A DOMINANT SELF-CONDITIONING DIRECTION DRIVES REPETITION IN UNCONDITIONAL CONTINUOUS DIFFUSION LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Continuous diffusion language models offer an alternative to autoregressive generation, but their generations may suffer from repetition. We find that unconditional generations from ELF, a recent family of continuous diffusion language models, are more repetitive than human text, while Gen-PPL, a common likelihood-based metric, gives lower perplexity to repetitive generations and can conceal this problem while biasing quality evaluation. Our analysis links this behavior to a self-conditioning feedback loop in which clean-embedding predictions are repeatedly carried into subsequent denoising steps, driving representations toward an effectively one-dimensional contractive attractor associated with repetition. Based on this mechanism, we introduce Attractor-Contrast-Escape (ACE), a training-free inference-time intervention that estimates a repetition direction by contrasting denoising paths trapped in repetition with paths relatively free of repetition and subtracts it from the self-conditioning feedback during denoising. Using a direction estimated only once on ELF-B, ACE reduces mean 4-gram self-repetition rate from 7.28% to 4.48%, while retaining competitive results on several text-quality metrics beyond Gen-PPL. The direction remains effective across ELF sizes and inference configurations, and ACE also generalizes to other unconditional self-conditioned continuous diffusion language models. These results identify self-conditioning feedback as a source of repetition in continuous diffusion language models and show that ACE can directly mitigate this repetition during inference.

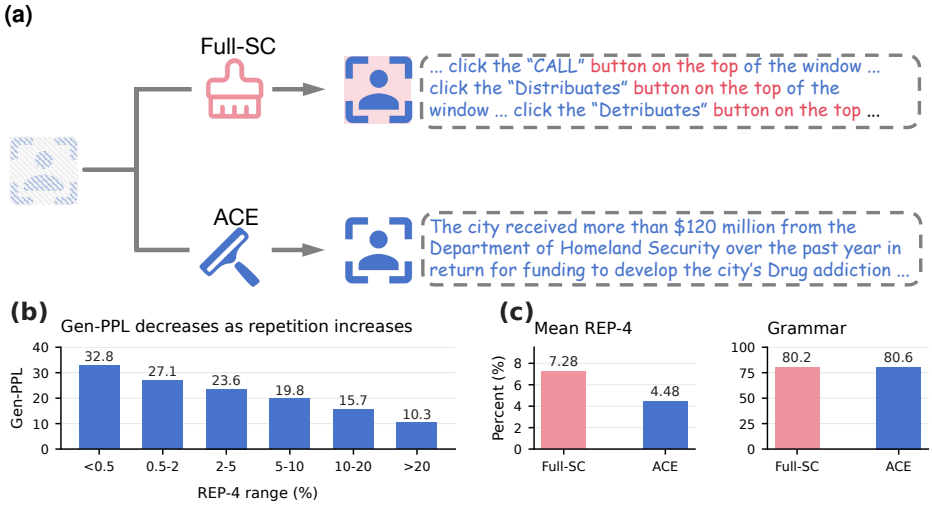


Figure 1: **Full-SC produces repetitive text and Gen-PPL favors it, whereas ACE reduces repetition with competitive text quality.** (a) An illustration of text generation from the same noisy state: Full-SC yields repetitive text, whereas ACE suppresses repetition during denoising. (b) Gen-PPL decreases across repetition bins. (c) ACE lowers mean REP-4 while preserving grammar.

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.

1 INTRODUCTION

Continuous diffusion language models (DLMs) generate text by denoising in a continuous embedding space. Relative to discrete DLMs, this continuous state exposes a differentiable denoising trajectory that provides a direct interface for gradient-based control and classifier-free guidance (Li et al., 2022; Hu et al., 2026). Self-conditioning improves generation quality by enabling the denoiser to refine its clean-embedding prediction across denoising steps (Chen et al., 2022). Related studies interpret self-conditioning as fixed-point iteration for few-step distillation (Yoo et al., 2026) and address its training–inference mismatch (Li & Xu, 2026).

However, recent continuous DLMs such as ELF (Hu et al., 2026) can produce highly repetitive text during unconditional generation under full self-conditioning (Full-SC), repeating words and phrases (Figure 1(a)). Gen-PPL, a commonly used likelihood-based quality proxy, gives lower perplexity to more repetitive generations (Figure 1(b)) and can conceal this problem while biasing quality evaluation (§3.1). We find that the self-conditioning feedback in the Full-SC regime drives repetition by biasing the denoising trajectory toward repetitive text along its dominant direction. With all other inference settings fixed, we observe that repetition rises as the self-conditioning strength increases, while a separately trained model without self-conditioning produces almost no repetition (Figure 2). These observations point to self-conditioning as a causal driver of repetition (§3.2). We further observe that repetition rises during denoising and then levels off (Figure 4), suggesting an attractor-like regime that favors repetitive text. This motivates a local fixed-point approximation of the self-conditioning response underlying this attractor-like repetitive regime. After we decompose each denoising update into a self-conditioning response toward the fixed point and an ordinary denoising drive, theoretically, we predict a dominant direction by locally linearizing around this fixed point. Empirically, we show that the scalar projection of a sample's self-conditioning feedback onto this direction tracks its repetition level, while the self-conditioning feedback dynamics develop a dominant mode aligned with it as the denoising trajectory settles into a repetitive regime. These findings show that the self-conditioning feedback forms a repetition attractor that biases denoising trajectories near the attractor toward its dominant direction, thereby producing repetitive text (§4.1).

Unlike autoregressive degeneration, where repeated tokens feed back through the generated prefix and accumulate along sequence order (Holtzman et al., 2020; Xu et al., 2022), self-conditioning in continuous DLMs feeds clean-embedding predictions back across denoising steps before token selection (Chen et al., 2022; Hu et al., 2026). Existing remedies for autoregressive repetition typically modify token-level decoding or training objectives, for example by changing sampling behavior or penalizing repetitive continuations (Welleck et al., 2020). In continuous DLMs, the repetition signal is formed upstream in the self-conditioning feedback of the clean-embedding prediction, so the intervention should target the denoising dynamics rather than token selection. Motivated by this effectively one-dimensional attractor, we propose Attractor-Contrast-Escape (ACE). ACE aims to escape the attractor by removing the self-conditioning feedback component aligned with its dominant repetition direction. ACE first estimates a repetition direction by contrasting the self-conditioning feedback from Full-SC denoising paths trapped in repetition with that from paths relatively free of repetition, then subtracts it from the feedback at each denoising step using a chosen steering strength (Figure 1(a)). This contrastive estimator is related to activation steering (Rimsky et al., 2024; Shnaidman et al., 2025), but ACE intervenes on self-conditioning predictions rather than internal residual-stream activations. Further discussion of activation steering and other related work is provided in App. H. ACE precedes token selection without retraining or an auxiliary model (§4.2).

At an appropriate steering strength, ACE reduces repetition in unconditional generation while retaining competitive quality. On ELF-B, it lowers mean REP-4 from 7.28% to 4.48% while Grammar remains essentially unchanged, from 80.21 to 80.57 (Figure 1(c)). More broadly, ACE reduces repetition across ELF-B/M/L while retaining competitive results on several text-quality metrics beyond Gen-PPL, covering both individual-generation quality and the quality of the generated set as a whole (§5.1). ACE is a simple, theory-motivated intervention that reduces repetition and offers a competitive repetition–quality trade-off against alternative self-conditioning methods (§5.2). Further analysis examines this trade-off across steering choices (§5.3). ACE is robust to changes in inference configurations and random seeds for both direction estimation and generation, and further generalizes to other unconditional self-conditioned continuous diffusion language models (§5.4).

2 BACKGROUND

Continuous diffusion language models generate text by evolving a noisy state through a denoising trajectory in embedding space. At each step, a denoiser processes the current state and a sampler advances it. Self-conditioning (Chen et al., 2022) augments this loop by feeding each step's clean-embedding prediction into the next step. Let $\boldsymbol{z}_{k-1}$ denote the current state at step k, t_k its denoising time, and $\boldsymbol{x}_k$ the clean-embedding prediction at that step. The prediction is updated as

$$\boldsymbol{x}_k = \boldsymbol{X}_\theta\big(\boldsymbol{z}_{k-1}, t_k; \boldsymbol{x}_{k-1}\big), \tag{1}$$

With either deterministic ODE or stochastic SDE updates, the joint denoising step is

$$(\boldsymbol{z}_k, \boldsymbol{x}_k) = \textsc{Step}\big(\boldsymbol{X}_\theta, \boldsymbol{z}_{k-1}, t_k, \boldsymbol{x}_{k-1}\big), \qquad k = 1, \ldots, T, \tag{2}$$

where T is the number of denoising steps and $\boldsymbol{x}_{k-1}$ is the self-conditioning signal.

Representative continuous DLMs with self-conditioning include Plaid (Gulrajani & Hashimoto, 2023), LangFlow (Chen et al., 2026), and ELF (Hu et al., 2026). More recently, ELF uses Flow Matching (Lipman et al., 2022) for language generation, where a learned velocity field transports the state from Gaussian noise toward the clean text-embedding distribution. It follows this path from $t = 0$ to $t = 1$ and decodes tokens only at the terminal step.

For ELF, let $\Delta t_k = t_k - t_{k-1}$. The clean-embedding prediction parameterizes the velocity, and the deterministic ODE sampler uses the Euler update:

$$\boldsymbol{v}_k = \frac{\boldsymbol{x}_k - \boldsymbol{z}_{k-1}}{1 - t_k}, \qquad \boldsymbol{z}_k = \boldsymbol{z}_{k-1} + \Delta t_k \boldsymbol{v}_k. \tag{3}$$

The SDE-inspired sampler first re-injects Gaussian noise and evaluates the prediction at a slightly earlier, noisier time:

$$\alpha_k = 1 - \gamma \Delta t_k, \qquad \bar{t}_k = \alpha_k t_k, \qquad \bar{\boldsymbol{z}}_{k-1} = \alpha_k \boldsymbol{z}_{k-1} + (1 - \alpha_k)\boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{4}$$

and then updates the original state as

$$\boldsymbol{z}_k = \boldsymbol{z}_{k-1} + \Delta t_k \frac{\boldsymbol{x}_k^{\mathrm{SDE}} - \boldsymbol{z}_{k-1}}{1 - t_k}, \tag{5}$$

where $\boldsymbol{x}_k^{\mathrm{SDE}}$ is predicted from $(\bar{\boldsymbol{z}}_{k-1}, \bar{t}_k)$. Generation starts from $\boldsymbol{z}_0 \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$; σ is the initial noise scale and γ controls noise reinjection.

ELF also supports classifier-free guidance (Ho & Salimans, 2021) with self-conditioning, linearly combining conditional and unconditional predictions with a guidance scale w to strengthen the conditioning signal. During unconditional generation, ELF uses its previous clean-embedding prediction as the condition for CFG. Through training-time CFG, $\boldsymbol{X}_\theta$ takes w as additional input and learns to predict the guided clean embedding directly, requiring only one forward pass per denoising step.

3 PROBLEM SETUP AND MOTIVATION

3.1 REPETITIVE GENERATION AND MISLEADING GEN-PPL

Under the standard Full-SC setting, ELF can produce text that repeats the same words and phrases many times (Figure 1(a)). We examine unconditional generations from ELF-B, ELF-M, and ELF-L and measure repetition using the 4-gram self-repetition rate (REP-4; Welleck et al., 2020), defined as the fraction of 4-grams that have already appeared in the same sequence. We tokenize by whitespace. *Mean REP-4* is the mean score across generated samples and summarizes average repetition severity. *Repetition share* is the fraction of samples above a human-calibrated threshold and measures how widespread repetition is. We use the 95th-percentile REP-4 of BBC/XSum (Narayan et al., 2018) articles, 1.92%, as this threshold. All three models exhibit repetition, with ELF-B showing the highest mean REP-4 and repetition share (Table 1). Gen-PPL is perplexity under a pretrained language-model scorer, with lower values interpreted as better generation. However, it decreases as repetition increases (Figure 1(b)), so it can conceal this repetition while biasing quality evaluation.

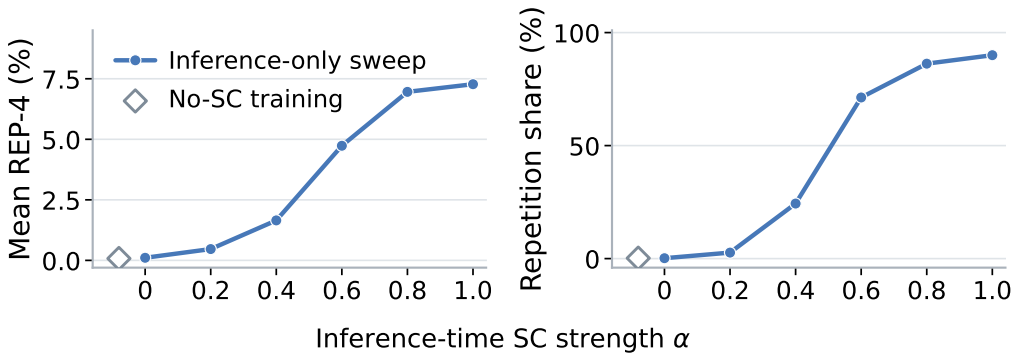


Figure 2: **Increasing self-conditioning strength increases repetition.** Mean REP-4 and repetition share are shown for the inference-only sweep and a model trained without self-conditioning.

Table 1: **ACE reduces repetition across ELF model sizes while maintaining competitive text quality.** Full-SC is compared with ACE using the ELF-B direction $\boldsymbol{d}_B$ shared across sizes or model-specific directions $\boldsymbol{d}_M$ and $\boldsymbol{d}_L$.

