A TRAIL FROM LOW-PRECISION SAMPLING: DIAGNOSING SUPER-DIFFUSIVE FP32–BF16 DRIFT IN DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Low-precision inference is now the default deployment path for diffusion image generators: BF16 is standard and FP8 is spreading fast. The rounding error this introduces is usually treated as small, zero-mean noise that averages out over the tens of denoising steps of a sampler. We show it does not average out. Running SDXL from an identical seed in FP32 and BF16, we find that the same-seed latent trajectories do *not* wander apart like independent noise: their per-step L_2 divergence grows *super-diffusively*, with a log–log growth exponent of 1.04 (diffusive random-walk null $= 0.5$) and a locatable takeoff around step 25 of 50. This is a robust same-seed phenomenon (n=64 pairs), and we pin down its cause with an unusually thorough elimination: five controls that rule out the tempting statistical explanations one at a time. It is *not* a mean bias (sign-bias statistic $|b| \approx 0.012$). It is *not* the error's step-to-step directional correlation: although consecutive error vectors are nearly collinear ($\cos(e_t, e_{t+1}) = 0.971$), that merely rides the trajectory's own velocity autocorrelation (0.979), and a magnitude-matched but *uncorrelated* Gaussian perturbation injected afresh at every step reproduces the super-diffusion (exponent $1.01 \approx 1.04$). And it is *not* a fixed-magnitude dynamical amplification: the same continual injection at *constant* per-step magnitude is merely diffusive (0.43), while a one-time kick is damped (0.19). What remains is a *growing-magnitude positive feedback*: the per-step rounding error scales with the accumulating separation, so as the same-seed trajectories diverge the error grows and drives them further apart—faster than $\sqrt{t}$, independent of the error's mean or correlation. Crucially, we do not stop at inferring this feedback but measure it directly: regressing the per-step divergence increment on the current divergence is decisively positive (Spearman 0.89, positive for all 64 of 64 prompts). We package these measurements as DRIFTPROBE, a training-free diagnostic—usable on any UNet or DiT block—that reports the per-step sign bias, the same-seed divergence curve and its onset t^*, and the injection controls that isolate the mechanism. We treat the drift as a lens, not a defect to correct, and lay out how t^* could serve as an early signal for downstream generation defects. Results are measured on SDXL/BF16 on a single 48 GB GPU (n=64 same-seed pairs; n=16 per injection control); the FP8 arm, the FLUX backbone, and the defect-prediction evaluation are scoped as future work.

1 INTRODUCTION

Diffusion models are almost never run in full precision in production. To fit a model on a single consumer GPU and to hit interactive latencies, practitioners cast weights and activations to BF16 as a matter of course, and increasingly to FP8. The received wisdom is that the resulting rounding error is a small, zero-mean perturbation: each denoising step rounds a slightly different value, the errors are independent across steps and channels, and over the tens of steps of a sampler they wash out. Under that model of the world, low precision costs a little fidelity and nothing structural.

We take the opposite view, and we back it with measurement. Running SDXL from an identical seed in FP32 and BF16, we track the per-step L_2 divergence $D_t = \|x_t^{\mathrm{BF16}} - x_t^{\mathrm{FP32}}\|_2$ between the two same-seed latent trajectories. Under the independent-noise model D_t should grow like a random

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.

walk, $\sqrt{\text{steps}}$ (log–log slope $\frac{1}{2}$). It does not: it grows *super-diffusively*, with a median log–log growth exponent of 1.04 (interquartile range 0.85–1.17 over $n{=}64$ prompt$\times$seed pairs) and a locatable takeoff at a median step $t^* \approx 25.5$ of 50. Same-seed FP32 and BF16 trajectories *bifurcate*; they do not diffuse apart. This is our headline empirical finding.

Why? We isolate the cause by elimination, ruling out the tempting statistical explanations with five controls (detailed in Section 5). It is *not* a mean bias: the sign-bias statistic is near zero ($|b| \approx 0.012$). It is *not* step-to-step correlation: consecutive error vectors *are* nearly collinear ($\cos(e_t, e_{t+1}) = 0.971$), but this merely rides the trajectory's own velocity autocorrelation (0.979), and an *uncorrelated* perturbation injected afresh at every step super-diffuses just the same (exponent $1.01 \approx$ the real 1.04; $n{=}16$). And it is *not* a fixed-magnitude effect: continual injection at *constant* magnitude is merely diffusive (0.43), and a one-time kick is damped (0.19). What is left is a *growing-magnitude positive feedback*: the per-step error scales with the accumulating separation, so as the trajectories diverge the error grows and drives them further apart, faster than $\sqrt{t}$. And we measure this loop directly rather than infer it: the per-step divergence increment grows with the accumulated divergence (Spearman 0.89; positive for all 64 of 64 prompts).

This paper makes that story operational and asks whether the bifurcation is *useful*—whether its onset step t^* could foretell the visible failures diffusion image generators are notorious for, the six-fingered hands and melted faces. We state and test hypotheses at three levels.

- **H1 (Mechanism of accumulation).** The super-diffusive accumulation is a *growing-magnitude positive feedback*—the per-step error scales with the accumulating separation—not any special statistical structure of the rounding. *Supported; the mean-bias, correlation-specific, and fixed-magnitude accounts are all ruled out by the injection controls above, and the feedback is measured directly (Spearman 0.89, 100% of prompts positive).*
- **H2 (Accumulation and divergence).** The accumulating error makes same-seed FP32/BF16 trajectories diverge super-diffusively (growth exponent $> \frac{1}{2}$) and separate at a locatable step t^*. *Supported.*
- **H3 (Prediction).** The magnitude and timing of the divergence correlate with the final image's defect rate, and t^* precedes visible failure—making drift an actionable early-stopping signal. *Scoped as future work; specified but not yet run.*

We package the measurements behind H1–H2 as DRIFTPROBE, a training-free diagnostic that runs a prompt from a fixed seed at two precisions and reports (i) a per-step sign-bias statistic, (ii) the L_2 divergence curve and its onset t^*, and (iii) the step-to-step error correlation with the injection controls that isolate the mechanism. Nothing is trained and nothing is corrected: a lens, not a fix. Correcting the quantization error is a legitimate but different goal (PTQD (He et al., 2023), quantization-aware drift correction (Ryu et al., 2026)); a corrector that removes the drift also removes the signal we want to read.

Contributions. (1) We show, by direct same-seed measurement on SDXL/BF16, that low-precision diffusion trajectories diverge *super-diffusively*, not like independent noise (Section 5). (2) We isolate the mechanism with an unusually thorough elimination—five controls: not mean bias, not step-to-step correlation (uncorrelated continual injection reproduces the growth), not amplification of an isolated kick (damped), and not a fixed-magnitude effect (constant-magnitude injection is diffusive)—then measure the driver directly (per-step increment vs. accumulated divergence: Spearman 0.89, all 64 prompts positive): a growing-magnitude positive feedback, the error scaling with the divergence it produces (Sections 3–5). (3) We give an operational, training-free probe defined on any tracked tensor (Section 4), and specify a t^*-based defect-prediction protocol with its remaining arms (FP8, FLUX, prediction) scoped as future work.

2 RELATED WORK

Diffusion samplers and their failures. From roots in nonequilibrium thermodynamics (Sohl-Dickstein et al., 2015) and score matching (Song & Ermon, 2019), denoising diffusion models (Ho et al., 2020; Nichol & Dhariwal, 2021; Dhariwal & Nichol, 2021) and the score-SDE view (Song et al., 2021b) generate an image by iterating a learned denoiser for tens of steps; deterministic

accelerations—DDIM (Song et al., 2021a), elucidated-design samplers (Karras et al., 2022), fast ODE solvers (Lu et al., 2022)—together with latent diffusion (Rombach et al., 2022), classifier-free guidance (Ho & Salimans, 2022), and large text-to-image systems (Ramesh et al., 2022; Saharia et al., 2022; Nichol et al., 2022) made this practical. Rectified-flow / flow-matching objectives (Lipman et al., 2023; Liu et al., 2023) underlie the newest UNet (Podell et al., 2024) and transformer/DiT backbones (Peebles & Xie, 2023; Vaswani et al., 2017; Black Forest Labs, 2024; Esser et al., 2024). Greater fidelity did not remove the *structured* failures these models are notorious for—malformed hands, distorted faces, duplicated objects—which our signal aims to foresee.

