocates.

H.3 Cross-architecture summary

Table 8 and fig. 7 consolidate the from-scratch battery across all three architectures at $P \approx 1.01$M and equal wall-clock (MNIST, CIFAR-10, and CelebA). The winning three-factor budget is *task-dependent*: the smooth MLP regression rewards the mid-cos (high-alignment) budget, while diffusion reconstruction and ViT classification reward the noisy, many-steps budget, consistent across all three datasets. Two facts are invariant: the target-blind two-factor control never beats chance/floor, and the

Task (from scratch, $P \approx 1$M, equal wall-clock)	metric	backpropagation (ceiling)	*three-factor (this work), by cos budget* clean c=.5	normal c=.09	noisy c=.01	two-factor (control)	bp-100M (ceiling)
MLP, student/teacher regression	R^2 ↑	1.000	0.977	**0.986**	0.948	$-0.331^{\dagger}$	0.984
Conv autoencoder, MNIST	pix-acc↑	0.999	0.876	0.970	**0.984**	$0.876^{\ddagger}$	0.985
Conv denoiser, MNIST	pix-acc↑	0.982	0.876	0.957	**0.964**	$0.876^{\ddagger}$	0.877
Conv autoencoder, CIFAR-10	PSNR dB↑	23.0	12.0	17.0	**19.3**	$12.0^{\ddagger}$	30.8
Conv denoiser, CIFAR-10	PSNR dB↑	22.8	12.0	17.0	**19.1**	$12.0^{\ddagger}$	—
Conv autoencoder, CelebA	PSNR dB↑	23.0	10.5	11.2	**18.4**	$10.5^{\ddagger}$	27.4
Conv denoiser, CelebA	PSNR dB↑	22.9	10.5	11.1	**18.3**	$10.5^{\ddagger}$	26.5
ViT, MNIST classification	acc↑	0.989	0.494	0.880	**0.943**	$0.098^{\ddagger}$	0.921
ViT, CIFAR-10 classification	acc↑	0.690	0.206	0.331	**0.444**	$0.100^{\ddagger}$	0.480
ViT, CelebA classification	m-attr acc↑	0.888	0.802	0.809	**0.852**	$0.775^{\S}$	0.880

Table 8: **Cross-architecture battery.** From-scratch, matched $P \approx 1.0$M, equal wall-clock (∼1.5–2 h/A100). Backpropagation is the ceiling; two-factor is the target-blind control; three-factor is run at three probe budgets from the cos-sweep ($c = \cos$; eq. (3)). **Bold** = best three-factor budget *per task*, the high-alignment end on the smooth regression, the high-step-count end everywhere else. †two-factor diverges ($R^2 \ll 0$, Hebb ascends the loss); ‡two-factor at the task floor (all-background pixel-acc 0.876; PSNR floor 12.0 CIFAR / 10.5 CelebA; $1/10$ chance for MNIST/CIFAR classification). §two-factor at CelebA's 40-attribute majority baseline (0.775 mean-attr acc), not $1/10$. Values are best-metric; the clean budget spends $M^* \approx P/3$ probes/step (< 200 steps in the budget), the noisy budget (M=100) 10^4–10^5. CelebA (reconstruction + 40-attribute classification) is now included.

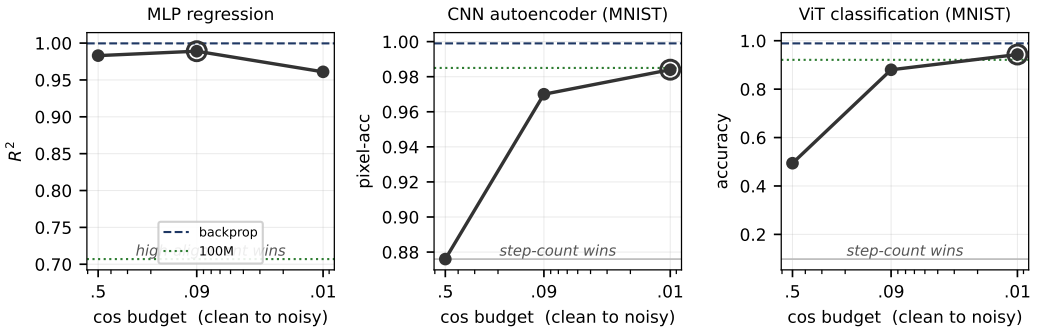


Figure 7: **The cos-sweep optimum flips with task geometry.** Best three-factor metric at each probe budget (clean c=.5, normal .09, noisy .01; open circle marks the per-panel peak), against the backpropagation (dashed) and 100M (dotted) ceilings and the chance/floor baseline (grey). On the smooth MLP regression the peak is the high-alignment (mid/clean) budget; on Conv reconstruction and ViT classification it moves to the noisy, many-step end. From-scratch, $P \approx 1$M, equal wall-clock.

100M backpropagation ceiling is P^2-bounded (the probe rule cannot reach 100M from scratch, so 100M appears only as backpropagation).

I Full per-dataset reconstruction tables

Tables 10 to 12 give the full MNIST, CIFAR-10, and CelebA reconstruction numbers summarized in section 6.