Model	Method	REP-4 mean ↓	Rep. share ↓	Gen-PPL	Grammar ↑	Fluency ↑	dist-1 ↑	dist-3 ↑	s-BLEU ↓	MAUVE ↑
ELF-B	Full-SC	$7.28_{\pm 0.35}$	$90.00_{\pm 0.9}$	$20.24_{\pm 0.3}$	$80.21_{\pm 0.2}$	$1.05_{\pm 0.03}$	$43.24_{\pm 0.4}$	$88.19_{\pm 0.4}$	$29.03_{\pm 0.3}$	$23.63_{\pm 1.23}$
	ACE ($\boldsymbol{d}_B$)	$4.48_{\pm 0.19}$	$74.02_{\pm 2.3}$	$23.66_{\pm 0.4}$	$80.57_{\pm 0.4}$	$1.08_{\pm 0.01}$	$46.75_{\pm 0.3}$	$91.93_{\pm 0.3}$	$29.61_{\pm 0.2}$	$19.74_{\pm 1.88}$
ELF-M	Full-SC	$2.99_{\pm 0.15}$	$52.42_{\pm 2.1}$	$21.92_{\pm 0.3}$	$85.97_{\pm 0.2}$	$1.20_{\pm 0.05}$	$47.88_{\pm 0.3}$	$94.04_{\pm 0.2}$	$27.10_{\pm 0.2}$	$25.05_{\pm 1.31}$
	ACE ($\boldsymbol{d}_B$)	$1.79_{\pm 0.07}$	$31.84_{\pm 1.4}$	$24.32_{\pm 0.4}$	$85.56_{\pm 0.2}$	$1.27_{\pm 0.04}$	$50.82_{\pm 0.2}$	$96.00_{\pm 0.1}$	$27.16_{\pm 0.2}$	$22.56_{\pm 2.58}$
	ACE ($\boldsymbol{d}_M$)	$1.80_{\pm 0.08}$	$31.78_{\pm 1.4}$	$24.80_{\pm 0.4}$	$85.15_{\pm 0.2}$	$1.26_{\pm 0.05}$	$51.16_{\pm 0.2}$	$96.02_{\pm 0.1}$	$26.90_{\pm 0.2}$	$21.04_{\pm 1.39}$
ELF-L	Full-SC	$2.50_{\pm 0.13}$	$43.61_{\pm 2.3}$	$23.70_{\pm 0.4}$	$85.83_{\pm 0.2}$	$1.43_{\pm 0.04}$	$50.14_{\pm 0.2}$	$94.94_{\pm 0.2}$	$27.27_{\pm 0.3}$	$23.04_{\pm 1.17}$
	ACE ($\boldsymbol{d}_B$)	$1.58_{\pm 0.04}$	$26.60_{\pm 1.0}$	$25.56_{\pm 0.3}$	$85.60_{\pm 0.2}$	$1.42_{\pm 0.04}$	$52.44_{\pm 0.2}$	$96.45_{\pm 0.1}$	$27.51_{\pm 0.2}$	$19.81_{\pm 1.16}$
	ACE ($\boldsymbol{d}_L$)	$1.56_{\pm 0.08}$	$26.46_{\pm 2.1}$	$25.48_{\pm 0.4}$	$85.63_{\pm 0.2}$	$1.45_{\pm 0.03}$	$52.27_{\pm 0.2}$	$96.49_{\pm 0.1}$	$27.57_{\pm 0.2}$	$19.63_{\pm 1.12}$

3.2 Self-conditioning drives repetition

ELF uses self-conditioning to improve generation quality by refining the clean-embedding prediction across denoising steps (Chen et al., 2022; Hu et al., 2026). We ask whether this quality gain comes with a cost in repetition. To test this, we scale the self-conditioning input as $\boldsymbol{x}_{k-1} \mapsto \alpha \boldsymbol{x}_{k-1}$ and vary α from 0 (off) to 1 (Full-SC), while keeping all other inference settings fixed. Both mean REP-4 and the repetition share increase with α (Figure 2). A separately trained model without self-conditioning also produces almost no repetition, corroborating its causal role.

4 Mechanism and Method

4.1 Mechanism: Self-conditioning forms a repetition attractor with one dominant direction

Section 3 establishes that self-conditioning is a causal driver of repetition. To understand how this self-conditioning feedback produces the defect, we decode intermediate states along a Full-SC trajectory and track their repetition. The Full-SC curve in Figure 4(d) rises rapidly and then levels off, suggesting an attractor-like regime. We model this regime locally with an attracting fixed point.

A local model of the self-conditioning loop. Near this fixed point, we view self-conditioning as a feedback loop. The key idea is to separate the response to the fed-back clean-embedding prediction from the remaining denoising drive. Let $\boldsymbol{x}_k \in \mathbb{R}^{L \times e}$ denote the clean-embedding prediction at denoising step k, and $\boldsymbol{u}_k \in \mathbb{R}^e$ its mean over L positions. With denoising conditions locally fixed,

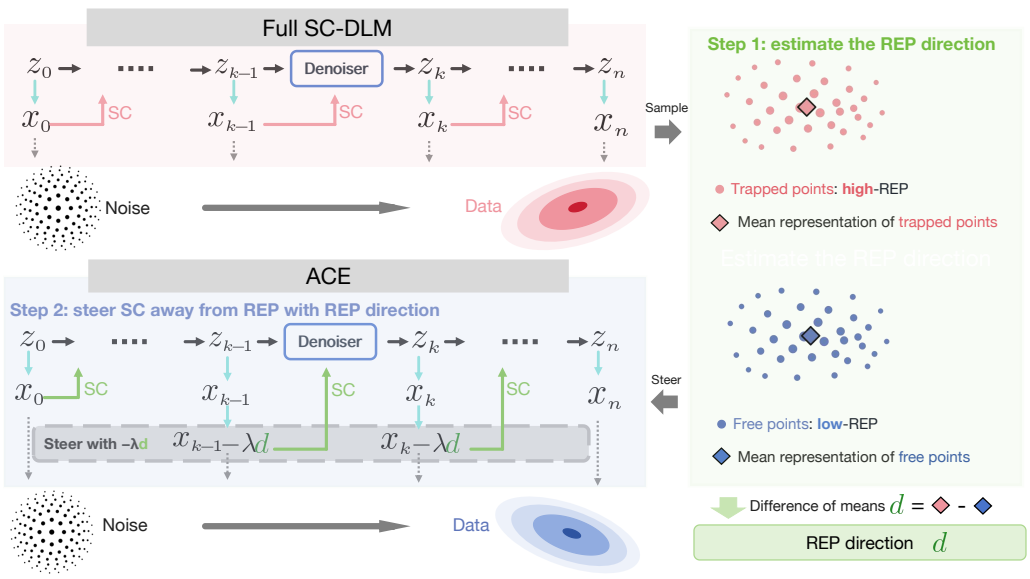


Figure 3: **ACE steers self-conditioning away from repetition.** Full-SC carries the self-conditioning feedback through denoising, whereas ACE first estimates a repetition direction from trapped (high-REP) and free (low-REP) trajectories, then subtracts it from feedback at each step.

we model the update of $\boldsymbol{u}_k$ as a self-conditioning response plus a denoising drive:

$$\boldsymbol{u}_k \;=\; \tfrac{1}{L}\textstyle\sum_{l=1}^{L} \boldsymbol{x}_k[l] \;\in\; \mathbb{R}^e, \qquad \boldsymbol{u}_{k+1} \;=\; \underbrace{\boldsymbol{s}(\boldsymbol{u}_k)}_{\text{self-conditioning}} \;+\; \underbrace{\boldsymbol{f}_k}_{\text{denoising drive}} . \tag{6}$$

Here $\boldsymbol{s}$ denotes the self-conditioning response defined around the fixed point, and $\boldsymbol{f}_k$ is the remaining, locally fixed denoising drive. This decomposition is defined in App. B.

Theoretically, linearization predicts a dominant convergence direction within the repetition attractor. To understand how this fixed point shapes nearby denoising updates, we use a local fixed-point approximation: we assume a smooth, contracting self-conditioning response and use first-order linearization to compare the contraction rates along different directions.

Assumption 1 (Local fixed point)**.** *The self-conditioning map $\boldsymbol{s}$ has a fixed point $\boldsymbol{u}^\star$ and is C^1 near it with a contracting Jacobian $\boldsymbol{J} = \mathrm{D}\boldsymbol{s}(\boldsymbol{u}^\star)$. For the analysis we idealize $\boldsymbol{J}$ as symmetric, with orthonormal eigenvectors $\{\boldsymbol{v}_i\}$ and real eigenvalues $1 > \mu_1 > \mu_2 \geq \cdots \geq 0$. The leading eigenvector $\boldsymbol{v}_1$ is the slowest-contracting direction, and $\rho := 1 - \mu_1 \in (0, 1)$ is its contraction rate.*

The Jacobian describes how perturbations in the fed-back prediction propagate through the self-conditioning loop. Linearization gives (Lemmas 1–2, App. B)

$$\Delta\boldsymbol{u}_k \;\approx\; \underbrace{\underbrace{\beta_k \boldsymbol{v}_1}_{\text{leading}} + \underbrace{\boldsymbol{r}_k}_{\text{decays}}}_{\text{dominant mode}} + \underbrace{\boldsymbol{f}_k}_{\text{denoising drive}} \tag{7}$$

Here $\boldsymbol{r}_k$ collects feedback components orthogonal to $\boldsymbol{v}_1$. To characterize convergence, we hold the drive fixed ($\boldsymbol{f}_k \equiv \boldsymbol{f}$). The linearized loop converges to a limiting pooled clean-embedding prediction $\boldsymbol{u}_\infty$ from an initial prediction $\boldsymbol{u}_0$. Lemma 3 gives its expression and derives the following eigenbasis expansion for deviations from the limit (App. B.1):

$$\boldsymbol{u}_k - \boldsymbol{u}_\infty \;=\; \sum_i \mu_i^k \,\langle \boldsymbol{u}_0 - \boldsymbol{u}_\infty, \boldsymbol{v}_i \rangle\, \boldsymbol{v}_i. \tag{8}$$

Under Assumption 1, the $\boldsymbol{v}_1$ component decays most slowly. If the initial deviation has a nonzero component in this direction, it eventually dominates the remaining deviation, yielding asymptotically one-dimensional convergence toward $\boldsymbol{u}_\infty$ along $\boldsymbol{v}_1$.

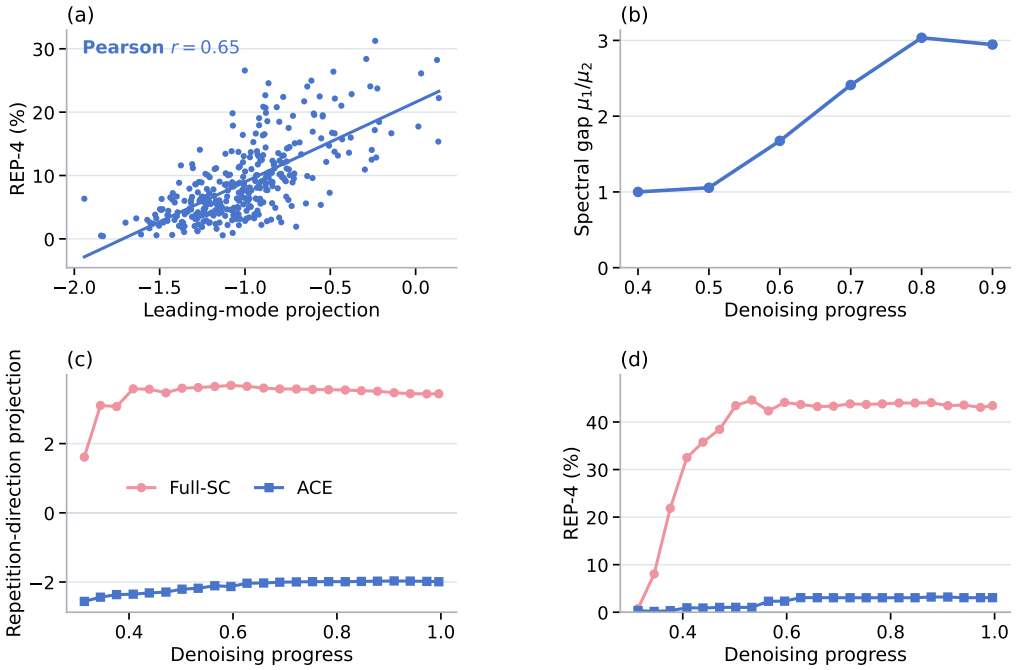


Figure 4: **A dominant self-conditioning mode accompanies repetition, while ACE redirects generation away from the estimated repetition direction.** (a) Across samples, REP-4 increases with the projection of the self-conditioning signal onto its leading mode; (b) the leading mode separates increasingly from the second mode as denoising progresses. (c) Full-SC gradually builds a stable positive projection along the estimated repetition direction associated with the leading mode, whereas ACE suppresses it. (d) Accordingly, REP-4 rises and plateaus under Full-SC but stays low under ACE.

Empirically, this dominant direction emerges with the attractor and predicts repetition. To characterize how this dominant direction relates to repetition and emerges during denoising, we analyze self-conditioning feedback collected along Full-SC trajectories, comparing its projection with repetition and tracking the local Jacobian spectrum. Across samples, the projection of self-conditioning feedback onto $\boldsymbol{v}_1$ tracks repetition (Fig. 4(a)), while the spectral gap rises as the denoising trajectory enters the repetitive regime and $\boldsymbol{v}_1$ becomes dominant (Fig. 4(b), Tab. 4). To test whether the feedback–repetition association and widening spectral gap are consistent with a local attractor, we hold the denoising conditions in Eq. 6 fixed and repeatedly apply the update. This brings the pooled prediction close to a fixed point (Fig. 11a). Across denoising fractions, the leading feedback gain remains below one, while the dominant mode accounts for a larger share of the local response in the later repetitive regime (Fig. 11b,c). Together, these findings support a local attractor mechanism with effectively one-dimensional repetition-related feedback. This structure motivates the directional intervention introduced in the next subsection.

4.2 Method: ACE redirects generation away from the dominant repetition direction

The one-dimensional structure suggests a minimal intervention: estimate the repetition direction and shift the self-conditioning feedback away from it, leaving orthogonal components unchanged. We recover this direction by contrasting average self-conditioning feedback across Full-SC trajectories. The resulting direction is fixed once and subtracted from the self-conditioning feedback at each denoising step with strength λ. We call this procedure **Attractor-Contrast-Escape (ACE)**. Figure 3 contrasts the unsteered Full-SC process with ACE's direction estimation and steering stages.

Concretely, we estimate the repetition direction $\boldsymbol{d}$ once from N Full-SC calibration samples. For each sample, we average its self-conditioning feedback over denoising steps and sequence positions.