Boundary 1: low-precision diffusion (*correction*, not reading). Network quantization is mature, from trained compression (Han et al., 2016; Esser et al., 2020) and post-training methods (Jacob et al., 2018; Nagel et al., 2020; Li et al., 2021; Nagel et al., 2021) to large-model quantization (Frantar et al., 2023; Xiao et al., 2023; Dettmers et al., 2022) and surveys (Gholami et al., 2022). For diffusion, PTQ4DM (Shang et al., 2023) and Q-Diffusion (Li et al., 2023) quantize the denoiser; temporal schemes (So et al., 2023; Huang et al., 2024) adapt to the changing per-step statistics; PTQD (He et al., 2023) models quantization noise as correlated plus residual and *corrects* it; and Q-Drift (Ryu et al., 2026) *compensates* the accumulation of quantization error along the trajectory. These share our starting observation—quantization error compounds over steps—but aim to *undo* the drift. We deliberately do not correct: we read the drift as a measurement channel. A corrector that succeeds *erases exactly the signal we read*.

Boundary 2: the accumulation mechanism (a controlled account). A common premise, imported from low-precision (BF16/FP8) attention and grounded in classical numerical analysis, is that rounding error is not zero-mean but *biased*, its direction-dependent bound growing with the accumulation length (Higham, 2002). Low-precision training practice—8-bit (Wang et al., 2018; Sun et al., 2019), BFLOAT16 (Kalamkar et al., 2019), modern FP8 formats (Micikevicius et al., 2022; Peng et al., 2023)—documents that naive accumulation is unstable without bias-aware scaling, consistent with the classical account (Goldberg, 1991). We carried this "accumulating error" lens into diffusion *inference* and *tested* it with injection controls rather than assuming a mechanism: the mean-bias premise does not hold (sign bias ≈ 0.012), and the natural fallback—directional correlation ($\cos(e_t, e_{t+1}) \approx 0.97$)—turns out non-causal, since a magnitude-matched *uncorrelated* continual injection reproduces the growth. We locate the driver instead in the sampler's response to *continual per-step* error injection. Neither this controlled account nor its use as a same-seed divergence signal has appeared before.

Boundary 3: early defect detection (*semantic*, not numerical). HEaD (Betti et al., 2024) and HEaD+ (Betti et al., 2026) detect hallucinations early by inspecting *cross-attention* maps for whether prompted objects have appeared—close in spirit (early internal-signal detection) but orthogonal in mechanism: they read *semantic*, prompt-conditioned evidence. DRIFTPROBE reads a *numerical* artifact—precision-induced trajectory divergence—that is agnostic to what the prompt asks and needs no cross-attention. The two are *complementary*: a present-but-malformed object is invisible to HEaD yet flagged by drift, and vice versa. In one line: prior quantization work (Boundary 1) *removes* the drift; the accumulating-error lens (Boundary 2) assumed a mean bias we find absent; early-detection work (Boundary 3) reads semantics. DRIFTPROBE instead quantifies the same-seed divergence, isolates its driver as the continual per-step re-injection of error (not its mean or correlation), and asks whether its onset t^* leads visible defects.

3 BACKGROUND

3.1 DIFFUSION SAMPLING AS A REPEATEDLY APPLIED OPERATOR

A diffusion generator iterates a learned denoiser ϵ_θ over a decreasing noise schedule. From $x_T \sim \mathcal{N}(0, I)$, a deterministic sampler such as DDIM (Song et al., 2021a) computes

$$x_{t-1} = \Phi_t(x_t, \epsilon_\theta(x_t, t, c)), \qquad t = T, T-1, \ldots, 1, \tag{1}$$

with text condition c and fixed update rule Φ_t (precision-agnostic in exact arithmetic). The structural fact we exploit: the *same* network is applied every step and x_{t-1} feeds the next evaluation, so any per-step perturbation is passed forward and re-processed, not reset—the sampler is a length-T

composition of one operator. We write $x_t^{(p)}$ for the latent under precision $p \in \{\text{FP32}, \text{BF16}, \text{FP8}\}$, at identical seed and schedule across p.

3.2 Low-Precision Arithmetic and Rounding Error

Casting to a low-precision format p replaces each exact value v by $\mathrm{rnd}_p(v) = v + \delta_p(v)$; for a single well-conditioned operation the rounding error δ_p is small and, under the standard model, treated as zero-mean noise bounded by the unit roundoff. The denoiser evaluation, however, is a long chain of fused multiply–adds—matrix products, attention softmaxes (Vaswani et al., 2017), normalizations—each accumulating many signed terms. Two *statistical* mechanisms are usually invoked to explain why such a chain would fail to average out over the sampler's steps—and a key finding of this paper is that *neither* is the operative cause.

(a) A non-zero mean (bias). Classical numerical analysis notes that the forward error of a length-n inner product can carry a systematic component whose bound accumulates with n (Higham, 2002), and low-precision quantization can add a rounding-direction bias (Nagel et al., 2021; Gholami et al., 2022). If the aggregate per-tensor error δ_p has a non-zero mean μ_p, then a length-T composition integrates μ_p into a drift that grows like $T\|\mu_p\|$. *(b) Temporal correlation.* Even a strictly zero-mean error fails to average out if its *direction* is correlated across the repeated applications: a sum of n zero-mean but strongly collinear increments has magnitude $\sim n$ (not $\sqrt{n}$). Either would give super-diffusive growth. But both are properties of the error's *statistics*, and we show by direct injection controls (Section 5) that the super-diffusion survives when the mean is removed *and* the correlation is destroyed, so long as the error is re-injected every step. The operative cause is therefore *dynamical*: continual per-step re-injection into an amplifying sampler, independent of the perturbation's mean or correlation. Testing and eliminating these two statistical routes rigorously is a central contribution.

3.3 Sign Bias and Directional Correlation

We make both notions precise at the level of a tracked tensor (a chosen latent or a chosen block's output). Let $a_{t,\ell}^{(p)}$ denote the value of tracked quantity ℓ at step t under precision p, and let $a_{t,\ell}^{(\text{FP32})}$ be the full-precision reference. The per-step, per-tensor rounding error is

$$r_{t,\ell}^{(p)} \;=\; a_{t,\ell}^{(p)} - a_{t,\ell}^{(\text{FP32})} \;\in\; \mathbb{R}^{d_\ell}, \tag{2}$$

a vector over the d_ℓ coordinates of the tensor.

Mean bias (route (a)). Under the classical independent-zero-mean model, the coordinatewise sign of $r_{t,\ell}^{(p)}$ is a fair coin: $\mathbb{E}[\mathrm{sign}(r_{t,\ell}^{(p)})] = 0$.

Definition 1 (Rounding-error sign bias)**.** For tracked tensor ℓ at step t and precision p, the *sign-bias statistic* is the mean coordinatewise sign of the rounding error,

$$b_{t,\ell}^{(p)} \;=\; \frac{1}{d_\ell}\sum_{i=1}^{d_\ell} \mathrm{sign}\big(r_{t,\ell,i}^{(p)}\big) \;\in\; [-1, 1]. \tag{3}$$

A consistent $|b_{t,\ell}^{(p)}| > 0$ across seeds—together with a large bias-vs-noise fraction $\|\mathbb{E}[r]\|/\mathbb{E}\|r\|$—would be the signature of route (a), a genuinely mean-biased rounding error.

Directional correlation (route (b)). The complementary statistic asks whether the error *direction* persists from one step to the next, independently of whether it has a mean. For consecutive per-step error vectors we define the step-to-step directional correlation

$$\rho_t^{(p)} \;=\; \cos\big(e_t^{(p)},\, e_{t+1}^{(p)}\big) \;=\; \frac{\langle e_t^{(p)},\, e_{t+1}^{(p)}\rangle}{\|e_t^{(p)}\|\,\|e_{t+1}^{(p)}\|}, \tag{4}$$

where $e_t^{(p)}$ is the per-step injected error (Section 4). The *independent-noise* null is $\mathbb{E}[\rho_t^{(p)}] = 0$; a value $\rho_t^{(p)} \to 1$ means each step's error points almost the same way as the last, so the errors add coherently even with zero mean.

The correlation is incidental: it rides the trajectory geometry. The error *is* directionally persistent, but this turns out to be incidental rather than causal. The sampler trajectory itself turns slowly: its velocity $\Delta x_t = x_{t+1} - x_t$ is strongly autocorrelated, and the error, being a perturbation of the same local update, lives in the same slowly-rotating frame. We quantify this with the FP32 *velocity autocorrelation*

$$\beta_t \;=\; \cos\big(\Delta x_t^{(\text{FP32})},\, \Delta x_{t+1}^{(\text{FP32})}\big). \tag{5}$$

Empirically $\rho_t^{(p)} \approx \beta_t$ (Section 5): the error's step persistence merely tracks the trajectory's own. The velocity control thus *explains away* the correlation as inherited geometry—and the uncorrelated-injection control below shows that removing the correlation entirely leaves the super-diffusion intact, so route (b) is not causal either.