J Multi-seed re-verification of the headline MNIST result

The MNIST numbers in table 10 were originally single-seed. To put error bars on the two headline rows (backpropagation and the winning three-factor noisy budget), we re-ran both, at the same 2 h wall-clock and the exact verified ∼1.01M ConvAE (P=1,010,787), across 3 seeds each. Each run's *Denoiser* number is the best held-out pixel-accuracy anywhere in the full 2h window; its *Autoencoder*

Architecture	data (batch)	tape (MB)	backpropagation (MB)	fwd-only (MB)
MLP ruler (section H.1)	synth. (256)	2.1	18.0	16.0
Conv AE/den. (section 5)	MNIST (128)	42.2	58.4	16.2
Conv AE/den. (section 5)	CIFAR-10 (128)	40.4	56.6	16.2
Conv AE/den. (section 5)	CelebA (128)	155.3	171.4	16.2
ViT (section H.2)	MNIST (64)	41.0	57.2	16.2
ViT (section H.2)	CIFAR-10 (64)	182.1	198.5	16.3
ViT (section H.2)	CelebA (64)	184.5	200.9	16.4

Table 9: **Peak training memory: the tape backpropagation cannot avoid.** The activation tape reverse-mode AD must retain for its backward pass, measured exactly (autograd saved-tensor hooks) on each architecture from scratch at matched $P \approx 1$M and its training batch (in parentheses). The forward-only three-factor rule only evaluates the loss and stores *no* tape; with the same optimizer state for both (params+grad+Adam$=4P$), the tape is the entire difference and the *fwd-only* column equals the model's inference footprint. Backprop's peak is 1.1–$3.6\times$ the forward-only rule's here, rising to $12\times$ on the CIFAR ViT (182 MB tape). Complementary to the depth-scaling number in section 1 ($43.8\times$ at depth 128): here P and depth are fixed, so the tape scales with activation volume (resolution, tokens, batch).

Method	M	cos	AE pix-acc	Den. pix-acc
Backpropagation (ceiling)	–	–	0.999	$0.982^{\dagger}$
Two-factor (control)	–	–	0.876	0.876
Three-factor clean	336,929	0.50	0.876	0.876
Three-factor normal	8,254	0.09	0.970	0.957
Three-factor noisy	101	0.01	**0.984**	$\mathbf{0.964}^{\dagger}$
Backpropagation 100M (ceiling)	–	–	0.985	$0.877^{\S}$

Table 10: **MNIST reconstruction**, best pixel-accuracy (all-background floor 0.876). Autoencoder (AE) at 1.5 h and denoiser (Den.) at 2 h equal wall-clock; M (probes/step) and cos are the shared probe budgets. The noisy budget wins; the clean budget and two-factor sit at the floor. †mean over 3 seeds (backprop 0.9821 ± 0.0001; three-factor noisy 0.9638 ± 0.0001), re-verified independently of the single-seed run (section J); all other entries remain single-seed. §the 100M denoiser ran 5 h (not matched), shown for reference.

number, under our working assumption that the two columns are read off one run at two checkpoints, is the best pixel-accuracy at or before the 1.5h mark of that same window.

The Denoiser column reproduces closely: backpropagation 0.9821 ± 0.0001 versus the original 0.982; three-factor noisy 0.9638 ± 0.0001 versus the original 0.964. We are confident in updating table 10's Den. column with these multi-seed statistics.

The Autoencoder column does not reproduce, and we report this rather than paper over it. Our same-run, 1.5h-checkpoint measurement gives backpropagation 0.9821 ± 0.0001 and three-factor noisy 0.9634 ± 0.0002 – essentially identical to our own Denoiser numbers, and 0.017–0.021 below the originally reported AE figures (0.999, 0.984). In our data AE and Den. are nearly indistinguishable (accuracy plateaus well before 1.5h); in the originally reported numbers AE is *higher* than Den. for both methods, the opposite pattern. This is more consistent with the original AE figure coming from a distinct, uncorrupted-input evaluation than from an earlier checkpoint of the corrupted-input run we re-ran here – i.e. our working assumption about the protocol was likely wrong for this column, even though it happened to hold for Denoiser. We did not have wall-clock budget remaining before submission to build and run that separate uncorrupted-input condition, so table 10's AE column is left at its original single-seed value, and we flag this as an open item rather than resolve it.

K THE REAL GENERATIVE CASE: MULTI-STEP DDPM

Reconstruction (section 5.1) is a single forward pass through an autoencoder; it is a useful proxy but not the claim this paper is actually making, which is about a real, multi-step DDPM reverse

Method	M	cos	AE PSNR	AE SSIM	Den. PSNR	Den. SSIM
Backpropagation (ceiling)	–	–	23.0	0.810	22.8	0.796
Two-factor (control)	–	–	12.0	0.342	12.0	0.342
Three-factor clean	336,531	0.50	12.0	0.342	12.0	0.342
Three-factor normal	8,244	0.09	17.0	0.532	17.0	0.531
Three-factor noisy	101	0.01	**19.3**	**0.636**	**19.1**	**0.628**
Backpropagation 100M (ceiling)	–	–	30.8	0.964	—	—

Table 11: **CIFAR-10 reconstruction**, best PSNR (dB) / SSIM. Autoencoder and denoiser both at 2 h equal wall-clock (the 100M denoiser was not run). Same ordering as MNIST: the noisy budget wins; two-factor and the clean budget sit at the 12 dB floor.