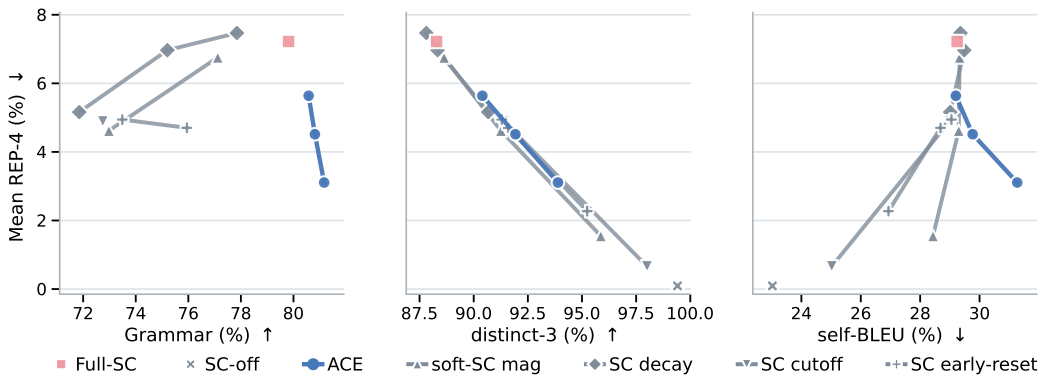


Figure 5: **Self-conditioning interventions exhibit repetition–quality trade-offs, with ACE achieving a favorable balance.** The panels compare repetition with complementary quality and diversity measures across methods; lines connect tested settings within each method. Only points within the displayed ranges are shown; full sweeps and additional metrics are reported in App. D.

Using a tail fraction $q \in (0, 1/2]$, we form trapped (high-REP) and free (low-REP) groups from the top and bottom q fractions of calibration samples ranked by REP-4, respectively. The direction is the normalized difference between the two group means. We assume that these groups differ mainly along the dominant mode, while the mean contribution of the remaining modes is shared between the groups. Under this separability assumption, the difference of means recovers the dominant mode $\boldsymbol{v}_1$ up to scale (Prop. 1, App. B). Empirically, $\boldsymbol{d}$ partially aligns with the leading mode measured along denoising trajectories (App. E); ACE uses this trajectory-level estimate for steering. During generation, we keep $\boldsymbol{d}$ fixed and pass $\tilde{\boldsymbol{x}}_k = \boldsymbol{x}_k - \lambda \mathbf{1}_L \boldsymbol{d}^\top$ to the next step. See Algorithm 1 (App. E).

5 Experiments

5.1 Main results: reducing repetition while preserving quality

We evaluate whether ACE reduces repetition while maintaining generation quality. We report mean REP-4 and repetition share to measure repetition, and assess generation quality at two levels. For individual generations, Grammar and Fluency assess linguistic quality, while distinct-1 and distinct-3 measure within-text diversity; for the generated set as a whole, self-BLEU measures cross-sample diversity and MAUVE measures similarity to the reference distribution. Evaluation details are provided in App. A. All results use a common 64-step SDE setting with initial noise scale σ=2.0, churn γ=1.0, and SC-CFG scale w=3. We choose ACE's calibration size N and tail fraction q by assessing the stability of the estimated direction, using adjacent-size cosine across N and cross-seed agreement across q. As N grows from 256 to 4096, the mean adjacent-size absolute cosine across the plotted curves rises from 0.917 to 0.995, indicating that the direction estimates stabilize at N=4096 (App. Fig. 8). Among the tested tail fractions $q \in \{1/8, 1/4, 1/2\}$, q=1/4 achieves the highest cross-seed agreement (App. Fig. 10a). We then select steering strength λ based on the repetition–quality trade-off; the strength-response analysis supports λ=1/2 as a quality-focused choice (Figure 6). We therefore use N=4096, q=1/4, and λ=1/2 for the main results. Metrics aggregate five held-out generation seeds with 1,024 samples per seed. Table 1 summarizes the main comparison. On ELF-B, ACE reduces mean REP-4 from 7.28% to 4.48% and repetition share from 90.00% to 74.02%. Gen-PPL rises after steering as repetition is reduced, while the other quality measures remain competitive. Panels (c,d) of Figure 4 show this pattern along one ELF-B trajectory: Full-SC's feedback projection onto $\boldsymbol{d}$ remains positive as REP-4 rises, whereas ACE shifts it to the opposite side and reduces repetition. Paired qualitative examples illustrating this contrast are provided in App. G. This quality preservation is consistent with the directional intervention: ACE shifts feedback along $\boldsymbol{d}$ while leaving orthogonal components unchanged at each intervention (App. B). As a complementary check, applying the same direction as a training-time penalty further reduces repetition beyond continued training alone; the training setup and full results are reported in App. C. Table 1 shows that ACE remains effective on ELF-M and ELF-L: using model-specific directions $\boldsymbol{d}_M$ and $\boldsymbol{d}_L$, it reduces repetition while retaining competitive quality. It also shows that

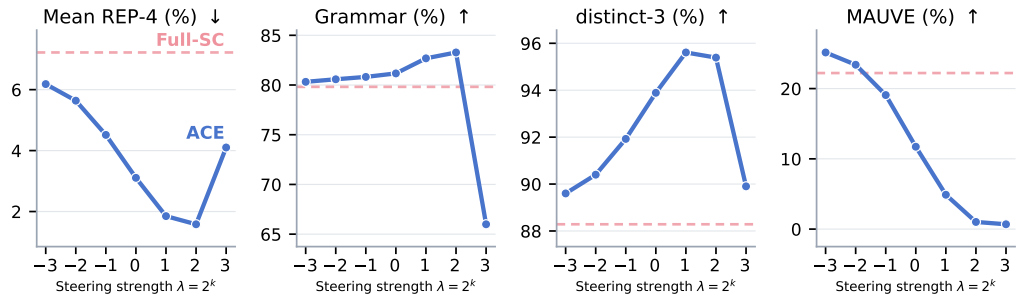


Figure 6: **A moderate steering strength gives a favorable repetition–quality trade-off.** The four panels compare repetition, single-generation quality, and sample-pool quality against Full-SC across steering strengths. Further strength and direction analyses are provided in App. E.

the ELF-B direction $\boldsymbol{d}_B$ can be shared with both models, where ACE likewise reduces repetition while retaining competitive quality. Fig. 8(c) further shows that the directions estimated separately for the three model sizes are aligned. This suggests a systematic repetition bias across unconditional continuous diffusion language models. Transfer across settings and seeds is examined in §5.4.

5.2 Comparison with alternative repetition interventions

Several alternative interventions modify the self-conditioning feedback and can also reduce repetition. SC-off disables the self-conditioning feedback during inference, while soft-SC mag rescales its magnitude. SC decay gradually attenuates the feedback over denoising, and SC cutoff disables the feedback for all samples after a fixed denoising fraction. SC early-reset checks for repetition midway through denoising, then sets the self-conditioning feedback to zero for the remaining steps of any sample that has already begun repeating. We include the Full-SC sampler for reference. We compare the methods under the same held-out generation seed, with $1{,}024$ direct generations per setting. Figure 5 plots mean REP-4 against Grammar, distinct-3, and MAUVE; the full parameter grid and complementary metrics are reported in App. D. These results show a trade-off: interventions that suppress feedback more broadly can reduce repetition sharply but may compromise generation quality, while softer changes can leave more repetition. Notably, ACE employs a simple theory-guided intervention to suppress repetition relative to Full-SC while retaining competitive quality, thereby yielding a favorable repetition–quality trade-off.

5.3 Analysis of ACE's steering strength and direction

With the estimated repetition direction fixed, we vary the steering strength λ and measure repetition and quality. Figure 6 shows that increasing λ initially suppresses repetition while slightly improving single-generation quality; sample-pool quality remains stable or slightly improves at weak strengths. At larger strengths, however, further repetition reduction can come at the cost of sample-pool quality, even when single-generation quality improves. When the steering strength becomes too large, single-generation quality also deteriorates and repetition rebounds, suggesting over-steering. Across the range of steering strengths, we compare direction estimators motivated by our repetition-attractor theory, including ACE, Logistic, LDA, and the Jacobian mode. Logistic and LDA provide discriminative estimates of the theory's repetition direction, whereas the Jacobian mode extracts the dominant local feedback mode predicted by the same analysis. For comparison, we also evaluate alternatives that do not use the repetition-attractor hypothesis. Top-PC extracts the leading singular direction of pooled feedback, using REP-4 to select its calibration subset and orient its sign. Label-permuted directions use the same REP-based estimator with the repetition labels shuffled, and random directions ignore the feedback structure. Figure 10 and Table 7 show that the reduction depends on alignment with the repetition contrast: at suitable steering strengths, theory-motivated estimators reduce repetition while retaining competitive quality, whereas theory-agnostic controls do not reliably reproduce this behavior. ACE provides a simple, stable, and effective implementation of this principle; further details appear in App. E.

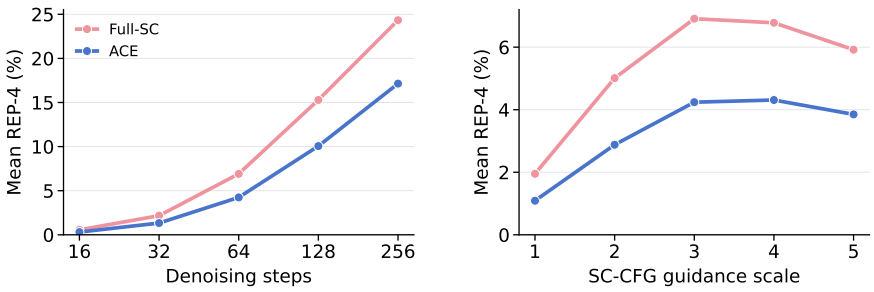


Figure 7: **ACE suppresses repetition across denoising steps and guidance scales.** Mean REP-4 for Full-SC (pink) and ACE (blue) as the denoising steps and SC-CFG guidance scale vary.

Table 2: **ACE reduces repetition across other self-conditioned continuous DLMs.** Full-SC is compared with ACE on LangFlow and Plaid using model-specific directions.

Model	Method	REP-4 mean ↓	Rep. share ↓	Grammar ↑	Fluency ↑	dist-1 ↑	dist-3 ↑	s-BLEU ↓	MAUVE ↑
LangFlow	Full-SC	3.98	35.74	61.56	1.00	52.31	93.76	18.90	14.44
	ACE	2.72	25.39	61.67	1.01	53.65	95.44	20.03	15.59
Plaid	Full-SC	3.95	30.30	89.48	2.04	51.09	93.93	19.68	15.52
	ACE	2.82	25.20	89.75	2.12	51.01	95.22	20.95	14.73

5.4 Robustness and Generalization

We first test the robustness of ACE to changes in inference configurations and random seeds within ELF. Keeping the shared ELF-B direction $\boldsymbol{d}_B$ fixed, we vary the number of denoising steps, SC-CFG guidance scale, initial noise scale σ, sampler, and churn γ, and evaluate selected settings across independent held-out generation seeds. We additionally re-estimate $\boldsymbol{d}_B$ from five independent calibration seeds to test the stability of the direction itself. ACE continues to reduce repetition while retaining competitive quality across these settings (Figure 7 and Table 8). The repetition gap is especially pronounced at larger denoising-step counts and guidance scales, where Full-SC trajectories are driven deeper into the repetition basin, while ACE keeps them away from it (Figure 7). Table 1 already shows that ACE generalizes across ELF model sizes. We next ask whether the same intervention extends beyond the ELF family. LangFlow (Chen et al., 2026) and Plaid (Gulrajani & Hashimoto, 2023) are self-conditioned continuous diffusion language models with distinct text representations. Using a direction estimated separately for each model, ACE reduces repetition while retaining competitive quality on most reported measures (Table 2). Detailed settings and steering-strength trade-offs are reported in App. F.3. Together, these results show that ACE is robust and can generalize across models.

6 Conclusion

We find that self-conditioned continuous diffusion language models can produce substantial repetition in unconditional generation, a defect that Gen-PPL can conceal. We trace this repetition to an effectively one-dimensional contractive attractor in the self-conditioning loop. Motivated by this mechanism, we propose Attractor-Contrast-Escape (ACE), a training-free inference-time intervention that estimates a repetition direction by contrasting denoising paths trapped in repetition with paths relatively free of repetition, then subtracts it from the self-conditioning feedback during denoising. At appropriate steering strengths, ACE reduces repetition while retaining competitive text quality across ELF model sizes, inference configurations, and random seeds. It also mitigates repetition in other self-conditioned continuous DLMs. These results identify self-conditioning feedback as a source of repetition and demonstrate that targeted intervention during denoising can mitigate it.

AI USE STATEMENT

In this work, we used generative AI tools to assist with literature discovery; research ideation, experimental design, and code implementation and debugging; interpretation of experimental results; mathematical reasoning and review of proof arguments; and drafting and editing portions of the manuscript. The authors reviewed the AI-assisted contributions and take responsibility for the final text, analyses, and claims.

REFERENCES

Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. Analog bits: Generating discrete data using diffusion models with self-conditioning. *arXiv preprint arXiv:2208.04202*, 2022.

Yuxin Chen, Chumeng Liang, Hangke Sui, Ruihan Guo, Chaoran Cheng, Jiaxuan You, and Ge Liu. Langflow: Continuous diffusion rivals discrete in language modeling. *arXiv preprint arXiv:2604.11748*, 2026.

Antonio Franca and Alexander Tong. Hacking generative perplexity: Why unconditional text evaluation needs distributional metrics. *arXiv preprint arXiv:2606.08417*, 2026.

Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.

Ishaan Gulrajani and Tatsunori B Hashimoto. Likelihood-based diffusion language models. *Advances in Neural Information Processing Systems*, 36:16693–16715, 2023.

Hongcan Guo, Qinyu Zhao, Yian Zhao, Shen Nie, Rui Zhu, Qiushan Guo, Feng Wang, Tao Yang, Hengshuang Zhao, Guoqiang Wei, et al. Continuous latent diffusion language model. *arXiv preprint arXiv:2605.06548*, 2026.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. The curious case of neural text degeneration. In *International Conference on Learning Representations*, 2020.

Keya Hu, Linlu Qiu, Yiyang Lu, Hanhong Zhao, Tianhong Li, Yoon Kim, Jacob Andreas, and Kaiming He. Elf: Embedded language flows. *arXiv preprint arXiv:2605.10938*, 2026.

Jaehyeong Jo and Sung Ju Hwang. Continuous diffusion model for language modeling. In *Advances in Neural Information Processing Systems*, volume 38, pp. 108049–108085, 2025. URL https://papers.nips.cc/paper_files/paper/2025/hash/8c4046a93293442e88f54e19a5849d68-Abstract-Conference.html.

Kalpesh Krishna, John Wieting, and Mohit Iyyer. Reformulating unsupervised style transfer as paraphrase generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 737–762, 2020.