These statistics frame the space of hypotheses. Route (a)—a non-zero mean sign bias—is refuted directly ($b_{t,\ell}^{(p)}$ is *near zero*). Route (b) is tempting because $\rho_t^{(p)}$ is *near one*, but that value is inherited from the trajectory geometry (5), and Section 4 designs two injection controls that discriminate "correlated rounding accumulates" from "continual re-injection accumulates": a *one-time* random kick (isolates whether an isolated perturbation grows) and a *continual uncorrelated* injection (isolates whether correlation is needed). Because (1) re-injects x_{t-1} into the next step, a perturbation supplied *every* step is integrated by the sampler while a one-time kick decays—and, as we will see, the coherence of that perturbation does not matter. Section 4 formalizes the accumulation as a trajectory-divergence quantity and defines the divergence step t^*.

4 METHOD

DRIFTPROBE has three components, matching the three hypotheses: a *bias meter* (H1), a *trajectory-divergence estimator* with a localization rule for t^* (H2), and a *defect-prediction protocol* (H3). All are training-free and run at inference time; the only requirement is the ability to run the same seed at two or more precisions and to log tracked tensors. We first state the hypotheses formally in terms of the statistics defined below, so that each is a decidable test rather than a slogan.

- **H1 (Mechanism of accumulation).** The failure of the rounding error to average out is a *growing-magnitude positive feedback*—the per-step error scales with the accumulating separation, so it grows as the trajectories diverge and drives further divergence—not the error's mean or correlation: the seed-expected sign bias is negligible ($\|\mathbb{E}[r]\|/\mathbb{E}\|r\| \ll 1$); the step correlation $\rho_t^{(p)}$ merely equals the velocity autocorrelation β_t ((5)); a magnitude-matched *uncorrelated* injection at the error's growing magnitude reproduces the super-diffusion, but at a *constant* per-step magnitude it is only diffusive, and a one-time kick is damped.
- **H2 (Accumulation).** The divergence $D_t^{(p)}$ of (6) grows super-diffusively: its log–log growth exponent exceeds $\frac{1}{2}$ (Proposition 1), and a finite onset step t_p^* ((7)) exists with the increment $\Delta_t^{(p)}$ persistently positive past it.
- **H3 (Prediction).** A drift summary ϕ predicts the defect label y above chance ($\mathrm{AUC}(\phi, y) > \frac{1}{2}$ with a margin), and the onset leads visible failure, $\Pr[t_p^* < t_{\text{vis}}] > \frac{1}{2}$. ***(H3 is specified here but scoped as future work; see Section 5.)***

4.1 MEASURING THE MECHANISM: FIVE CONTROLS (H1)

For a set of prompts $\mathcal{P}$ and seeds $\mathcal{S}$, we run (1) in FP32 (the reference) and in each low-precision format p, logging, at every step t and every tracked block ℓ, the rounding error $r_{t,\ell}^{(p)}$ of (2). We tracked two families of tensors: (i) the sampler latent x_t itself, and (ii) per-block outputs of the denoiser (UNet residual blocks for SDXL; transformer block outputs for FLUX). H1 asks what makes the error accumulate; we answer by eliminating the statistical explanations (Section 3.3) with a sign-bias meter and three injection controls.

Route (a): mean bias. We compute the sign-bias statistic $b_{t,\ell}^{(p)}$ (Definition 1) and ask whether the seed-averaged $|b_{t,\ell}^{(p)}|$ exceeds the zero-mean null (an average of d_ℓ symmetric signs concentrates at 0

with sd $\approx d_\ell^{-1/2}$). We also report the bias-vs-noise fraction $\|\mathbb{E}[r]\|/\mathbb{E}\|r\|$ (0 for pure noise, 1 for a purely systematic offset); a large fraction confirms route (a).

Directional correlation and its dynamical null. From the same trajectories we recover the per-step injected error $e_t^{(p)}$ (the residual between the low-precision update and the exact update transported from the shared state) and compute the step-to-step directional correlation $\rho_t^{(p)}$ of (4), together with the trajectory's own velocity autocorrelation $\beta_t = \cos(\Delta x_t^{(\text{FP32})}, \Delta x_{t+1}^{(\text{FP32})})$ (Eq. 5), the state-alignment $\cos(e_t^{(p)}, x_t^{(\text{FP32})})$, and the long-range $\cos(e_{\text{early}}^{(p)}, e_{\text{final}}^{(p)})$. If $\rho_t^{(p)} \approx \beta_t$, the error is no more persistent than the trajectory itself, so its correlation is inherited geometry rather than a rounding-specific property—but this alone does not prove correlation is irrelevant. The injection controls do.

Three injection controls: what actually drives the accumulation? We inject a perturbation ξ_t into an otherwise-FP32 trajectory and fit its divergence growth exponent. **(C1) One-time:** ξ at a single step, at the real per-step magnitude—if sub-diffusive the sampler is locally *contractive* to an isolated kick. **(C2) Continual, growing magnitude:** an independent uncorrelated Gaussian ξ_t added *afresh every step* at the BF16 error's growing magnitude profile (zero mean, zero step correlation)—if C2 reproduces the real exponent, neither mean nor correlation is needed. **(C3) Continual, constant magnitude:** the same continual uncorrelated injection at a *fixed* per-step magnitude—the difference between C2 and C3 isolates whether the *growth* of the magnitude (rather than continual injection per se) is what produces the super-diffusion. With the sign-bias meter, C1–C3 isolate the mechanism by elimination (Section 5).

4.2 Trajectory Divergence and the Divergence Step t^* (H2)

The quantity a downstream user cares about is not the per-step error but how far the low-precision trajectory has *drifted* from the reference by step t. Because seeds and schedules are shared, we can compare states directly:

$$D_t^{(p)} \;=\; \big\|x_t^{(p)} - x_t^{(\text{FP32})}\big\|_2, \qquad \Delta_t^{(p)} \;=\; D_t^{(p)} - D_{t+1}^{(p)}, \tag{6}$$

the cumulative divergence and its per-step increment. The growth rate of $D_t^{(p)}$ discriminates the accounts. A first-order propagation argument (Prop. 1, Appendix A) writes the divergence vector update as $\eta_{t-1}^{(p)} \approx J_t \eta_t^{(p)} + e_t^{(p)}$—a running sum of per-step errors $e_t^{(p)}$ transported by the Jacobian J_t of the exact update—and predicts, for *fixed-variance* independent zero-mean increments, $\mathbb{E}\|\eta^{(p)}\|_2 \sim \sqrt{n}\,\sigma$ (slope $\frac{1}{2}$). That prediction is borne out by our constant-magnitude control (C3, slope 0.43). The real error and the growing-magnitude control (C2) exceed it (slope ≈ 1) for a distinct reason: the per-step increment is *not* fixed-variance—it scales with the accumulating separation $\|\eta_t^{(p)}\|$ (measured directly in Section 5, and *not* through a latent-magnitude scaling), so a larger divergence feeds a larger increment, a positive feedback that lifts the exponent above $\frac{1}{2}$. A one-time kick, lacking this sustained growing drive, is damped (J_t contracts an isolated perturbation). H2 tests the exponent against $\frac{1}{2}$; H1's controls (Section 5) identify what supplies the slope-1 growth. We summarize the onset by a change-point on the $D_t^{(p)}$ curve:

$$t_p^* \;=\; \min\Big\{\, t : D_t^{(p)} \geq \kappa \cdot D_{\text{final}}^{(p)} \text{ and } \Delta_{t'}^{(p)} > 0 \;\forall\, t' \leq t \,\Big\}, \tag{7}$$

i.e. the earliest step by which the trajectory has accumulated a fraction κ of its total drift under sustained positive increments. (κ is a readout threshold, swept in the appendix; the qualitative claims are κ-robust.) We compute t_p^* both in latent space and in the tracked feature spaces, since a defect may first register in a specific block before surfacing in the latent.

4.3 From Drift to Defect Prediction (H3, future work)

The final component couples the drift summaries to a defect label $y \in \{0, 1\}$ for the generated image (malformed hand or face, from off-the-shelf detectors audited on a human subset). We evaluate H3 three ways: *association* (AUC of each drift summary—$D_{\text{final}}^{(p)}$, t_p^*, peak $\Delta_t^{(p)}$, peak $|b|$—and of a logistic combination, for predicting y); *lead time* (whether the numerical onset precedes the first

visibly defective decoded step, $t_p^* < t_{\text{vis}}$); and *application* (a t_p^*-triggered early-stop / re-seed decision, traded off against defect rate on a quality–compute frontier). H3 is specified here but scoped as future work; the detector/audit protocol and artifact tables are in Appendices F–G, and the full offline diagnostic pass is given as Algorithm 1 in Appendix C. At deployment the online early-stop variant evaluates the t_p^* rule during a single low-precision run against a learned threshold, with no FP32 reference at run time.