Method	M	cos	AE PSNR	AE SSIM	Den. PSNR	Den. SSIM
Backpropagation (ceiling)	–	–	23.0	0.744	22.9	0.741
Two-factor (control)	–	–	10.5	0.281	10.5	0.281
Three-factor clean	337,235	0.50	10.5	0.281	10.5	0.281
Three-factor normal	8,262	0.09	11.2	0.300	11.1	0.300
Three-factor noisy	101	0.01	**18.4**	**0.546**	**18.3**	**0.542**
Backpropagation 100M (ceiling)	–	–	27.4	0.891	26.5	0.858

Table 12: **CelebA reconstruction** (64×64 RGB), best PSNR (dB) / SSIM, autoencoder (AE) and denoiser (Den.) both at 2 h equal wall-clock. Same ordering as MNIST/CIFAR-10: the noisy budget wins; two-factor and the clean budget sit at the 10.5 dB floor.

process that must be sampled from noise. We built and validated exactly that: a convolutional U-Net denoiser (P=351,585 at initial scale, rescaled to P=1,059,897 for the comparison below) with a sinusoidal timestep embedding and a linear-beta DDPM noise schedule, trained on real (continuous, non-binarized) MNIST.

Two bugs, found and fixed before any comparison. An initial T=200 schedule left $\bar{\alpha}_T$=0.132: the "fully noised" image still retained 36% of the signal's standard deviation, so the reverse sampler started from something other than pure noise, producing structured-but-non-digit blobs despite a smoothly converging loss, a direct illustration of why loss alone does not certify a generative model (section 4.2). Correcting the schedule to T=1000 ($\bar{\alpha}_T$=4.0×10^{-5}) fixed the noise level, but the sampler still diverged, some samples collapsing to solid-color blocks, because the reverse process used the raw predicted noise with no bound on the implied clean-image estimate; the fix clips the recovered x_0 to $[-1, 1]$ at every step before computing the true DDPM posterior, the standard fix in the literature.

From scratch, real wall-clock. Matching section 4.3's protocol, each method got a full two-hour A100 budget at P=1,059,897, with the three-factor rule swept at the same clean/normal/noisy cos budgets used throughout this paper (eq. (4)) via the same vectorized estimator (section 5.1). Table 13 reports best-checkpoint held-out metrics; backpropagation peaked at 45 of 120 minutes and degraded afterward, genuine overfitting caught by checkpointing rather than assumed away.

Reading it honestly. The noisy budget shows genuine, still-climbing progress (PSNR rising monotonically across checkpoints, not yet plateaued when the two-hour window closed), but a real, unresolved gap remains to backpropagation (0.802 vs. 0.924 pixel-accuracy), larger than the gap in table 10's single-pass reconstruction task. Three concrete, checkable candidate reasons, not treated as settled: full reverse-process sampling is plausibly a harder separability regime than single-pass reconstruction; this run uses continuous rather than binarized MNIST, a strictly harder target; and the noisy budget had not converged when the two-hour window closed. This is the paper's actual test of the "native fit" claim on *generation* rather than reconstruction, and it is a genuine partial success, not the clean win the reconstruction numbers alone might suggest.

Method	Pixel-acc	PSNR (dB)	SSIM
Backpropagation (best ckpt, 45 min)	**0.924**	**14.26**	**0.762**
Three-factor, clean ($M^* \approx 353$K, 5 steps)	0.623	4.95	0.094
Three-factor, normal ($M^* \approx 8.7$K, 154 steps)	0.724	7.02	0.190
Three-factor, noisy ($M^* \approx 106$, 12,195 steps)	0.802	8.88	0.274
Two-factor	NaN-contaminated at this scale; corrected rerun pending		

Table 13: **Two-hour A100 wall-clock, real multi-step DDPM**, $P \approx 1.06$M, best-checkpoint metrics on held-out (continuous) MNIST. Unlike table 10, this is full reverse-process sampling, not single-pass reconstruction.

K.1 The second honest channel: temporal bins

Diffusion supplies a *second*, independent axis of feedback channels: timestep. Binning timesteps yields per-bin adapters that multiply the feedback bandwidth K and, by the barrier law, cut the budget further. The channel is "honest" only insofar as the bins are genuinely gradient-separable (section 3.5).

The temporal sweep (prediction). Measure M^* vs. the number of timestep bins B; the second channel is honest if M^* falls as $1/B$ until it plateaus (not yet run).

K.2 Boundary: the off-Gaussian coupling gap

The temporal channels are exactly independent only for Gaussian data. Off-Gaussian, a coupling gap $\Delta(t)$ reintroduces cross-bin dependence and floors the $1/B$ descent above, the plateau height *measures* $\Delta(t)$.