Bocheng Li and Linli Xu. Improving few-step language flows with untied self-conditioning. *arXiv preprint arXiv:2608.22244*, 2026.

Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. Diffusion-lm improves controllable text generation. *Advances in neural information processing systems*, 35: 4328–4343, 2022.

Jiajun Liang, Yucheng Liao, Yukang Cao, Jiazhe Wei, Ken Li, Wende Tan, Jiankun Zhang, ZY Cui, Jingkang Yang, Liucheng Guo, et al. Aurora-lm: Autoencoding unified representation for continuous-latent diffusion language modeling. *arXiv preprint arXiv:2608.02602*, 2026.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Shashi Narayan, Shay B Cohen, and Mirella Lapata. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 1797–1807, 2018.

Dhruvesh Patel, Aishwarya Sahoo, Avinash Amballa, Tahira Naseem, Tim GJ Rudner, and Andrew McCallum. Insertion language models: Sequence generation with arbitrary-position insertions. *arXiv preprint arXiv:2505.05755*, 2025.

Krishna Pillutla, Swabha Swayamdipta, Rowan Zellers, John Thickstun, Sean Welleck, Yejin Choi, and Zaid Harchaoui. Mauve: Measuring the gap between neural text and human text using divergence frontiers. *Advances in Neural Information Processing Systems*, 34:4816–4828, 2021.

Patrick Pynadath, Jiaxin Shi, and Ruqi Zhang. Generative frontiers: Why evaluation matters for diffusion language models. *arXiv preprint arXiv:2604.02718*, 2026.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Nina Rimsky, Nick Gabrieli, Julian Schulz, Meg Tong, Evan Hubinger, and Alexander Turner. Steering llama 2 via contrastive activation addition. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15504–15522, 2024.

Junzhe Shen, Jieru Zhao, Ziwei He, and Zhouhan Lin. Codar: Continuous diffusion language models are more powerful than you think. *arXiv preprint arXiv:2603.02547*, 2026.

Adi Shnaidman, Erin Feiglin, Osher Yaari, Efrat Mentel, Amit Levi, and Raz Lapid. Activation steering for masked diffusion language models. *arXiv preprint arXiv:2512.24143*, 2025.

Robin Strudel, Corentin Tallec, Florent Altché, Yilun Du, Yaroslav Ganin, Arthur Mensch, Will Grathwohl, Nikolay Savinov, Sander Dieleman, Laurent Sifre, et al. Self-conditioned embedding diffusion for text generation. *arXiv preprint arXiv:2211.04236*, 2022.

Yixuan Su, Tian Lan, Yan Wang, Dani Yogatama, Lingpeng Kong, and Nigel Collier. A contrastive framework for neural text generation. *Advances in Neural Information Processing Systems*, 35: 21548–21561, 2022.

Yequan Wang, Jiawen Deng, Aixin Sun, and Xuying Meng. Perplexity from plm is unreliable for evaluating text quality. *arXiv preprint arXiv:2210.05892*, 2022.

Alex Warstadt, Amanpreet Singh, and Samuel R Bowman. Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, 7:625–641, 2019.

Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. Neural text generation with unlikelihood training. In *International Conference on Learning Representations*, 2020.

Jin Xu, Xiaojiang Liu, Jianhao Yan, Deng Cai, Huayang Li, and Jian Li. Learning to break the loop: Analyzing and mitigating repetitions for neural text generation. *Advances in Neural Information Processing Systems*, 35:3082–3095, 2022.

Jaehoon Yoo, Wonjung Kim, Floor Eijkelboom, Chanhyuk Lee, Nicholas M Boffi, Seunghoon Hong, and Jinwoo Kim. Self-conditioned flow map language models via fixed-point flows. *arXiv preprint arXiv:2607.00714*, 2026.

A Evaluation Protocol and Metric Definitions

Unless specified otherwise, we evaluate each model on $1{,}024$ samples generated with each held-out seed, using directions estimated from separate calibration samples. For multi-seed evaluations, we report mean scores over five seeds; accompanying $\pm$ values denote sample standard deviations across seeds. The canonical ELF setting is 64 SDE denoising steps with initial noise scale $\sigma{=}2.0$, churn $\gamma{=}1.0$, and SC-CFG scale $w{=}3$. We consider a generated sample repetitive if its REP-4 exceeds the 95th-percentile REP-4 of BBC/XSum articles (1.92%). Table 3 summarizes the implementation and aggregation of every metric reported in the paper. REP-4, repetition share, Grammar, distinct-1, distinct-3, self-BLEU, and MAUVE are reported in percentage points in the tables; Gen-PPL is left on its native scale, and Fluency is reported on its 1–5 judge scale. Following the

reference-free unconditional-generation evaluation setup of Patel et al. (2025), we adapt its prompt and fluency rubric to GPT-4o. The box below shows the full judging prompt, with {generation} replaced by the generated text. We extract the integer following [RESULT] as the score. GPT-4o uses temperature 0, top-p 0.9, and a maximum completion length of 1,024 tokens. All metric implementations and the raw per-sample Fluency records are retained with the corresponding evaluation manifests.

Fluency evaluation prompt

System message

You are a fair judge assistant tasked with providing clear, objective feedback based on specific criteria, ensuring each assessment reflects the absolute standards set for performance.

User message

###Task Description:

An unconditional generation to evaluate, and a score rubric representing an evaluation criteria are given.

1. Write a detailed feedback that assesses the quality of the generation strictly based on the given score rubric, not evaluating in general.
2. After writing a feedback, write a score that is an integer between 1 and 5. You should refer to the score rubric.
3. The output format should look as follows: "(write a feedback for criteria) [RESULT] (an integer number between 1 and 5)".
4. Please do not generate any other opening, closing, or explanations.

Generation to evaluate:

{generation}

Score Rubrics:

Fluency (Is the text fluent and easy to read?)

Score of 1: The text is disjointed and lacks fluency, making it hard to follow.

Score of 2: The text has limited fluency with frequent awkward phrasing.

Score of 3: The text is moderately fluent, with some awkward phrasing but generally easy to follow.

Score of 4: The text is fluent with smooth transitions and rare awkward phrases.

Score of 5: The text is highly fluent, with natural and smooth expression throughout.

Feedback:

B Formal analysis and empirical validation of the repetition attractor

This appendix formalizes the local picture developed in §4. We first construct a local model of the self-conditioning loop, linearize it near a fixed point, separate the self-conditioning response from the denoising drive, and characterize the resulting driven fixed point. We also extend the fixed-point result locally to the nonlinear map and discuss its quasi-static interpretation when the local coefficients vary slowly (§B.1). Under a separability assumption, we then show how the difference between mean self-conditioning feedback from trapped and free Full-SC trajectories recovers the dominant direction used to steer repetition; empirical alignment is approximate (§B.2). Finally, we test numerical convergence of the frozen pooled map and estimate the local Jacobian on the trained model to assess contraction, mode dominance, and separation of the denoising drive from the self-conditioning response (§B.3). These conclusions are local: the linear analysis assumes a symmetric Jacobian and treats the latent trajectory as exogenous, with the drive frozen when characterizing the fixed point. Direction recovery additionally relies on separability; the analysis does not provide a global decomposition of the denoising trajectory.

Table 3: Evaluation metrics and their aggregation protocols.

Metric	Implementation	Aggregation and interpretation
REP-4 and repetition share	Whitespace-tokenized 4-grams; REP-4 is the fraction of occurrences that duplicate an earlier 4-gram.	We report the mean REP-4 over all generated samples and the fraction of samples above the 1.92% human-calibrated threshold.
Gen-PPL	GPT-2 Large (Radford et al., 2019) scores each generated token with a 1,024-token evaluation context.	The exponential of the corpus-level mean token negative log-likelihood. Lower values mean that the scorer finds the sample corpus more predictable.
Grammar	A RoBERTa-large classifier trained on CoLA (Warstadt et al., 2019; Krishna et al., 2020), loaded from the `roberta-large-cola-krishna2020` checkpoint.	We split each generation into sentences of at least three words, truncate each sentence to the classifier's 512-token input limit (including special tokens), score its probability of acceptability, average within each text, and then average over texts.
Fluency	GPT-4o directly assesses the fluency of each unconditional generation.	We average the parsed scores over generations and then held-out seeds; scores range from 1 to 5, with higher values indicating greater fluency.
distinct-1 and distinct-3	For each text, unique whitespace-token n-grams divided by total n-grams.	We first average the per-text ratios and then average across the sample pool; larger values indicate greater within-text diversity.
self-BLEU	Smoothed BLEU-4 against remaining texts, using whitespace tokens and up to 1,024 leave-one-out references.	We average the per-text scores over the pool. Because this is a similarity measure, lower values indicate less cross-sample similarity.
MAUVE	GPT-2-Large features and MAUVE divergence-frontier estimator (Pillutla et al., 2021); 1,024 OpenWebText references; generation cap of 1,024 tokens.	We report the resulting distributional score on a 0–100 scale as an auxiliary comparison with the chosen reference corpus.

B.1 Local model of the self-conditioning loop

Local feedback map. To analyze the feedback at the sequence level, we pool the clean-embedding predictions in equation 1 and use Assumption 1. Let $\overline{(\cdot)} = \frac{1}{L}\sum_{l=1}^{L}(\cdot)[l]$ denote position pooling and define the pooled model map

$$\overline{\boldsymbol{X}}_\theta(\boldsymbol{z}, t; \boldsymbol{u}) := \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{X}_\theta(\boldsymbol{z}, t; \mathbf{1}_L \boldsymbol{u}^\top)[l] \in \mathbb{R}^e, \tag{9}$$

where $\mathbf{1}_L \boldsymbol{u}^\top$ denotes the self-conditioning feedback broadcast to all positions. Under this pooled approximation, the feedback update at step $k+1$ is:

$$\boldsymbol{u}_{k+1} = g_{k+1}(\boldsymbol{u}_k) := \overline{\boldsymbol{X}}_\theta(\boldsymbol{z}_k, t_{k+1}; \boldsymbol{u}_k). \tag{10}$$

The subscript on g_{k+1} records the changing state and denoising time. In the local linearized model, we treat the denoising conditions as exogenous to the feedback input and use a common Jacobian, while allowing the drive to vary across steps. The fixed-point result below additionally assumes a constant drive. Because $\boldsymbol{u}_k$ enters through the fed-back channel, we expand the pooled map about the fixed point $\boldsymbol{u}^\star$:

$$\overline{\boldsymbol{X}}_\theta(\boldsymbol{z}_k, t_{k+1}; \boldsymbol{u}_k) \;=\; \underbrace{\overline{\boldsymbol{X}}_\theta(\boldsymbol{z}_k, t_{k+1}; \boldsymbol{u}^\star)}_{\text{zeroth order}} \;+\; \underbrace{\boldsymbol{J}\,(\boldsymbol{u}_k - \boldsymbol{u}^\star)}_{\text{first order}} \;+\; o(\|\boldsymbol{u}_k - \boldsymbol{u}^\star\|). \tag{11}$$

The three objects used throughout are the terms of equation 11: the *denoising drive* $\boldsymbol{f}_k := \overline{\boldsymbol{X}}_\theta(\boldsymbol{z}_k, t_{k+1}; \boldsymbol{u}^\star) - \boldsymbol{u}^\star$, which is independent of $\boldsymbol{u}_k$ under the frozen local model, the self-conditioning Jacobian $\boldsymbol{J} := \partial\overline{\boldsymbol{X}}_\theta/\partial\boldsymbol{u}\big|_{\boldsymbol{u}^\star}$, and the self-conditioning map $\boldsymbol{s}(\boldsymbol{u}) := \boldsymbol{u}^\star + \boldsymbol{J}(\boldsymbol{u} - \boldsymbol{u}^\star) + o(\|\boldsymbol{u} - \boldsymbol{u}^\star\|)$, which fixes $\boldsymbol{u}^\star$ by construction. Collecting them gives the stepwise split, with the local-map index suppressed:

$$g(\boldsymbol{u}_k) \;=\; \boldsymbol{s}(\boldsymbol{u}_k) + \boldsymbol{f}_k. \tag{12}$$

Within this local model, the latent trajectory $\{\boldsymbol{z}_k\}$ is treated as exogenous, so $\boldsymbol{f}_k$ acts as a $\boldsymbol{u}_k$-independent forcing. This decoupling approximation, together with the local Jacobian assumptions, is evaluated empirically in App. B.3.

Let $\boldsymbol{a}_k = \boldsymbol{u}_k - \boldsymbol{u}^\star$ be the residual, with coordinates $a_k^{(i)} = \langle \boldsymbol{a}_k, \boldsymbol{v}_i \rangle$ in the eigenbasis of Assumption 1; by the symmetric-$\boldsymbol{J}$ idealization $\boldsymbol{v}_1$ is both a left and right eigenvector, so the $\boldsymbol{v}_1$-coordinate decouples below.

Lemma 1 (Linearized loop)**.** *Near the fixed point the residual obeys the linear recursion*

$$\boldsymbol{a}_{k+1} \;=\; \boldsymbol{J}\,\boldsymbol{a}_k + \boldsymbol{f}_k, \tag{13}$$

the residual form of the loop equation 6 near $\boldsymbol{u}^\star$.

Proof. Subtract $\boldsymbol{u}^\star$ from the split equation 12 and use $\boldsymbol{s}(\boldsymbol{u}_k) = \boldsymbol{u}^\star + \boldsymbol{J}\boldsymbol{a}_k + o(\|\boldsymbol{a}_k\|)$ from equation 11: $\boldsymbol{a}_{k+1} = \boldsymbol{J}\boldsymbol{a}_k + \boldsymbol{f}_k + o(\|\boldsymbol{a}_k\|)$, which is equation 13 once the higher-order remainder is dropped (exact for the idealized linear model). □

Lemma 2 (Per-step decomposition)**.** *Within the linearized model, the one-step change splits as*

$$\Delta\boldsymbol{u}_k \;=\; \beta_k\,\boldsymbol{v}_1 \;+\; \boldsymbol{r}_k \;+\; \boldsymbol{f}_k, \tag{7}$$

with repetition coefficient $\beta_k = (\mu_1 - 1)\,a_k^{(1)}$ and remainder $\boldsymbol{r}_k = \sum_{i\geq 2}(\mu_i - 1)\,a_k^{(i)}\boldsymbol{v}_i$. The remainder collects self-conditioning modes orthogonal to $\boldsymbol{v}_1$, which contract faster ($\mu_i \leq \mu_1$ for $i \geq 2$); the denoising drive $\boldsymbol{f}_k$ remains a separate term.