5 EXPERIMENTS

We report the same-seed measurements that establish H2 (super-diffusive divergence) and settle its mechanism (H1) by elimination—all on **SDXL**, **BF16 vs. FP32**, identical seeds, a single 48 GB L40S, computed from logged trajectories (nothing hand-entered). The FP8 arm and FLUX backbone are marked *TBD*; the defect-payoff study (H3) is run in part (§5.3) and the online application (E4) is future work.

5.1 SETUP

Model and precisions. We run SDXL (Podell et al., 2024) (UNet backbone) under deterministic DDIM sampling (Song et al., 2021a) with 50 steps at 1024×1024. Each prompt is executed twice from a bit-identical initial latent x_T and schedule: once in FP32 (the reference) and once in BF16. Because x_T and the schedule are shared, the latent states $x_t^{(p)}$ are directly comparable step-by-step, which is what makes the same-seed divergence $D_t^{(p)}$ meaningful. FP8 is supported by the hardware but not yet in this run; FLUX (Black Forest Labs, 2024) (DiT backbone) is the planned second backbone. Both are scoped as future work below.

Prompts, seeds, and scale. We use $n{=}64$ prompt$\times$seed pairs. This is a deliberately small, single-GPU pilot: it is enough to establish the divergence law and its mechanism with tight interquartile ranges, but the scale to $\sim$1,000 images needed for the defect-prediction study (E3) is future work. We report medians and interquartile ranges (IQR) over the 64 pairs rather than Gaussian error bars, since the growth-exponent and correlation statistics are not symmetric.

What is logged. Per step we log the latent x_t and the per-block outputs of the seven UNet residual stages (`down0–down2`, `mid`, `up0–up2`), and derive the sign-bias $b_{t,\ell}^{(p)}$ (Def. 1), the per-step directional correlation $\rho_t^{(p)}$ (Eq. 4), the cumulative divergence $D_t^{(p)}$ (Eq. 6), and its onset t^* (Eq. 7).

5.2 E1: WHAT DRIVES THE ACCUMULATION? (H1)

We isolate the cause by elimination, ruling out the statistical explanations with a sign-bias meter and three injection controls. The verdict: *the error is near-zero-mean, its correlation is incidental, and the super-diffusion is a growing-magnitude positive feedback—the per-step error scales with the accumulating separation, and under a constant per-step magnitude the same continual injection is merely diffusive.* Table 1 collects the numbers.

(1) Not a mean bias. The per-block sign-bias statistic $b_{t,\ell}^{(p)}$ (averaged over steps and the $n{=}60$ pairs with logged block outputs; full table in Appendix E, Table 4) is weakly negative in every block but tiny: overall mean $|b|$ is 0.012, no block exceeds $|b|/\mathrm{std} = 1.7$, and the systematic component is $\sim$2% of the noise on average. The error is, to a good approximation, *zero-mean*; a mean this small cannot integrate into the divergence E2 measures.

(2) Correlation is real but incidental. Consecutive error vectors are almost collinear ($\cos(e_t, e_{t+1}) = 0.971$; null 0), which would let a zero-mean error accumulate coherently—but this persistence is *inherited*: the FP32 trajectory's own velocity autocorrelation is 0.979, matching the error's 0.971 (-0.007), and the error is orthogonal to the motion (-0.076) and to the current latent (-0.062). Whether this correlation *causes* the super-diffusion is settled by the controls below, not its value.

Table 1: E1 — isolating the mechanism by elimination (SDXL/BF16). The error's step correlation (0.971) merely matches the trajectory velocity autocorrelation (0.979), so it is incidental. The injection controls settle causation: a one-time kick is damped (0.19); a continual uncorrelated injection at the error's *growing* magnitude reproduces the real super-diffusion ($1.01 \approx 1.04$), but at *constant* magnitude it is merely diffusive (0.43). The driver is the growth of the error magnitude with the separation—a positive feedback—not the error's mean or correlation. Diffusive null $= 0.5$; correlation null $= 0$.

Diagnostic	value	IQR / null
sign bias $\lvert b \rvert$	0.012	—
$\cos(e_t, e_{t+1})$ (error step-to-step)	0.971	$[0.969, 0.974]$
$\cos(\Delta x_t, \Delta x_{t+1})$ (trajectory velocity)	0.979	null 0
difference (error − trajectory)	−0.007	—
$\cos(e_t, \Delta x_t)$ (error vs. motion)	−0.076	null 0
$\cos(e_{\text{early}}, e_{\text{final}})$ (long-range)	0.125	null 0
growth exp., real BF16 error (n=64)	**1.04**	$[0.85, 1.17]$
growth exp., C1 one-time injection (n=16)	**0.19**	$[0.03, 0.35]$
growth exp., C2 continual, *growing* mag. (n=16)	**1.01**	$[1.006, 1.014]$
growth exp., C3 continual, *constant* mag. (n=16)	**0.43**	$[0.41, 0.44]$

(3–5) Injection controls isolate the feedback. We inject matched perturbations into an otherwise-FP32 trajectory and fit the growth exponent (n=16 each). **C1 (one-time):** a single kick at the real per-step magnitude is *damped*, exponent **0.19** (below the null of 0.5)—the sampler contracts an isolated perturbation. **C2 (continual, growing magnitude):** an *uncorrelated* Gaussian added afresh every step at the BF16 error's *growing* magnitude profile super-diffuses at **1.01**, matching the real 1.04—so neither the mean nor the correlation is needed. **C3 (continual, constant magnitude):** the same injection at a *fixed* magnitude grows only at **0.43**—*sub*-diffusive. The C2/C3 contrast is the crux: continual injection alone is not enough; the super-diffusion requires the per-step *magnitude to grow* with the accumulating divergence—not a mean bias, not a correlation effect, not a fixed-magnitude amplification, but a growing-magnitude positive feedback, which we now measure directly.

(6) The loop, measured directly. Beyond the elimination controls we measure the feedback head-on: regressing the per-step divergence increment $\Delta D_t = D_{t+1} - D_t$ on the current divergence D_t (over 3,136 step-points) is decisively positive—Spearman **0.89**, pooled slope $+$**0.043**, and *all* 64 *of* 64 *prompts (*100%*)* have a positive per-prompt slope—so a larger accumulated divergence produces a larger next increment: the growing-magnitude positive feedback is now *directly measured, not merely asserted*. **Nuance:** the amplifier is the accumulated *separation*, not latent magnitude—ΔD_t grows even as the latent norm $\|x_t\|$ *shrinks* over denoising (Appendix E).

5.3 E2: Do Same-Seed Trajectories Diverge Super-Diffusively? (H2)

This is the headline. Fig. 1 plots the mean per-step divergence $D_t = \|x_t^{\text{BF16}} - x_t^{\text{FP32}}\|_2$ on log–log axes against the diffusive null (slope $\frac{1}{2}$). The measured curve leaves the null decisively: the fitted growth exponent has median **1.04** (IQR 0.85–1.17 over n=64 pairs), against the diffusive 0.5. Same-seed FP32 and BF16 trajectories *bifurcate* super-diffusively; they do not diffuse apart. The divergence takes off at a median onset $t^* = 25.5$ of 50 and reaches terminal $D_{\text{final}} \approx 42$ (Table 2). **H2 is supported.**

The full divergence table with the planned FP8/FLUX arms is deferred to Appendix B (Table 2).

Grounded in visible output (H3, partial payoff). Decoding the n=64 FP32/BF16 *final* latents through the SDXL VAE, D_{final} tracks the *magnitude* of the visible image difference (Spearman **0.97**, AUC **0.99**; median SSIM 0.917), so the drift reaches the pixels. Two honest limits: this 0.97 is partly tautological (the near-isometric decode makes latent and pixel L_2 near-equivalent, so it *certifies* the drift surfaces rather than revealing a surprising predictor); and predicting a *specific* defect is weak (D_{final} correlates only 0.25 with a sharpness change, t^* only −0.29). A targeted detector and the lead-time test remain future work.

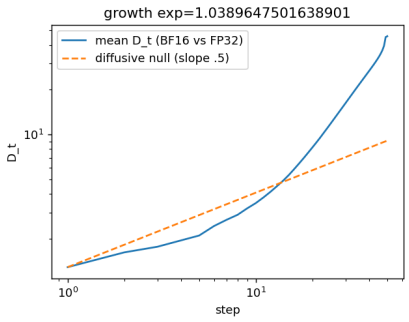

Figure 1: E2 (H2) — same-seed per-step divergence $D_t = \|x_t^{\mathrm{BF16}} - x_t^{\mathrm{FP32}}\|_2$ on SDXL, log–log, mean over n=64 pairs (blue), against the diffusive null of slope $\frac{1}{2}$ (orange dashed). Fitted exponent 1.04—super-diffusive, crossing above the $\sqrt{\text{steps}}$ null around the onset $t^* \approx 25$.