Proof. From Lemma 1, $\Delta\boldsymbol{u}_k = \boldsymbol{u}_{k+1} - \boldsymbol{u}_k = \boldsymbol{a}_{k+1} - \boldsymbol{a}_k = (\boldsymbol{J} - \boldsymbol{I})\boldsymbol{a}_k + \boldsymbol{f}_k$. Expanding the residual in the orthonormal eigenbasis (Assumption 1), $\boldsymbol{a}_k = \sum_i a_k^{(i)}\boldsymbol{v}_i$, gives $(\boldsymbol{J} - \boldsymbol{I})\boldsymbol{a}_k = \sum_i(\mu_i - 1)a_k^{(i)}\boldsymbol{v}_i$; separating the i=1 term as $\beta_k\boldsymbol{v}_1$ from the rest as $\boldsymbol{r}_k$ yields the stated identity. Projecting equation 13 on $\boldsymbol{v}_i$ gives the scalar recursion $a_{k+1}^{(i)} = \mu_i a_k^{(i)} + f_k^{(i)}$. Its homogeneous transient decays as μ_i^k, fastest for the smallest μ_i; consequently, the orthogonal transients ($\mu_i \leq \mu_1$, $i \geq 2$) die out at least as fast as the $\boldsymbol{v}_1$ transient. When the drive is frozen, each coordinate settles at the driven offset $f^{(i)}/(1 - \mu_i)$, as made explicit in Lemma 3. Thus $\boldsymbol{v}_1$ is the slowest-decaying self-conditioning mode, while the drive contribution remains separate. □

Lemma 3 (Where the driven loop settles)**.** *With the drive frozen at $\boldsymbol{f}_k \equiv \boldsymbol{f}$, the recursion equation 13 converges geometrically (at rate μ_1) to the unique fixed point $\boldsymbol{a}_\infty = (\boldsymbol{I} - \boldsymbol{J})^{-1}\boldsymbol{f}$, i.e.*

$$\boldsymbol{u}_\infty = \boldsymbol{u}^\star + (\boldsymbol{I} - \boldsymbol{J})^{-1}\boldsymbol{f}. \tag{14}$$

Its distance from $\boldsymbol{u}^\star$ along the repetition axis is $|a_\infty^{(1)}| = |f^{(1)}|/(1 - \mu_1) = |f^{(1)}|/\rho$ with $f^{(1)} = \langle \boldsymbol{f}, \boldsymbol{v}_1 \rangle$. For fixed $\rho > 0$, a small but nonzero $f^{(1)}$ yields a correspondingly small but nonzero offset along $\boldsymbol{v}_1$, rather than collapse along this mode. The empirical size of this axial component is evaluated in App. B.3. For fixed $f^{(1)}$, this offset decreases as ρ increases; a near-marginal mode permits a larger offset. The repetition interpretation is supported by the empirical association between leading-mode projection and REP (Fig. 4(a)).

Proof. $\boldsymbol{J}$ is contracting, so its spectral radius $\mu_1 < 1$ and $\boldsymbol{I} - \boldsymbol{J}$ is invertible (eigenvalues $1 - \mu_i > 0$); the Neumann series $\sum_{j\geq 0}\boldsymbol{J}^j = (\boldsymbol{I} - \boldsymbol{J})^{-1}$ converges. Unrolling equation 13 with $\boldsymbol{f}_k \equiv \boldsymbol{f}$ from $\boldsymbol{a}_0$,

$$\boldsymbol{a}_k = \boldsymbol{J}^k\boldsymbol{a}_0 + \sum_{j=0}^{k-1}\boldsymbol{J}^j\boldsymbol{f} \;\xrightarrow[k\to\infty]{}\; (\boldsymbol{I} - \boldsymbol{J})^{-1}\boldsymbol{f}, \tag{15}$$

the transient $\boldsymbol{J}^k\boldsymbol{a}_0$ vanishing at rate μ_1^k. Equivalently $\boldsymbol{a}_\infty$ solves $(\boldsymbol{I} - \boldsymbol{J})\boldsymbol{a}_\infty = \boldsymbol{f}$; projecting on the unit eigenvector $\boldsymbol{v}_1$, $(1 - \mu_1)a_\infty^{(1)} = f^{(1)}$, so $|a_\infty^{(1)}| = |f^{(1)}|/(1 - \mu_1)$. The map $\rho \mapsto |f^{(1)}|/\rho$ is decreasing on $(0, 1)$, so a larger ρ shrinks the leading-mode offset.

At the fixed point, $\boldsymbol{a}_\infty = \boldsymbol{J}\boldsymbol{a}_\infty + \boldsymbol{f}$. Subtracting this relation from equation 13 and iterating gives $\boldsymbol{a}_k - \boldsymbol{a}_\infty = \boldsymbol{J}^k(\boldsymbol{a}_0 - \boldsymbol{a}_\infty)$. Since $\boldsymbol{a}_k = \boldsymbol{u}_k - \boldsymbol{u}^\star$ and $\boldsymbol{a}_\infty = \boldsymbol{u}_\infty - \boldsymbol{u}^\star$, this is $\boldsymbol{u}_k - \boldsymbol{u}_\infty = \boldsymbol{J}^k(\boldsymbol{u}_0 - \boldsymbol{u}_\infty)$. Expanding in the orthonormal eigenbasis gives equation 8. If $\langle \boldsymbol{u}_0 - \boldsymbol{u}_\infty, \boldsymbol{v}_1 \rangle \neq 0$, each other eigencomponent decays relative to the leading component at rate $(\mu_i/\mu_1)^k$, $i \geq 2$. The strict spectral gap in Assumption 1 makes these ratios vanish, proving asymptotically one-dimensional convergence along $\boldsymbol{v}_1$. □

Nonlinear local existence and uniqueness. The fixed-point expression in Lemma 3 is exact for the linearized loop. For the nonlinear map, the implicit function theorem gives a nearby fixed point. Write its condition as $\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{f}) = \boldsymbol{s}(\boldsymbol{u}) - \boldsymbol{u} + \boldsymbol{f} = \boldsymbol{0}$; it is C^1 near $(\boldsymbol{u}^\star, \boldsymbol{0})$ (Assumption 1) with $\boldsymbol{G}(\boldsymbol{u}^\star, \boldsymbol{0}) = \boldsymbol{0}$, and its $\boldsymbol{u}$-Jacobian there is $\partial_{\boldsymbol{u}}\boldsymbol{G} = \boldsymbol{J} - \boldsymbol{I}$, invertible whenever $\mu_1 \neq 1$. The contraction $\mu_1 < 1$ supplies this, so the theorem gives a neighborhood of $\boldsymbol{f}{=}\boldsymbol{0}$ on which the driven fixed point $\boldsymbol{u}_\infty(\boldsymbol{f})$ exists, is locally unique, and is C^1 in $\boldsymbol{f}$, with first-order expansion $\boldsymbol{u}_\infty(\boldsymbol{f}) = \boldsymbol{u}^\star + (\boldsymbol{I} - \boldsymbol{J})^{-1}\boldsymbol{f} + o(\|\boldsymbol{f}\|)$: the linear formula equation 14 is its leading term. The theorem upgrades the linear solution to a locally unique branch of the nonlinear loop; it does not by itself produce $\boldsymbol{u}^\star$, whose existence is Assumption 1. The empirical check of the near-fixed point is given in App. B.3.

Local frozen-coefficient approximation. Lemma 3 freezes $\boldsymbol{J}$ and $\boldsymbol{f}$ (the standard quasi-static linearization). When $\boldsymbol{J}$ and $\boldsymbol{f}$ vary sufficiently slowly, the local model describes tracking of an instantaneous driven fixed point rather than convergence to a stationary one.

B.2 Difference-of-means estimation of the repetition direction

The local model identifies an idealized dominant eigenvector $\boldsymbol{v}_1$, but this vector is not available during generation. We therefore estimate an operational direction from observed self-conditioning feedback: trajectories with more repetition should lie farther along the attractor than trajectories with less repetition. The following definition makes this contrast precise and specifies how the resulting direction is used for steering.

Definition 1 (Difference-of-means direction and steering)**.** *Let $a^{(1)} = \langle \boldsymbol{a}, \boldsymbol{v}_1 \rangle$ be the signed repetition-axis coordinate, oriented so $+\boldsymbol{v}_1$ points toward the repetition fixed point $\boldsymbol{u}^\star$; locally, larger $a^{(1)}$ is associated with higher REP-4. The trapped and free groups $\mathcal{T}, \mathcal{F}$ contain the top and bottom q fractions of calibration trajectories ranked by REP-4, respectively, where $q \in (0, 1/2]$ is the fraction assigned to each tail. They are separated along signed $a^{(1)}$ with $\bar{a}^{(1)}_{\mathcal{T}} > \bar{a}^{(1)}_{\mathcal{F}}$, so the difference-of-means gap $\bar{a}^{(1)}_{\mathcal{T}} - \bar{a}^{(1)}_{\mathcal{F}} \neq 0$. The estimator*

$$\boldsymbol{d} \;=\; \frac{\boldsymbol{m}_{\mathcal{T}} - \boldsymbol{m}_{\mathcal{F}}}{\|\boldsymbol{m}_{\mathcal{T}} - \boldsymbol{m}_{\mathcal{F}}\|}, \qquad \boldsymbol{m}_{\mathcal{T}} = \mathbb{E}[\boldsymbol{u} \mid \mathcal{T}], \quad \boldsymbol{m}_{\mathcal{F}} = \mathbb{E}[\boldsymbol{u} \mid \mathcal{F}], \tag{16}$$

defines the attractor direction, and steered self-conditioning feeds back $\tilde{\boldsymbol{x}}_k = \boldsymbol{x}_k - \lambda \boldsymbol{1}_L \boldsymbol{d}^\top$ with strength $\lambda \geq 0$.

Proposition 1 (Difference of means recovers the attractor direction)**.** *Split the self-conditioning feedback into the repetition axis and the rest, $\boldsymbol{u} = \boldsymbol{u}^\star + a^{(1)}\boldsymbol{v}_1 + \boldsymbol{w}$, where $\boldsymbol{w} \perp \boldsymbol{v}_1$ is the orthogonal component. Assume* separability*: the trapped/free split acts only through the on-axis coordinate $a^{(1)}$, so the orthogonal component $\boldsymbol{w}$ has a group-independent mean, $\mathbb{E}[\boldsymbol{w} \mid \mathcal{T}] = \mathbb{E}[\boldsymbol{w} \mid \mathcal{F}]$. Then the difference of means is parallel to the repetition axis, $\boldsymbol{d} \parallel \boldsymbol{v}_1$.*

Proof. Take the group-conditional mean under the model: for $\mathcal{G} \in \{\mathcal{T}, \mathcal{F}\}$,

$$\boldsymbol{m}_{\mathcal{G}} = \mathbb{E}[\boldsymbol{u} \mid \mathcal{G}] = \boldsymbol{u}^\star + \bar{a}^{(1)}_{\mathcal{G}}\, \boldsymbol{v}_1 + \mathbb{E}[\boldsymbol{w} \mid \mathcal{G}], \qquad \bar{a}^{(1)}_{\mathcal{G}} := \mathbb{E}[a^{(1)} \mid \mathcal{G}],$$

the group-mean on-axis coordinate. Subtracting the two groups,

$$\boldsymbol{m}_{\mathcal{T}} - \boldsymbol{m}_{\mathcal{F}} = \underbrace{(\boldsymbol{u}^\star - \boldsymbol{u}^\star)}_{=\boldsymbol{0}} + \big(\bar{a}^{(1)}_{\mathcal{T}} - \bar{a}^{(1)}_{\mathcal{F}}\big)\boldsymbol{v}_1 + \underbrace{\big(\mathbb{E}[\boldsymbol{w} \mid \mathcal{T}] - \mathbb{E}[\boldsymbol{w} \mid \mathcal{F}]\big)}_{=\boldsymbol{0}} : \tag{17}$$

the fixed point $\boldsymbol{u}^\star$ is shared by both groups and cancels, and the orthogonal term $\mathbb{E}[\boldsymbol{w} \mid \mathcal{T}] - \mathbb{E}[\boldsymbol{w} \mid \mathcal{F}]$ vanishes by the modeling assumption. Only the on-axis term survives, so $\boldsymbol{m}_{\mathcal{T}} - \boldsymbol{m}_{\mathcal{F}} = (\bar{a}^{(1)}_{\mathcal{T}} - \bar{a}^{(1)}_{\mathcal{F}})\boldsymbol{v}_1 \parallel \boldsymbol{v}_1$, with scalar $\bar{a}^{(1)}_{\mathcal{T}} - \bar{a}^{(1)}_{\mathcal{F}} \neq 0$ since the trapped group sits deeper along $\boldsymbol{v}_1$ (Definition 1). Normalizing, $\boldsymbol{d} \parallel \boldsymbol{v}_1$ (up to sign). □

On the trained model, the estimated direction only partially aligns with the leading mode (Tab. 4).

Table 4: **A dominant mode emerges as repetition settles into an attractor.** Jacobian diagnostics across denoising trajectory fractions.

Denoising fraction	REP-4 mean (%)	Spectral gap μ_1/μ_2	$\lvert\cos(\boldsymbol{v}_1, \boldsymbol{d})\rvert$	Amp. ($\boldsymbol{d}$/random)
0.40	3.2	1.00	0.51	0.87
0.50	7.5	1.06	0.45	1.23
0.60	10.1	1.67	0.56	1.78
0.70	11.3	2.41	0.56	2.20
0.80	11.6	3.04	0.56	2.06
0.90	11.7	2.95	0.54	2.16

B.3 Empirical validation of the attractor model

We now test the assumptions and predictions of the local attractor model on the trained loop. The validation covers three questions: whether the feedback has a locally contracting fixed point, whether one mode becomes dominant as the repetition basin forms, and whether the observed drift direction differs from the self-conditioning response. For the fixed-point and Jacobian diagnostics, trajectory base points are captured with the SDE sampler, while the frozen map and Jacobian are evaluated using the ODE update; the drift-direction comparison below uses separate deterministic ODE runs.

The frozen pooled loop converges to a fixed point. At fraction 0.85, we capture 100 SDE trajectories per seed. For each sample, we hold its latent state and denoising interval fixed, initialize the pooled iterate from its captured pooled feedback, and apply the deterministic pooled ODE map for 200 iterations. Across five seeds, the mean relative residual $\|\boldsymbol{g}(\boldsymbol{u}) - \boldsymbol{u}\|/\|\boldsymbol{u}\|$ decreases from 0.116 to approximately 9×10^{-5} (Fig. 11a), supporting numerical convergence of this reduced feedback map.

A dominant contracting mode emerges as the basin forms. To determine whether the local contraction is concentrated in a single mode, at each trajectory fraction we capture a batch of 100 SDE trajectories and estimate the batch-averaged Jacobian-vector product of the deterministic ODE update at their captured position-wise feedback by central finite differences, perturbing uniformly across positions. We use 30 matrix-free power iterations for the leading mode and 30 deflated iterations for the second (Alg. 2). Table 4 reports the mean REP-4 of the 100 intermediate predictions at each fraction, alongside the corresponding Jacobian diagnostics. The five-seed Jacobian-gain and response-share results are shown in Fig. 11b,c. We repeat the measurement at trajectory fractions from 0.4 through 0.9. The spectral gap μ_1/μ_2 increases from 1.00 at fraction 0.4 to 3.04 at 0.8, before slightly decreasing to 2.95 at 0.9. This indicates that one mode becomes increasingly separated as the trajectory enters the repetition regime (Tab. 4). Across five calibration seeds, trajectory-based probes show leading gains below one and a larger leading-mode response share in the later repetition regime (Fig. 11b,c). These results support the emergence of a dominant contracting mode as the repetition basin forms.

The observed drift direction is distinct from the feedback mode. Across three deterministic ODE seeds, the leading principal direction of the smoothed per-step drift over the final third of denoising, estimated from 200 trajectories per seed, has mean absolute cosine 0.15 with the leading Jacobian mode, estimated from 30 trajectories per seed, and 0.09 with the ACE direction. This empirical drift direction summarizes observed updates; it is not a direct estimate of the theoretical drive $\boldsymbol{f}_k$. At fixed latent state and denoising time, finite differencing the local map cancels its input-independent drive term and isolates the response to the fed-back clean prediction. Thus, ACE directly modifies the self-conditioning input; this local analysis does not imply that the full denoising trajectory is unchanged.

C Training-Time Mitigation of the Repetition Attractor

The inference-time intervention raises a complementary training-time question: can the model be trained to avoid producing self-conditioning feedback toward the repetition attractor? We continue-train ELF-B from the released checkpoint on OpenWebText using AdamW, with weight decay 0,

effective batch size 512, learning rate 10^{-4}, and 128 optimizer steps. The anti-attractor run uses four 32-step blocks: it re-estimates $\boldsymbol{d}$ from $N{=}4096$ Full-SC calibration samples at $q{=}1/4$ between blocks, and initializes each block from the preceding EMA checkpoint. The first block uses a 50-step warmup; subsequent blocks do not repeat it. The penalty weight is $w{=}8$, while the continued-training control uses $w{=}0$. We evaluate the EMA weights by plain unsteered sampling on a held-out pool of 1,024 samples, matched to the Full-SC pool from the main-results run (Table 1). The EMA decay is reduced from the released value 0.9999 to 0.999 so that the moving average can track this short run (Table 5).

Rather than modifying the feedback at inference, the anti-attractor variant trains the model not to produce clean-latent predictions along the repetition direction. Let $\boldsymbol{X}_\theta(\boldsymbol{z}_k, t_k) \in \mathbb{R}^{L\times e}$ denote the model's clean-latent prediction at noise level t_k, and let $\boldsymbol{u}_\theta = \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{X}_\theta[l] \in \mathbb{R}^e$ be its position-pooled prediction, corresponding to the pooled self-conditioning feedback $\boldsymbol{u}$ in §4 and equation 6. Using the frozen difference-of-means direction $\boldsymbol{d}$ from equation 16, we add the following regularizer to the standard flow-matching loss:

$$\mathcal{L}_{\text{attr}}(\theta) \;=\; \mathbb{E}_{\boldsymbol{z}_0,\,k}\Big[\,\mathrm{ReLU}\big(\langle \boldsymbol{u}_\theta,\, \boldsymbol{d}\rangle\big)^2\,\Big], \tag{18}$$

and we continue-train against

$$\mathcal{L}(\theta) \;=\; \mathcal{L}_{\text{base}}(\theta) \;+\; w\,\mathcal{L}_{\text{attr}}(\theta), \qquad w \geq 0, \tag{19}$$

where $\mathcal{L}_{\text{base}}$ is the standard flow-matching loss. The one-sided ReLU targets only the $+\boldsymbol{d}$ half-line associated with the repetitive basin (Prop. 1), leaving the $-\boldsymbol{d}$ side and all components orthogonal to $\boldsymbol{d}$, denoted by $\boldsymbol{w} \perp \boldsymbol{d}$, unchanged. Because the regularizer acts on $\boldsymbol{X}_\theta$ itself, the resulting model requires no test-time intervention.

At 128 steps, the anti-attractor variant lowers both REP-4 mean and repetition share relative to the released Full-SC checkpoint and continued-training control (Table 5). The continued-training control changes REP-4 only marginally relative to Full-SC. This result is consistent with $\boldsymbol{d}$ reflecting a trainable property of the self-conditioning dynamics, rather than merely an inference-time handle.

D Alternative Self-Conditioning Interventions for Repetition

We detail the inference-time alternatives evaluated in §5.2. They operate without retraining by modifying the clean-embedding prediction before it re-enters the self-conditioning loop. For ELF-B, under the common 64-step SDE setting ($\sigma{=}2.0$, $\gamma{=}1.0$, and SC-CFG $w{=}3$), we generate 1,024 samples directly for each setting using the same held-out generation seed, without rejection sampling, and report REP-4, repetition share, Grammar, distinct-$1/3$, self-BLEU, and MAUVE. The Full-SC sampler is retained as a reference. The compact repetition–quality comparison is shown in Fig. 5, and the complete grid over tested alternatives and parameter settings is given in Tab. 6.

The alternatives modify the feedback in different ways. *soft-SC mag* feeds back α times the clean-embedding prediction at every denoising step. Here $\alpha{=}1$ recovers Full-SC, while $\alpha{=}0$ gives SC-off by disabling self-conditioning feedback during inference; the full α sweep is in Figure 2. *SC decay* multiplies the feedback by a coefficient that decreases linearly from 1 at the first denoising step to a floor α at the last. *SC cutoff* feeds back $\mathbf{0}$ for all samples after a fraction τ of the denoising steps has elapsed. Finally, *SC early-reset* performs one additional decoder pass on the current clean-embedding prediction after a fraction τ of the denoising steps. It computes the fraction of token-ID 4-gram occurrences that repeat an earlier one in the intermediate decoded sequence and, for samples above threshold θ, feeds back $\mathbf{0}$ for the remaining steps; other samples retain their original feedback.

ACE instead subtracts $\lambda\boldsymbol{d}$ from the clean-embedding prediction (§4.2), while retaining the remaining feedback. The alternatives occupy different repetition–quality trade-offs: broad feedback suppression can reduce repetition sharply but may lower Grammar, while less disruptive interventions can retain more repetition. ACE is a simple, theory-motivated option that targets the estimated repetition direction without adding an auxiliary decoder pass; it is not uniformly best on every reported metric.

E Steering Strength and Direction Analysis

This appendix supplements §5.3 with analyses of ACE's direction and steering strength. Algorithm 1 summarizes the two-stage procedure: estimating the repetition direction from calibration

Table 5: **Penalizing the estimated REP direction during training reduces repetition.** The released ELF-B checkpoint is compared with matched continued-training variants with and without the anti-attractor penalty on repetition.

Variant	REP-4 mean ↓	Rep. share ↓	Grammar ↑	Fluency ↑	dist-1 ↑	dist-3 ↑	s-BLEU ↓	MAUVE ↑
Released Full-SC	6.85	89.36	80.35	1.04	43.46	88.67	29.16	22.50
Training control	6.74	86.91	79.87	1.03	43.48	88.81	28.98	23.65
Anti-attractor	5.42	84.77	80.35	1.02	44.41	90.40	29.59	23.15

Table 6: **Self-conditioning interventions trade repetition against text quality.** Full-SC, ACE across steering strengths, and the evaluated alternatives are compared across repetition and quality metrics.

Method	Setting	REP-4 mean ↓	Rep. share ↓	Grammar ↑	Fluency ↑	dist-1 ↑	dist-3 ↑	s-BLEU ↓	MAUVE ↑
Full-SC	–	7.22	88.67	79.81	1.04	43.26	88.29	29.24	22.21
SC-off	–	0.09	0.00	51.99	1.00	56.58	99.41	23.02	8.17
ACE λ	.25	5.64	82.42	80.58	1.07	45.28	90.40	29.20	23.39
	.50	4.52	75.59	80.82	1.06	46.76	91.93	29.77	19.08
	1	3.11	55.37	81.17	1.09	48.59	93.89	31.27	11.72
	2	1.85	32.62	82.68	1.08	49.79	95.61	35.88	4.89
	4	1.58	23.54	83.27	1.00	47.48	95.39	44.17	1.02
soft-SC mag α	.2	0.42	2.15	59.68	1.00	52.15	98.29	25.84	17.43
	.4	1.55	21.58	67.48	1.03	47.72	95.87	28.43	23.78
	.6	4.62	69.14	72.98	1.03	43.56	91.25	29.30	26.48
	.8	6.76	86.33	77.12	1.05	42.59	88.63	29.33	26.16
SC decay α	0	2.72	39.36	67.09	1.02	47.44	94.31	27.80	23.42
	.25	5.16	71.00	71.85	1.00	43.63	90.67	29.01	24.14
	.5	6.97	84.86	75.20	1.02	42.03	88.33	29.48	26.26
	.75	7.47	87.89	77.85	1.02	42.17	87.81	29.36	26.96
SC cutoff τ	.25	0.09	0.10	51.83	1.00	56.39	99.42	23.06	6.51
	.50	0.66	7.32	56.29	1.01	53.16	98.00	25.02	17.38
	.75	4.89	73.44	72.75	1.03	45.90	91.40	28.98	27.72
SC early-reset (τ, θ)	$(.25, .05)$	2.52	31.54	61.72	1.01	51.81	95.64	24.36	15.16
	$(.50, .02)$	0.97	12.99	58.80	1.02	52.37	97.46	25.39	18.96
	$(.50, .05)$	2.27	39.75	67.23	1.03	49.40	95.24	26.93	24.83
	$(.50, .10)$	4.70	74.02	75.95	1.05	45.71	91.59	28.68	23.64
	$(.75, .05)$	4.94	75.29	73.49	1.04	45.73	91.29	29.05	26.60

trajectories and applying it during generation. We first justify the calibration and steering settings used in the main results, then compare alternative directions and controls, and finally examine the local dynamics underlying the repetition attractor.

Algorithm 1 ACE: estimate and subtract the repetition direction

Require: N Full-SC trajectories summarized by $\{(\boldsymbol{s}_n, r_n)\}_{n=1}^{N}$, where $\boldsymbol{s}_n$ is mean self-conditioning feedback and r_n is sample REP-4; tail fraction $q \in (0, 1/2]$; one-step denoising transition function STEP, denoising schedule $\{t_k\}_{k=1}^{T}$, initial noise scale σ, steering strength λ, and token decoder DEC

Ensure: repetition direction $\boldsymbol{d}$ and one steered sample

Stage 1: Estimate the REP direction from trapped and free trajectories

$\mathcal{T} \leftarrow \{n : r_n$ is in the top q fraction by REP-4$\}$

$\mathcal{F} \leftarrow \{n : r_n$ is in the bottom q fraction by REP-4$\}$

$\boldsymbol{d} \leftarrow \text{normalize}\Big(\frac{1}{|\mathcal{T}|}\sum_{n\in\mathcal{T}} \boldsymbol{s}_n - \frac{1}{|\mathcal{F}|}\sum_{n\in\mathcal{F}} \boldsymbol{s}_n\Big)$

Stage 2: Steer denoising away from repetition

$\boldsymbol{z}_0 \sim \mathcal{N}(0, \sigma^2 I), \quad \tilde{\boldsymbol{x}}_0 \leftarrow \mathbf{0}$

for $k = 1, \ldots, T$ **do**

  $(\boldsymbol{z}_k, \boldsymbol{x}_k) \leftarrow \text{STEP}(\boldsymbol{z}_{k-1}, t_k, \tilde{\boldsymbol{x}}_{k-1})$

  $\tilde{\boldsymbol{x}}_k \leftarrow \boldsymbol{x}_k - \lambda \mathbf{1}_L \boldsymbol{d}^\top$ ▷ broadcast over positions

end for

return $\text{DEC}(\boldsymbol{z}_T)$

E.1 SELECTING ACE'S CALIBRATION AND STEERING SETTINGS

We select the calibration size N from direction stability across sample sizes, the tail fraction q from cross-seed agreement, and the steering strength λ from the held-out repetition–quality trade-off. As N grows from 256 to 4096, the mean adjacent-size absolute cosine across the plotted curves rises from 0.917 to 0.995, indicating that the direction stabilizes at the largest tested size (Fig. 8a). Among $q \in \{1/8, 1/4, 1/2\}$, q=1/4 has the highest cross-seed agreement (Fig. 8b and Fig. 10a). Panel (c) compares the model-specific directions estimated at the selected setting, N=4096 and q=1/4. Their pairwise absolute cosines range from 0.86 to 0.94, showing substantial alignment and supporting the shared-direction results in Table 1.

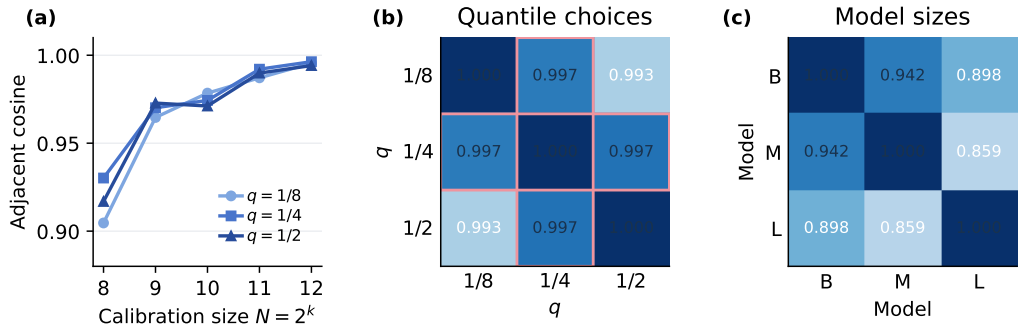


Figure 8: **The estimated repetition direction is stable across calibration choices and model sizes.** (a) Adjacent-size absolute cosine, approaching a stable value at larger N; (b) pairwise cosine among the three repetition-quantile choices at N=4096; and (c) pairwise absolute cosine among the model-local directions $\boldsymbol{d}_B$, $\boldsymbol{d}_M$, and $\boldsymbol{d}_L$, estimated with N=4096 and q=1/4.