6 Discussion

What we established, and what it corrects. The headline (H2) is measured, not conjectured: same-seed FP32/BF16 trajectories on SDXL diverge *super-diffusively* (1.04 vs. a diffusive 0.5)—structured accumulation, not washout noise. The mechanism (H1) is, honestly, more mundane than a "correlated rounding" story: the five controls (Table 1) rule out mean bias, incidental step-correlation, and fixed-magnitude amplification alike, and what remains is a *growing-magnitude positive feedback* that we *measure directly*: the per-step increment grows with the accumulated divergence (Spearman 0.89; 100% of prompts positive), so separation—not latent magnitude (the increment grows even as $\|x_t\|$ shrinks)—enlarges the next increment. The exponent 1.04 is a useful divergence law, but its *explanation* is this feedback, which a constant-magnitude injection does not exhibit.

Honest failure modes. We report the bounds plainly (full list in Appendix J). (i) The per-step error may fail to accumulate on a sufficiently contractive sampler—it held here (1.04) but need not everywhere. (ii) The controls are at n=16 on SDXL/BF16; a closed-form account of the feedback and its cross-backbone replication are future work (Appendix I). (iii) H3 is only *partly* paid off: the drift predicts the *magnitude* of the visible image difference (Spearman 0.97)—near-isometric with the latent drift—while *defect-specific* prediction (sharpness 0.25) and lead time remain open. (iv) The FP32 reference is not exact, so every statistic concerns the FP32→BF16 gap, not truth. (v) All numbers are one configuration (SDXL/BF16, n=64); FP8 and FLUX are future work.

Why the signal, if real, would be cheap and general. The drift channel is *training-free* (a second precision, or a learned threshold), *architecture-general* (any tracked tensor, UNet or DiT), and *semantics-agnostic* (a numerical artifact, not cross-attention)—hence *complementary* to semantic early-detectors. The remaining arms (FP8, FLUX/DiT, stochastic samplers, and a defect-specific detector) are future work.

7 Conclusion

The rounding error of low-precision diffusion inference is usually dismissed as zero-mean noise that washes out. By direct same-seed measurement on SDXL/BF16 we showed it does not: the FP32 and BF16 latent trajectories *bifurcate super-diffusively* (growth exponent 1.04 vs. a diffusive 0.5, onset $t^* \approx 25$ of 50; H2). Five controls (Table 1) rule out biased, correlated, and fixed-magnitude accounts, and regressing the per-step increment on the accumulated divergence measures the mechanism *directly*: a *growing-magnitude positive feedback* (Spearman 0.89, 100% of prompts positive) in which separation—not latent magnitude—enlarges the next increment (H1). DriftProbe packages this as a training-free diagnostic that *reads* the drift on any UNet or DiT block; defect-specific prediction (H3), lead time, and the FP8/FLUX arms remain future work.

Reproducibility Statement

DriftProbe is a purely diagnostic, training-free procedure, which makes reproduction cheap and deterministic given a fixed seed. The core statistics—the per-step sign bias, the pairwise rounding-error correlation, and the second-moment divergence rate that underlies Proposition 1—are defined in Section 4 (Eq. (6), Eq. (7)) and formalized in Appendix E. The prediction protocol and detector/audit harness are specified in Section 4.3 and Appendix F, and the sampler and precision configuration (SDXL backbone, BF16 arm, scheduler, step count, seeds) is pinned down in Appendix H. All reported E1/E2 numbers (Tables 4, 1, 2) come from single-GPU, same-seed, pure-inference runs on a publicly available generator; we release the trajectory-logging hooks, the statistic implementations, and the evaluation scripts so every measured table can be regenerated from the released artifacts without any training. The pre-registered failure analysis and honest caveats (including which arms remain future work) are recorded in Appendices I and J.

Ethics Statement

This work is a diagnostic for numerical drift in low-precision diffusion sampling; it does not train or release a generative model and creates no new generative capability. Its intended use is to make low-precision inference more trustworthy by flagging when accumulated rounding error pushes a trajectory off its expected path. The study uses only publicly available models and model-generated images under their respective licenses, and involves no human subjects or personal data beyond pre-registered ratings of generated content.

Use of Large Language Models

Large language models were used as writing and engineering assistance: drafting and copy-editing prose, refactoring LaTeX and analysis scaffolding, and literature triage — surfacing candidate related work, which the authors then verified against primary sources. Entries that could not be confirmed against a paper page, project page or proceedings entry were removed rather than retained on the strength of a model's summary. The research contribution — the same-seed super-diffusion diagnostic, the five-control elimination of the competing statistical mechanisms, the direct measurement of the growing-magnitude feedback, and the experimental design — is the authors' own, and the authors take full responsibility for all content, including any errors. No results, data, or citations were generated by a language model.

REFERENCES

Federico Betti, Lorenzo Baraldi, Rita Cucchiara, and Nicu Sebe. Optimizing resource consumption in diffusion models through hallucination early detection. *arXiv preprint arXiv:2409.10597*, 2024.

Federico Betti, Lorenzo Baraldi, Rita Cucchiara, and Nicu Sebe. Hallucination early detection in diffusion models. *arXiv preprint arXiv:2604.20354*, 2026.

Black Forest Labs. FLUX.1: Text-to-image generation with rectified flow transformers. `https://github.com/black-forest-labs/flux`, 2024.

Tim Dettmers, Mike Lewis, Younes Belkada, and Luke Zettlemoyer. LLM.int8(): 8-bit matrix multiplication for transformers at scale. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *International Conference on Machine Learning (ICML)*, 2024.

Steven K Esser, Jeffrey L McKinstry, Deepika Bablani, Rathinakumar Appuswamy, and Dharmendra S Modha. Learned step size quantization. In *International Conference on Learning Representations (ICLR)*, 2020.

Elias Frantar, Saleh Ashkboos, Torsten Hoefler, and Dan Alistarh. GPTQ: Accurate post-training quantization for generative pre-trained transformers. In *International Conference on Learning Representations (ICLR)*, 2023.

Amir Gholami, Sehoon Kim, Zhen Dong, Zhewei Yao, Michael W Mahoney, and Kurt Keutzer. A survey of quantization methods for efficient neural network inference. *arXiv preprint arXiv:2103.13630*, 2022.

David Goldberg. What every computer scientist should know about floating-point arithmetic. *ACM Computing Surveys*, 23(1):5–48, 1991.

Song Han, Huizi Mao, and William J Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. In *International Conference on Learning Representations (ICLR)*, 2016.

Yefei He, Luping Liu, Jing Liu, Weijia Wu, Hong Zhou, and Bohan Zhuang. PTQD: Accurate post-training quantization for diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics (SIAM), 2nd edition, 2002.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

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

Yushi Huang, Ruihao Gong, Jing Liu, Tianlong Chen, and Xianglong Liu. TFMQ-DM: Temporal feature maintenance quantization for diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Dhiraj Kalamkar, Dheevatsa Mudigere, Naveen Mellempudi, Dipankar Das, Kunal Banerjee, Sasikanth Avancha, Dharma Teja Vooturi, Nataraj Jammalamadaka, Jianyu Huang, Hector Yuen, et al. A study of BFLOAT16 for deep learning training. *arXiv preprint arXiv:1905.12322*, 2019.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Xiuyu Li, Yijiang Liu, Long Lian, Huanrui Yang, Zhen Dong, Daniel Kang, Shanghang Zhang, and Kurt Keutzer. Q-Diffusion: Quantizing diffusion models. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Yuhang Li, Ruihao Gong, Xu Tan, Yang Yang, Peng Hu, Qi Zhang, Fengwei Yu, Wei Wang, and Shi Gu. BRECQ: Pushing the limit of post-training quantization by block reconstruction. In *International Conference on Learning Representations (ICLR)*, 2021.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Paulius Micikevicius, Dusan Stosic, Neil Burgess, Marius Cornea, Pradeep Dubey, Richard Grisenthwaite, Sangwon Ha, Alexander Heinecke, Patrick Judd, John Kamalu, et al. FP8 formats for deep learning. *arXiv preprint arXiv:2209.05433*, 2022.

Markus Nagel, Rana Ali Amjad, Mart van Baalen, Christos Louizos, and Tijmen Blankevoort. Up or down? adaptive rounding for post-training quantization. In *International Conference on Machine Learning (ICML)*, 2020.