At N=4096, the held-out curves show how repetition and quality change with q and steering strength (Fig. 9); the dashed lines mark the Full-SC references. Increasing λ initially reduces repetition while quality remains competitive, whereas stronger steering exposes a repetition–quality trade-off. We therefore use λ=1/2 as the quality-focused setting in the main results. Together, these analyses support the settings used in §5.1: N=4096, q=1/4, and λ=1/2.

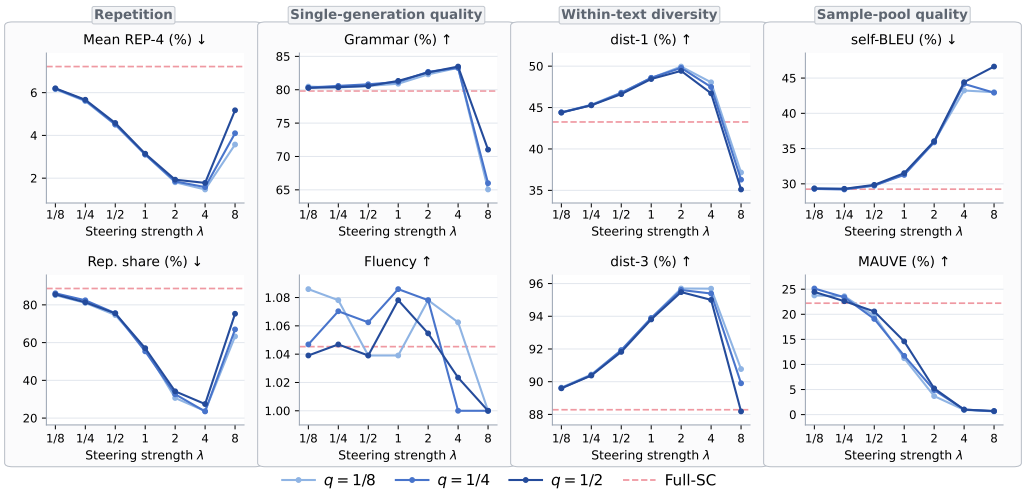


Figure 9: **ACE exhibits a repetition–quality trade-off across steering strengths.** The curves show repetition, single-generation quality, within-text diversity, and sample-pool quality as steering strength λ varies for the tested calibration tail fractions q.

E.2 Comparing direction estimators and controls

We evaluate how held-out steering depends on the direction estimator. ACE subtracts the mean self-conditioning feedback of the bottom-q REP-4 tail from that of the top-q tail, omitting the middle $1 - 2q$. Logistic and LDA fit scikit-learn's default `LogisticRegression` and `LinearDiscriminantAnalysis` models to the same tails after feature standardization; their coefficients are mapped back to the original feature space before normalization. Thus, q denotes the fraction in each tail for ACE, Logistic, and LDA. Top-PC uses the leading right singular direction of uncentered pooled feedback from the top-q trajectories. The Jacobian estimator averages local self-conditioning Jacobian-vector products over the same trajectories and takes the leading mode at $\tau = 7/8$. For both estimators, $q = 1$ includes all calibration trajectories; signs are set by the high-minus-low REP-4 feedback contrast, using $q = 1/2$ for this sign convention when $q = 1$. A random unit vector and an ACE direction estimated after permuting REP-4 labels serve as controls.

Figure 10 shows direction agreement across q in panel (a) and held-out responses across steering strengths in panel (b). Panel (a) tests $q \in \{1/8, 1/4, 1/2\}$ for ACE, Logistic, and LDA, plus $q = 1$ for Top-PC and the Jacobian estimator. Panel (b) uses each estimator's q with the highest mean pairwise cosine across five calibration directions: $1/4$ for ACE, $1/8$ for Logistic, $1/4$ for LDA, and 1 for Top-PC and the Jacobian estimator. Both panels use N=4096 samples and five calibration seeds; panel (b) varies $\lambda \in \{1/4, 1/2, 1, 2, 4\}$, with standard-error bands across the five directions on a shared 1,024-sample held-out pool. Table 7 gives a matched-strength comparison under the 64-step SDE setting (σ=2.0, γ=1.0, w=3), using a separate N=4096 calibration set and a common held-out generation seed with 1,024 direct generations per condition. ACE and the permuted-label control use q=1/4; the random direction does not use q. All steered rows use λ=1/2, with Full-SC as the reference.

Table 7: **Repetition-aligned steering reduces repetition relative to unsteered and placebo controls.** Held-out comparison of Full-SC, ACE, random-direction, and permuted-label conditions.

Method	REP-4 mean ↓	Rep. share ↓	Grammar ↑	Fluency ↑	dist-1 ↑	dist-3 ↑	s-BLEU ↓	MAUVE ↑
Full-SC	7.22	88.67	79.81	1.04	43.26	88.29	29.24	22.21
ACE	4.52	75.59	80.82	1.05	46.76	91.93	29.77	19.08
Random direction	8.62	91.89	79.29	1.07	41.36	86.44	30.10	26.36
Permuted labels	7.82	90.14	76.12	1.02	40.26	87.12	32.14	13.72

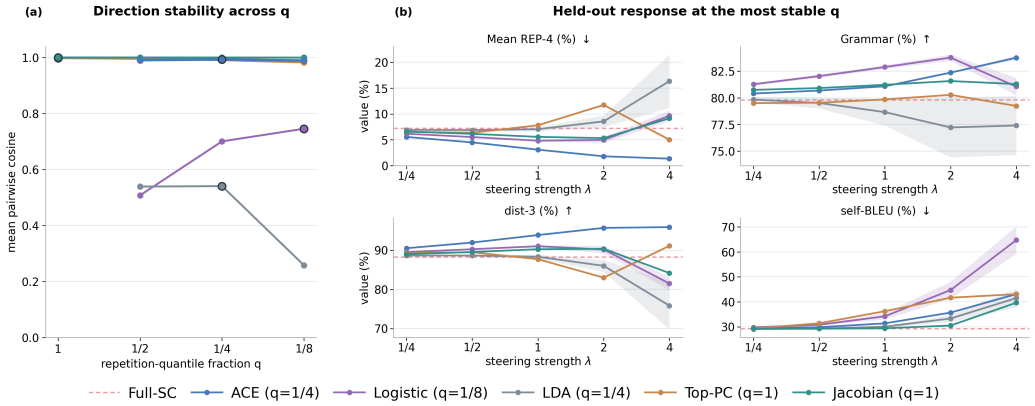


Figure 10: **Calibration stability and held-out responses distinguish repetition-aligned directions.** (a) Mean pairwise cosine across calibration directions as q varies. (b) Held-out repetition and quality as steering strength λ varies at each estimator's selected q.

At this strength, ACE lowers both REP-4 measures relative to Full-SC, whereas neither control does. Across strengths, ACE and Logistic reduce REP-4 at moderate values while maintaining competitive Grammar and distinct-3, although self-BLEU rises with stronger steering. Top-PC also lowers REP-4 at λ=4, but with a marked increase in self-BLEU; LDA and the Jacobian mode show less consistent REP-4 reduction. Together, these results detail the §5.3 finding that theory-motivated estimators can reduce repetition while retaining competitive quality at suitable strengths, whereas theory-agnostic alternatives do not reliably reproduce this joint behavior. ACE provides a simple, stable, and effective implementation of this principle.

E.3 Dynamical validation and interpretation of the repetition direction

The fixed-point and mode-dominance diagnostics are detailed in App. B.3; here we focus on how ACE's estimated direction relates to the local feedback mode. We estimate Jacobian modes from central finite-difference products of the deterministic ODE transition at SDE-captured trajectory points, retaining position-wise feedback. We evaluate fractions $f \in \{0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, average over 100 trajectories per seed and five calibration seeds, and use 30 power iterations for the leading mode and 30 deflated iterations for the second (Alg. 2). The compared ACE direction uses N=4096 and q=1/4; q is not a Jacobian-estimator parameter. By contrast, the fixed-point check applies a frozen pooled map, which broadcasts pooled feedback (App. B.3). We orient $\boldsymbol{v}_1$ so its feedback projection increases with repetition. The finite-iteration gains are approximate mode diagnostics; the orthogonal eigenmode interpretation relies on the symmetric idealization.

As the trajectory enters the repetitive regime, the spectral gap rises from 1.00 at fraction 0.40 to 3.04 at 0.80, then is 2.95 at 0.90 (Tab. 4; Fig. 4b). Across these fractions, the absolute cosine between $\boldsymbol{d}$ and $\boldsymbol{v}_1$ is 0.45–0.56, above the random reference $1/\sqrt{e} = 0.044$ but far from identity; the amplification ratio along $\boldsymbol{d}$ relative to random perturbations reaches 2.20. The held-out curves show less consistent repetition reduction for the Jacobian estimator than for ACE (Fig. 10). Thus, $\boldsymbol{d}$ is partially aligned with, but is not identical to, the leading local mode. This is consistent with ACE capturing a direction that recurs in the evolving feedback dynamics, rather than a single static Jacobian eigenvector.

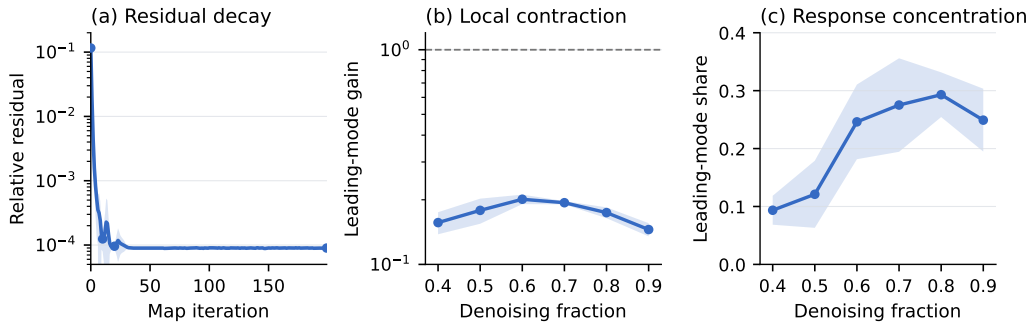


Figure 11: **The frozen pooled map converges, while the trajectory-level response concentrates in one leading mode.** (a) Frozen pooled-map iteration reduces the relative residual. (b) The leading finite-difference gain remains below one. (c) The leading-mode response share increases in the later repetition regime. Shaded bands show standard deviations across five calibration seeds.

Algorithm 2 Attractor modes $\boldsymbol{v}_1, \boldsymbol{v}_2$: matrix-free deflated power iteration on the self-conditioning feedback Jacobian

Require: one-step map $\boldsymbol{g}$ at trajectory fraction f; power-iteration count $K{=}30$; probe scale ε
Ensure: modes $\boldsymbol{v}_1, \boldsymbol{v}_2$ and gains μ_1, μ_2 (spectral gap μ_1/μ_2)

iterate the loop $\boldsymbol{g}$ forward (baseline run) and capture its self-conditioning feedback $\boldsymbol{x}_k$ at step $k{=}\lfloor fT \rfloor$ (near-converged)
$\boldsymbol{u}^\star \leftarrow \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{x}_k[l]$ ▷ captured from the run, not solved
$\varepsilon \leftarrow 0.01\ \mathrm{std}(\boldsymbol{u}^\star)$
$\boldsymbol{J}\boldsymbol{v} := \big[\boldsymbol{g}(\boldsymbol{u}^\star{+}\varepsilon\boldsymbol{v}) - \boldsymbol{g}(\boldsymbol{u}^\star{-}\varepsilon\boldsymbol{v})\big]/(2\varepsilon)$ ▷ drive $\boldsymbol{f}_k$ cancels
$\boldsymbol{v}_1 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, normalized
for $j = 1, \ldots, K$ **do**
    $\boldsymbol{w} \leftarrow \boldsymbol{J}\boldsymbol{v}_1; \quad \mu_1 \leftarrow \|\boldsymbol{w}\|; \quad \boldsymbol{v}_1 \leftarrow \boldsymbol{w}/\mu_1$ ▷ (Golub & Van Loan, 2013)
end for
$\boldsymbol{v}_2 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, orthogonalized against $\boldsymbol{v}_1$ and normalized
for $j = 1, \ldots, K$ **do**
    $\boldsymbol{w} \leftarrow \boldsymbol{J}\boldsymbol{v}_2 - \langle \boldsymbol{J}\boldsymbol{v}_2, \boldsymbol{v}_1\rangle \boldsymbol{v}_1; \quad \mu_2 \leftarrow \|\boldsymbol{w}\|; \quad \boldsymbol{v}_2 \leftarrow \boldsymbol{w}/\mu_2$ ▷ deflated
end for
return $\boldsymbol{v}_1, \boldsymbol{v}_2, \mu_1, \mu_2$

To measure how much of the local feedback response is captured by the leading mode, we draw P random unit perturbations $\{\boldsymbol{\delta}_j\}_{j=1}^{P}$ and evaluate the finite-difference responses $\boldsymbol{J}\boldsymbol{\delta}_j$. Under the local symmetric approximation, we summarize the response energy along $\boldsymbol{v}_1$ by

$$\rho_1 = \frac{\sum_{j=1}^{P} \langle \boldsymbol{J}\boldsymbol{\delta}_j, \boldsymbol{v}_1 \rangle^2}{\sum_{j=1}^{P} \|\boldsymbol{J}\boldsymbol{\delta}_j\|^2}. \tag{20}$$

A value near one would mean that nearly all local response energy lies along $\boldsymbol{v}_1$. The measured share rises late in denoising but remains below one (Fig. 11c), supporting an increasingly concentrated, rather than exactly rank-one, response. The leading gains remain below one, and the separate frozen pooled-map iteration converges numerically (Fig. 11a,b; App. B.3). Together, the two diagnostics support contraction in the local maps and increasing concentration of the trajectory-level response in one mode; they do not establish an exactly one-dimensional system.