Markus Nagel, Marios Fournarakis, Rana Ali Amjad, Yelysei Bondarenko, Mart van Baalen, and Tijmen Blankevoort. A white paper on neural network quantization. *arXiv preprint arXiv:2106.08295*, 2021.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning (ICML)*, 2021.

Alexander Quinn Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob Mcgrew, Ilya Sutskever, and Mark Chen. GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. In *International Conference on Machine Learning (ICML)*, 2022.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Houwen Peng, Kan Wu, Yixuan Wei, Guoshuai Zhao, Yuxiang Yang, Ze Liu, Yifan Xiong, Ziyue Yang, Bolin Ni, Jingcheng Hu, et al. FP8-LM: Training FP8 large language models. *arXiv preprint arXiv:2310.18313*, 2023.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations (ICLR)*, 2024.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with CLIP latents. *arXiv preprint arXiv:2204.06125*, 2022.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Sooyoung Ryu, Mathieu Salzmann, and Saqib Javed. Q-Drift: Quantization-aware drift correction for diffusion model sampling. *arXiv preprint arXiv:2603.18095*, 2026.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al. Photorealistic text-to-image diffusion models with deep language understanding. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Yuzhang Shang, Zhihang Yuan, Bin Xie, Bingzhe Wu, and Yan Yan. Post-training quantization on diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Junhyuk So, Jungwon Lee, Daehyun Ahn, Hyungjun Kim, and Eunhyeok Park. Temporal dynamic quantization for diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning (ICML)*, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations (ICLR)*, 2021a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

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)*, 2021b.

Xiao Sun, Jungwook Choi, Chia-Yu Chen, Naigang Wang, Swagath Venkataramani, Vijayalakshmi Srinivasan, Xiaodong Cui, Wei Zhang, and Kailash Gopalakrishnan. Hybrid 8-bit floating point (HFP8) training and inference for deep neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Naigang Wang, Jungwook Choi, Daniel Brand, Chia-Yu Chen, and Kailash Gopalakrishnan. Training deep neural networks with 8-bit floating point numbers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Guangxuan Xiao, Ji Lin, Mickael Seznec, Hao Wu, Julien Demouth, and Song Han. SmoothQuant: Accurate and efficient post-training quantization for large language models. In *International Conference on Machine Learning (ICML)*, 2023.

APPENDIX

This appendix collects the formal definitions of the sign-bias, correlation, and divergence statistics (E), the defect-detector and human-audit protocol for the future-work prediction study (F), the sampler and precision configuration (H), a failure analysis (I), and reproducibility notes with honest caveats (J). The E1/E2 entries (Tables 4, 1, 2) are measured on SDXL/BF16; entries for the FP8 arm, the FLUX backbone, and the E3/E4 defect-prediction study remain *TBD* as future work.

A GROWTH-RATE PROPOSITION (FIRST-ORDER SCAFFOLD) AND ITS LIMITS

The growth law used in Section 4 is a first-order propagation argument. It motivates the growth exponent as the diagnostic, but our controls show it is *incomplete*: it treats J_t as inert, whereas the sampler amplifies each injected increment. Writing the one-step map as $x_{t-1}^{(p)} = \Phi_t(x_t^{(p)}, \epsilon_\theta(x_t^{(p)}, t, c)) + e_t^{(p)}$ with $e_t^{(p)}$ the per-step injected rounding error and J_t the Jacobian of the exact update at $x_t^{(\mathrm{FP32})}$, the divergence vector $\eta_t^{(p)} = x_t^{(p)} - x_t^{(\mathrm{FP32})}$ propagates as $\eta_{t-1}^{(p)} \approx J_t\, \eta_t^{(p)} + e_t^{(p)}$: a running sum of past errors transported by the sampler.

Proposition 1 (Two routes to super-diffusive divergence)**.** *Suppose the transported per-step errors $\tilde{e}_t^{(p)}$ (the $e_t^{(p)}$ mapped to the terminal frame) have common mean $\mu_p = \mathbb{E}[\tilde{e}_t^{(p)}]$, per-coordinate variance σ_p^2, and lag-one directional correlation $\rho_p = \cos(\tilde{e}_t^{(p)}, \tilde{e}_{t+1}^{(p)})$. Over n elapsed steps, $\mathbb{E}\|\eta^{(p)}\|_2$ scales as*

$$\mathbb{E}\|\eta^{(p)}\|_2 \;\sim\; \begin{cases} c\,\sqrt{n}\,\sigma_p, & \mu_p{=}0,\ \rho_p{\approx}0 \textit{ (diffusive)}, \\ n\,\|\mu_p\| + O(\sqrt{n}\,\sigma_p), & \mu_p \neq 0 \textit{ (biased drift)}, \\ c'\,n\,\sigma_p\sqrt{\rho_p}, & \mu_p{=}0,\ \rho_p{\to}1 \textit{ (correlated)}. \end{cases}$$

Hence the log–log slope of $D_t^{(p)}$ is $\frac{1}{2}$ only under the independent, zero-mean *null; it tends to* 1 *if* either *a persistent mean $\mu_p \neq 0$ dominates* or *the increments are strongly positively correlated ($\rho_p \to 1$), even when $\mu_p = 0$. The intuition: an independent zero-mean error is a random walk ($\sqrt{n}$); a non-zero mean adds a drift $n\|\mu_p\|$; a correlated zero-mean error also grows like n because the positive cross-step terms of $\sum_{s,t} \mathbb{E}\langle \tilde{e}_s, \tilde{e}_t \rangle$ sum to $O(n^2)$. The crossover where the n-term overtakes the $\sqrt{n}$-term is the onset t^*.*

Why this scaffold is not the mechanism. The proposition predicts that independent zero-mean increments ($\mu_p{=}0, \rho_p{\approx}0$) give the diffusive slope $\frac{1}{2}$. Our continual uncorrelated-injection control (C2) satisfies exactly those hypotheses yet grows at slope ≈ 1.01, not $\frac{1}{2}$. The discrepancy is the transported-error assumption: the proposition takes $\tilde{e}_t$ as a fixed running sum, but J_t is neither the identity nor uniformly contractive—an increment injected at step s is repeatedly re-processed and amplified along the remaining trajectory, so continual injection of *any* increment (correlated or not) accumulates super-diffusively. A one-time kick, by contrast, is damped, indicating J_t contracts an *isolated* perturbation while sustained per-step driving is amplified. A quantitative account of this amplification (the time-varying spectrum of J_t and the growing latent scale over the denoising schedule) is the natural next step; we report the controls as the empirical isolation of the mechanism and leave the exact dynamical law open.

B FULL DIVERGENCE TABLE (E2, WITH PLANNED ARMS)

Table 2 is the full version of the E2 divergence result summarized in Section 5, including the FP8 and FLUX arms scoped as future work.

C DRIFTPROBE DIAGNOSTIC PASS (ALGORITHM)

Algorithm 1 summarizes the offline diagnostic pass referenced in Section 4.3.

Table 2: E2 (H2) — same-seed trajectory divergence (SDXL/BF16, n=64; medians, IQR in brackets). Growth exponent $\gg \frac{1}{2}$ (diffusive null) confirms super-diffusive accumulation. FP8/FLUX: *TBD* (future work).

Backbone	Prec.	D_{final}	growth exp.	median t^*
SDXL	BF16	42.0	**1.04** [.85,1.17]	25.5
SDXL	FP8	*TBD*	*TBD*	*TBD*
FLUX	BF16	*TBD*	*TBD*	*TBD*
FLUX	FP8	*TBD*	*TBD*	*TBD*
diffusive null		—	0.50	—

Algorithm 1 DRIFTPROBE diagnostic pass (per prompt, per seed)

Require: prompt c, seed s, precisions $\{\text{FP32}, p_1, \dots\}$, tracked layers $\{\ell\}$

for each precision q **do**
  run (1) from seed s; log $x_t^{(q)}$ and $a_{t,\ell}^{(q)}$ for all t, ℓ
end for
for each low-precision p, each step t, layer ℓ **do**
  $r_{t,\ell}^{(p)} \leftarrow a_{t,\ell}^{(p)} - a_{t,\ell}^{(\text{FP32})}$; $b_{t,\ell}^{(p)} \leftarrow$ mean sign ▷ H1(a)
  $\rho_t^{(p)} \leftarrow \cos(e_t^{(p)}, e_{t+1}^{(p)})$ ▷ H1(b)
  $D_t^{(p)} \leftarrow \|x_t^{(p)} - x_t^{(\text{FP32})}\|_2$ ▷ H2
end for
$t_p^* \leftarrow$ change-point of $D_t^{(p)}$ via (7)
decode final image; obtain defect label y ▷ H3
return $\{b_{t,\ell}^{(p)}\}, \{D_t^{(p)}\}, t_p^*, y$

D EXTENDED EXPERIMENTAL DESIGN

The executable plan is staged as E0–E4, each mapped to the exact table/figure it fills. Table 3 records that mapping. E0–E2 are done (SDXL/BF16); E3–E4 are future work.

Table 3: Stage-to-artifact map. E0 is infrastructure (no artifact). Status is for the present SDXL/BF16 pilot.

Stage	Hypothesis	Fills	Status
E0	— (pipeline)	determinism / logging	done
E1	H1 (mechanism)	T. 4, 1	done
E2	H2 (divergence)	Fig. 1, T. 2	done
E3	H3 (prediction)	Table 6	future
E4	application	Table 7	future

E0 (determinism and logging). The load-bearing precondition is that the initial latent x_T and the noise schedule are *bit-identical* across precisions, otherwise $D_t^{(p)}$ is not comparable. We seed the RNG in FP32 and cast the sampled x_T down to each format, fix the sampler and step count per backbone, and disable any nondeterministic kernels. Logging stores *summaries* (per-tensor sign counts, norms, and the latent itself), never full activation tensors, to keep disk within budget.

E1–E4 follow the metric definitions below. In the present pilot only BF16-vs-FP32 on SDXL is run (n=64 pairs, 50 DDIM steps, 1024^2). When the FP8 arm is added it will use a larger sample budget than BF16 to offset its higher defect rate, so that per-cell significance and per-class AUC are estimated at comparable precision.

E SIGN-BIAS AND DIVERGENCE STATISTICS

Recall the per-step per-tensor rounding error $r_{t,\ell}^{(p)} = a_{t,\ell}^{(p)} - a_{t,\ell}^{(\text{FP32})}$ (Eq. 2). We track the following:

- **Sign bias** $b_{t,\ell}^{(p)} = \frac{1}{d_\ell} \sum_i \text{sign}(r_{t,\ell,i}^{(p)})$ (Def. 1). Null: under i.i.d. symmetric signs, $b \sim$ mean of d_ℓ Rademacher variables, so $\text{sd}(b) \approx d_\ell^{-1/2}$; we flag a cell significant at $|b| > z_{0.975}\, d_\ell^{-1/2}$ and additionally require consistency of sign across seeds.
- **Bias-vs-noise fraction** $\|\mathbb{E}[r_{t,\ell}^{(p)}]\| / \mathbb{E}\|r_{t,\ell}^{(p)}\| \in [0,1]$, estimated by averaging over seeds; 0 = pure noise, 1 = pure systematic offset. Measured overall value 0.022 (Table 4): the error is effectively zero-mean.
- **Step-to-step directional correlation** $\rho_t^{(p)} = \cos(e_t^{(p)}, e_{t+1}^{(p)})$ (Eq. 4), read against the trajectory's own velocity autocorrelation $\beta_t = \cos(\Delta x_t^{(\text{FP32})}, \Delta x_{t+1}^{(\text{FP32})})$ (Eq. 5), with the state-alignment control $\cos(e_t^{(p)}, x_t^{(\text{FP32})})$, the motion-alignment control $\cos(e_t^{(p)}, \Delta x_t^{(\text{FP32})})$, and the long-range $\cos(e_{\text{early}}, e_{\text{final}})$. Independent null 0. Measured $\rho = 0.971$ (IQR $[0.969, 0.974]$), $\beta = 0.979$ so $\rho - \beta = -0.007$, motion-alignment -0.076, state-alignment -0.062, long-range 0.125 (Table 1). The near-equality $\rho \approx \beta$ shows the correlation is *incidental*—the error rides the trajectory's smooth geometry—and the near-zero motion- and state-alignments show it is not a trivial state-proportional artifact. Whether this coherence *causes* the growth is settled by the injection controls (Appendix I), which show it does not: a magnitude-matched uncorrelated continual injection super-diffuses equally.
- **Trajectory velocity autocorrelation** β_t of Eq. (5), computed on the FP32 trajectory increments. Measured 0.979; it explains away the error's step correlation ($\rho = 0.971 \approx \beta$) as inherited geometry rather than a rounding-specific property.
- **Cumulative divergence** $D_t^{(p)} = \|x_t^{(p)} - x_t^{(\text{FP32})}\|_2$ and increment $\Delta_t^{(p)} = D_t^{(p)} - D_{t+1}^{(p)}$ (Eq. 6). Growth exponent estimated by regressing $\log D_t^{(p)}$ on $\log(\text{steps elapsed})$; ≈ 0.5 = diffusive, > 0.5 = super-diffusive. Measured median 1.04 (IQR $[0.85, 1.17]$; Table 2); the driver is isolated by the injection controls (Appendix I), not by the exponent alone.
- **Divergence step** t_p^* via Eq. (7); threshold κ swept over $\{0.1, 0.25, 0.5\}$ with a robustness table (*TBD*; Table 5).

Table 4: E1 — per-block rounding-error *sign bias* (SDXL/BF16, n=60). All blocks are weakly negative but near zero: overall mean $|b|$=0.012, overall bias/noise fraction 0.022. The mean bias is *not* the mechanism (cf. Table 1). FP8/FLUX: *TBD*.

Block	mean b	$\lvert b\rvert/\text{std}$	bias/noise frac.
`down0`	-0.011	1.07	0.020
`down1`	-0.007	1.38	0.012
`down2`	-0.001	0.25	0.006
`mid`	-0.001	0.31	0.006
`up0`	-0.012	1.71	0.022
`up1`	-0.011	0.65	0.024
`up2`	-0.026	1.69	0.066
overall	$-0.012^\dagger$	—	0.022

†mean of $|b|$ across blocks; the signed mean is negative but $|b| \approx 0.012$.

Concentration of the null. Under the zero-mean model the coordinate signs $\text{sign}(r_{t,\ell,i}^{(p)})$ are i.i.d. Rademacher, so $b_{t,\ell}^{(p)} = \frac{1}{d_\ell}\sum_i s_i$ has $\mathbb{E}[b] = 0$ and $\text{Var}[b] = 1/d_\ell$; by a Hoeffding bound, $\Pr[|b| \geq \tau] \leq 2\exp(-d_\ell \tau^2/2)$, so a cell with $|b_{t,\ell}^{(p)}| > z_{0.975}\, d_\ell^{-1/2}$ is significant at the 5% level and the wide layers (large d_ℓ) have the tightest null bands—meaning a *small* $|b|$ there can already be significant, and conversely raw magnitude must be compared against $d_\ell^{-1/2}$, never in absolute terms.

We additionally require sign consistency of $b_{t,\ell}^{(p)}$ across seeds to rule out a single seed's fluctuation masquerading as bias. Empirically the measured $|b|$ (≈ 0.012) sits close to this null band—the blocks are weakly, consistently negative but not materially mean-biased—which rules out the mean-bias route. The velocity-autocorrelation control ($\rho \approx \beta$, -0.007) shows the error's step correlation is inherited trajectory geometry, and the injection controls (Appendix I) show the correlation is not causal at all: a magnitude-matched uncorrelated continual injection super-diffuses just as much, so the driver is per-step re-injection, not any statistical structure of the rounding.

Growth-exponent estimate. The exponent in Proposition 1 is estimated by ordinary least squares of $\log D_t^{(p)}$ on $\log(\text{elapsed steps})$ over the post-onset segment, with the IQR taken over the n=64 pairs. H2's super-diffusive claim is the one-sided test that this slope exceeds $\frac{1}{2}$; the measured median 1.04 with IQR $[0.85, 1.17]$ clears it (the lower quartile is well above 0.5).

Direct measurement of the feedback loop. Beyond the injection controls, we measure the closed loop of Section 5 head-on. For each prompt p and step t we form the per-step divergence increment $\Delta D_t^{(p)} = D_{t+1}^{(p)} - D_t^{(p)}$ and regress it on the current divergence $D_t^{(p)}$. Pooling the 3,136 step-points over the 64 prompts, the relation is strongly positive: rank correlation $\mathrm{Spearman}(\Delta D_t, D_t) = 0.89$ and pooled OLS slope $+0.043$. The effect is not carried by a few prompts—fitting the slope *within* each prompt, **all** 64 **of** 64 **prompts** (100%) **have a positive slope** (median per-prompt slope ≈ 0.059). This is the growing-magnitude positive feedback measured directly, rather than inferred from the injection contrast alone: a larger accumulated divergence produces a larger next increment.

The amplifier is the separation, not the latent magnitude. A natural alternative is that the increment grows simply because the rounding error scales with the latent's magnitude $\|x_t\|$. This is *ruled out* by the sign of the denoising trend: over DDIM sampling $\|x_t\|$ *shrinks*, yet both the increment and the divergence *grow*. Concretely, $\mathrm{Spearman}(\Delta D_t, \|x_t\|) = -0.68$ and $\mathrm{Spearman}(D_t, \|x_t\|) = -0.87$: the increment and the accumulated divergence both rise as the latent norm falls. The driver is therefore the sampler's sensitivity to the current *separation* between the two trajectories, not a scaling of the error with the latent magnitude. We state this explicitly so the feedback is not overclaimed as a latent-magnitude effect.

Table 5: κ-robustness of t^* (median over seeds). All *TBD*; the qualitative ordering of t^* across precisions is expected to be κ-stable.

Prec.	κ=0.1	κ=0.25	κ=0.5
BF16	*TBD*	*TBD*	*TBD*
FP8	*TBD*	*TBD*	*TBD*

F Defect Detectors and Human Audit

Hand defects are labeled by an off-the-shelf hand keypoint / finger-count detector; face defects by a landmark-based symmetry/consistency detector. Thresholds, versions, and per-detector precision/recall on a human-audited subset of N images are reported here (*TBD*).

Labeling protocol. (1) Every generated final image is scored by both detectors, giving an automatic label $y_{\text{auto}} \in \{0, 1\}$ (defect if either detector fires). (2) A random subset (stratified by backbone $\times$ precision $\times$ y_{auto}, with oversampling of the positive class) is sent to human raters. (3) Each image is rated by ≥ 2 raters against a fixed rubric—for hands: incorrect finger count, fused/extra digits, impossible articulation; for faces: gross asymmetry, misplaced or duplicated features, melted structure. (4) Inter-rater agreement (Cohen's κ; *TBD*) and detector precision/recall against the human consensus (*TBD*) quantify label noise. (5) This detector-vs-human error is propagated into the AUC confidence intervals of Table 6 by a label-noise-aware bootstrap, so a reported AUC is never cleaner than the labels that produced it.

Why a human subset is mandatory. The automatic detectors are themselves imperfect and can share failure modes with the generator (e.g., a keypoint detector may miscount fingers on exactly the

blurry hands the model also fails on). Reporting AUC only against y_{auto} would let the metric look better than the underlying truth; the human subset is the anchor that bounds that gap. This is a hard caveat, not a footnote: without it, H3's numbers would be uninterpretable.

Intermediate-step decoding for t_{vis}. The lead-time test needs the earliest step at which a *decoded* intermediate is judged defective. Decoding every step through the VAE is costly, so we decode at a fixed subset of steps (geometrically spaced) and take t_{vis} as the earliest decoded step flagged by the same detectors; the resulting t_{vis} is therefore an upper bound on the true first-visible step, which makes the lead-time claim $t^* < t_{\text{vis}}$ *conservative* rather than inflated.

G Planned E3/E4 Artifact Tables (Future Work)

The defect-prediction (E3) and application (E4) stages are specified but not yet run; their artifact tables are reproduced here with honest *TBD* entries so the executable plan is complete.

Table 6: E3 (H3) — defect prediction from drift summaries (AUC). *Future work; not yet run.*

Drift summary	SDXL/BF16	SDXL/FP8	FLUX/BF16	FLUX/FP8
D_{final}	*TBD*	*TBD*	*TBD*	*TBD*
t^*	*TBD*	*TBD*	*TBD*	*TBD*
peak Δ_t	*TBD*	*TBD*	*TBD*	*TBD*
logistic combination	*TBD*	*TBD*	*TBD*	*TBD*

Table 7: E4 — compute saved vs. defect-rate change from a drift-triggered early stop / re-sample. *Future work; not yet run.*

Policy	compute Δ	defect rate	net saving
baseline (full)	*TBD*	*TBD*	*TBD*
drift early-stop	*TBD*	*TBD*	*TBD*
drift re-sample	*TBD*	*TBD*	*TBD*

H Sampler and Precision Configuration

Per backbone we fix: sampler (DDIM-style deterministic), number of steps, guidance scale, resolution, and RNG seeding so that x_T is bit-identical across precisions. Low-precision execution uses [BF16: native autocast; FP8: format and scaling scheme TBD]. Tracked layers: for SDXL, the down/mid/up residual-block outputs; for FLUX, per-transformer-block outputs. Exact layer indices and the logging hook are listed here (*TBD*).

I Failure Analysis: How Each Hypothesis Can Break

We enumerate the ways the story can fail and how each is diagnosed, so that a negative result is still an informative one.

The two injection controls (isolating the mechanism). The measured statistics—near-zero mean ($|b| \approx 0.012$) and high step correlation ($\rho \approx 0.971$, itself equal to the trajectory velocity autocorrelation $\beta \approx 0.979$)—leave two candidate drivers: the error's directional coherence, or the mere fact that it is re-injected every step. We discriminate with matched-magnitude injections into an otherwise-FP32 trajectory, each perturbation scaled to the measured BF16 per-step error profile. *C1 (one-time).* A single random kick grows at exponent 0.19 (IQR $[0.03, 0.35]$; $n{=}16$)—*sub*-diffusive, below the null 0.5: the sampler *contracts* an isolated perturbation, so raw amplification of any one perturbation is ruled out. *C2 (continual, uncorrelated).* An independent Gaussian perturbation of the matched

magnitude, added *afresh at every step* (zero mean and zero step-to-step correlation by construction), grows at exponent 1.01 (IQR $[1.006, 1.014]$; n=16)—statistically indistinguishable from the real error's 1.04. Since C2 has neither mean nor correlation yet reproduces the super-diffusion, **the directional coherence of the rounding error is not causal**: the driver is the continual per-step re-injection itself, amplified by a sampler that nonetheless contracts a one-time kick. Two caveats bound this: both controls are at n=16 on SDXL/BF16, and a full account of why sustained injection exceeds the first-order $\sqrt{t}$ prediction (the time-varying Jacobian spectrum; Appendix A) is left open. A ρ-matched continual injection (to confirm correlation adds nothing even at matched magnitude *and* matched ρ) and cross-backbone replication are future work.

H2 holds but H3 fails (divergence without prediction). Trajectories may diverge super-diffusively yet the divergence may not sort defective from clean images—if, say, the drift saturates into a benign global shift rather than a structured corruption. Symptom: AUC near $\frac{1}{2}$ despite a clean t^*. We report it as such.

H3 association holds but the lead fails. Even a high AUC does not guarantee $t^* < t_{\text{vis}}$: the numerical divergence could coincide with or lag the visible defect, which would kill the early-stop value while leaving a valid post-hoc correlate. This is why the lead-time distribution will be reported in full when E3 is run, negative mass included, and why the application table (Table 7) charges re-sampling cost honestly.

Confounding by prompt difficulty. Both drift and defects might merely track how hard a prompt is. The same-seed design and per-prompt conditioning defend against this, but a residual confound is possible; we treat any prediction result as associational unless the early-stop intervention (E4) actually moves the defect rate, which is the causal-flavored test.

J Scope, Forward Notes, and Reproducibility

Out of scope (left for future work). We study two backbones (SDXL, FLUX) under deterministic sampling. Stochastic samplers re-inject fresh noise each step and would blur the same-seed comparison; INT8 and mixed-precision schemes have different rounding-direction structure; video and 3D diffusion add a temporal axis over which drift may compound differently. Each is a natural extension of the same measurement channel.

Reproducibility. The diagnostic analysis (bias, divergence, AUC) regenerates from logged trajectories without a GPU; only the trajectory generation needs the single 4090/5090. Released artifacts (*TBD*): logging hooks, trajectory summaries, detector configs, human-audit rubric, and analysis scripts.

Consolidated honest caveats. (i) FP32 is the *operational* reference, not exact real arithmetic—every "rounding error" is defined relative to it, so a statistic measured against FP32 is a statement about the FP32→low-precision gap, not about truth. (ii) The measured results cover *one* configuration: SDXL, BF16-vs-FP32, deterministic DDIM, n=64 same-seed pairs. FP8, FLUX, and larger scale are future work; the corresponding table cells are honest *TBD*, not implied. (iii) Both statistical accounts of H1 are *not* supported—the sign bias is ≈ 0.012 (not mean-biased) and the step correlation, though ≈ 0.97, is inherited trajectory geometry and shown non-causal by the continual uncorrelated-injection control ($1.01 \approx 1.04$). The driver we report is the continual per-step re-injection of error into an amplifying sampler; a closed-form account of that amplification (the time-varying Jacobian spectrum) is left open (Appendix A). (iv) The defect-prediction claims (H3) are *not yet tested*; automatic defect labels are imperfect and will need the human-audited subset when E3 is run. (v) n=64 is a single-GPU pilot—enough for tight IQRs on the divergence law and its mechanism, but not for the per-class AUC estimates E3 needs. None of these caveats is optional; each is surfaced in the main text as well.