The temporal probes provide a complementary view: across samples, projection onto $\boldsymbol{v}_1$ tracks REP-4, while along Full-SC trajectories the projection onto $\boldsymbol{d}$ builds as REP-4 rises; ACE suppresses both the projection and repetition (Fig. 4a,c,d). This scalar projection trend is distinct from the direction of the per-step drift, whose leading principal direction has low mean absolute cosine similarity with both the Jacobian mode and $\boldsymbol{d}$; as discussed in App. B.3, it is not a direct estimate of the theoretical drive. The association suggests that $\boldsymbol{d}$ tracks repetition-related feedback during denoising, though it

does not by itself validate a detector or intervention based on that signal. As a token-level probe, a linear logit lens through the output unembedding ranks subword pieces such as *tri*, *ction*, *iding*, and *lect* among its top-scoring tokens, while whole words such as *emotion*, *scholarship*, and *stretched* rank among its lowest. This pattern is consistent with $\boldsymbol{d}$ encoding repetitive sub-lexical content that steering removes.

F Robustness and Generalization

This appendix supplements §5.4. Within ELF, we test transfer of the ELF-B direction across model sizes and inference configurations, and separately assess robustness to calibration and generation seeds. For these ELF experiments, the canonical setting uses 64 SDE denoising steps, SC-CFG guidance scale w=3, noise scale σ=2.0, and churn γ=1.0; ACE uses λ=$1/2$. Each ELF robustness-table run contains $1{,}024$ direct generations, and we report repetition and text-quality metrics for the Full-SC baseline and ACE with the shared direction. The direction-seed block averages five independently estimated directions on a common held-out pool; size and robustness rows average five held-out seeds, while step and guidance rows use a single held-out seed.

F.1 Robustness across inference configurations

We vary the number of denoising steps, guidance scale, initial noise scale $\sigma \in \{1.5, 2.5\}$, ODE and SDE sampling, SDE churn, and random seed. Figure 7 displays five denoising-step values, $16, 32, 64, 128, 256$, and five SC-CFG guidance scales, $1, 2, 3, 4, 5$; each point reports mean REP-4 from $1{,}024$ samples. The step panel fixes the SC-CFG scale at 3, while the guidance panel fixes the number of denoising steps at 64. Table 8 reports the corresponding quality measures for these settings and for the additional noise, churn, and ODE configurations.

F.2 Transfer across model sizes

The direction estimated on ELF-B also transfers to ELF-M and ELF-L. Table 8 reports the shared direction on five held-out seeds and shows the same reduction in repetition while the quality measures remain competitive. Model-specific direction estimates are reported separately in the main results table.

F.3 Transfer to other self-conditioned models

To test cross-model transfer, we select unconditional continuous-diffusion models that use self-conditioning and have released checkpoints. We exclude models that combine left-to-right block-wise generation with within-block diffusion, such as AURORA-LM (Liang et al., 2026) and ByteDance's CoLA-DLM (Guo et al., 2026). The resulting test models are Plaid (Gulrajani & Hashimoto, 2023) and LangFlow (Chen et al., 2026).

For Plaid, direction calibration uses 1024 Full-SC samples and q=$1/3$; the table reports λ=2. Plaid uses 4096 denoising steps and score temperature 0.9. For LangFlow, we use q=$1/2$ and its 1024-step Euler–EDM sampler. Like ELF, ACE exhibits a repetition–quality trade-off on these external models. Across $\lambda \in \{1/4, 1/2, 1, 2, 4\}$, LangFlow's REP-4 falls as strength increases through λ=2 while single-generation quality remains near Full-SC; at stronger steering, sample-pool metrics worsen and REP-4 rebounds. This parallels the strength-dependent pattern discussed for ELF in §5.3. Table 2 reports λ=$1/2$, matching the main-text reference strength.

Together, these results show that ACE is robust across the tested ELF configurations and model sizes and also reduces repetition in two external self-conditioned continuous diffusion models using model-specific directions.

G Qualitative Examples

We provide two paired ELF-B generations to make the repetition defect and the effect of ACE concrete. Within each pair, Full-SC and ACE use the same inference setting and random seed. For

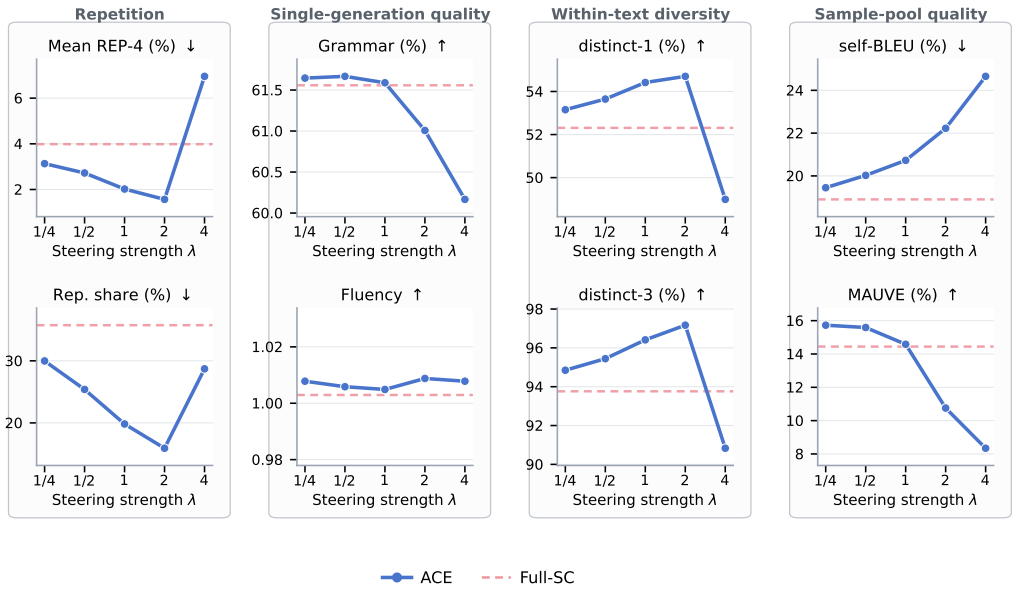


Figure 12: **LangFlow shows a strength-dependent repetition–quality trade-off.** Mean REP-4 and repetition share, Grammar and Fluency, distinct-1/3, and self-BLEU and MAUVE across ACE steering strengths; dashed lines mark Full-SC.

readability, red boxes highlight all occurrences of one repeated 4-gram in each Full-SC output. The examples illustrate the repetition effect; aggregate text quality is evaluated in Table 1.

G.1 Paired generation 1

Full-SC REP-4: 23.3%*; ACE REP-4:* 1.5%*. The highlighted 4-gram, "button on the top", occurs 8 times in Full-SC.*

Full-SC (baseline)

Want more the top college football news you've been watching? Make be sure to follow us on Twitter: www.leadyball.com If you're a beginner, today is the perfect time to create a custom logo for your website. Before you start doing anything else, consider also designing your logo for your website. This post will overview the best ways to create a logo, how to design for your website, and picking the best one you have. tips: 1. Design The Plams Creating a custom logo is a bit of a tedious process, but we can usually do everything we can to provide a design so that a logo can be much more visible and better shared with the public. Photos and videos are essential to creating a logo, so it's always important to keep your website's logo visible as well. Remember to make your logo a single photo. Read a tutorial on how to do this below. 2. Create The Logo The first step of creating the logo, click the "CALL" button on the top of the window and to select the website logo. Click on "Define" to create the new logo. To create the logo, click the "Distribuates" button on the top of the window and select to select the website logo. Click on "Distribuates" to create the new logo. The next step, select to be your favorite type of website and select to be your favorite type of website. To design the new logo, click the "Defate" button on the bottom of the window. After you start creating the logo, you can select the type of websites you would love to use. To create the new logo, click the "Detribuates" button on the top of the window and select the links at the fillin of the next window. Click on logo in the step below. 3. Customate The Logo Once you've created your own logo, you can click the Design button above your window. Now, you can design your own logo, and be sure to select the website that you like the most. Here's where you start to create: 4. Design The Logo First step of creating the logo, you first step by selecting the appropriate type of websites you would love to use. Next step is to select the appropriate type of websites you'd love to use. In this video, be sure to design your logo so that it remains visible. The image below will show you everything you really need to do to design the logo. You can design a custom website logo to make sure it remains very visible on your

Table 8: **ACE transfers across model sizes and inference configurations, remains stable across independent seeds, and maintains competitive text quality.** Full-SC and ACE are compared across model sizes, inference configurations, and independent seeds.

Setting	Method	REP-4 mean ↓	Rep. share ↓	Grammar ↑	Fluency ↑	dist-1 ↑	dist-3 ↑	s-BLEU ↓	MAUVE ↑
Canonical ELF-B configuration (64-step SDE, σ=2, γ=1, w=3)									
Reference	Full-SC	7.28	90.00	80.21	1.04	43.24	88.19	29.03	23.63
	ACE	4.48	74.02	80.57	1.08	46.75	91.93	29.61	19.74
Direction estimation seeds									
5 $\boldsymbol{d}_B$ estimates	Full-SC	7.22	88.67	79.81	1.04	43.26	88.29	29.24	22.21
	ACE	$4.50_{\pm 0.06}$	$74.57_{\pm 0.60}$	$80.69_{\pm 0.06}$	$1.05_{\pm 0.02}$	$46.81_{\pm 0.05}$	$91.94_{\pm 0.07}$	$29.84_{\pm 0.13}$	$19.27_{\pm 0.73}$
Transfer across model sizes									
ELF-M	Full-SC	2.99	52.42	85.97	1.20	47.88	94.04	27.10	25.05
	ACE	1.79	31.84	85.56	1.27	50.82	96.00	27.16	22.56
ELF-L	Full-SC	2.50	43.61	85.83	1.43	50.14	94.94	27.27	23.04
	ACE	1.58	26.60	85.60	1.42	52.44	96.45	27.51	19.81
Denoising steps									
32 steps	Full-SC	2.17	42.58	77.88	1.04	48.03	95.07	28.34	18.73
	ACE	1.33	22.85	77.18	1.03	51.23	96.69	27.50	15.95
128 steps	Full-SC	15.29	99.02	79.71	1.04	38.29	78.94	28.61	25.51
	ACE	10.06	94.73	80.56	1.02	42.46	85.08	29.88	23.46
SC-CFG guidance scale									
$w = 2$	Full-SC	5.01	76.56	79.20	1.04	44.80	91.07	28.74	22.13
	ACE	2.88	55.66	79.14	1.04	48.53	94.22	28.88	21.10
$w = 4$	Full-SC	6.78	89.06	80.81	1.04	43.33	88.64	29.92	23.66
	ACE	4.31	76.56	81.00	1.09	46.80	92.05	30.34	20.93
Additional inference configurations									
$\sigma = 1.5$	Full-SC	15.80	99.45	81.96	1.04	34.19	77.25	40.56	6.92
	ACE	9.99	94.75	82.15	1.03	39.41	84.43	38.17	8.16
$\sigma = 2.5$	Full-SC	8.05	79.20	72.81	1.04	43.95	87.34	22.03	35.07
	ACE	4.36	62.81	75.25	1.09	47.64	92.12	24.93	28.96
$\gamma = 0.5$	Full-SC	6.34	83.46	77.16	1.04	45.11	89.60	26.34	26.49
	ACE	3.85	65.43	77.75	1.06	48.23	92.94	27.30	21.83
$\gamma = 1.5$	Full-SC	8.57	93.57	82.40	1.04	41.21	86.34	31.88	18.15
	ACE	5.33	81.15	82.49	1.07	45.17	90.65	31.99	16.15
ODE sampler	Full-SC	5.55	76.58	72.50	1.04	47.01	90.78	23.74	29.46
	ACE	3.34	58.01	73.72	1.05	49.85	93.75	24.93	25.56

website. The first step of creating the logo, click the “Distribuates” button to select to be your favorite type of website. Click the “Define” button on the top of the window to select the your favorite type of website and then click on logo to select the website. Click the “Detribuates” button on the top of the window Your next step is to select the type of websites you want to use. Click the “Distribuates” button to select the type of websites. Click the “Defate” button on the top of the window to select your favorite type of website. 4. Customate The Website Logo The next step involves selecting the type of websites you want to use. To start creating the logo, click the “Conturn” button on the bottom of the window to select the favorite website logo you like. Your next step is to select the type of websites you would love to use on your website. After you start creating the logo, you can select the appropriate websites you would love to use and click the Printdown button in the next window. 5. Design Your Own Delete The Website Logo The first step of creating the logo, you can custom a new logo. Click the “Delete” button on the top of the window to select your favorite website logo and the “Printdown” button to select to remove the website that you like. After you create the logo, you can select a logo that you like. In order to select your favorite website logo to change, click the Printdown button in the next window. To select your website logo to change into a new design, click the “Delete” button on the top of the window to select your favorite website logo. Click on logo in step below. 6. Customate The Website Logo The next step involves selecting the favorite website you would love to use.

Click the “Tain” button to select your favorite website logo. Click the “Define” button to select your favorite website logo. To select the logo you like, click the “ Like” button above your window to learn about your favorite website logo. Click the “Re Remove” button to select the most popular website you love. Click the “Delete” button to select your favorite website logo. Click “Conturn” button to select the most popular website you love. Select a website logo you like.

ACE

LONDIN (Reuters) - Thousands of people rallied in a building in Bench City, the financial sector capital of the Western capital, on Tuesday for a march on White House Day, as many of their opposition supporters battle over their government. Supporters of Syrian President Bashar al-Assad rally on the streets of Bench City during a demonstration against President Assad in the West African capital, July 19, 2015. REUTERS/Carel Jadder Al-Assad’s ouster of Syrian president on Tuesday sparked a smearing of opposition by U.S. forces in Syria over the Turkish government, something that has ringed in the country for more than two decades. Opponent had said that Assad’s ouster hadposed a direct threat to the U.S. security system. “The elections are taken very seriously.” Interior Minister John Haden said a defender could not provide further details about the event. The march also became the largest opposition march in Damascus to be carried out by more than 700 people. Nearly 32,000 rallied outside the Sidran conference centre, which houses more than 1,100 people. Trade unions, pro-government opposition groups, workers’ rights, and business and television networks were also participants. Nearly 2,000 people had attended the event of the opposition march. Among the participants, Rafel Khajdi, a member of the National Security Council, a right-wing organization in Washington and Moscow, was among those who attended the march. Seri Belocki, a former U.S. diplomat who backed Russia’s Republican presidential nominee Donald Trump, was among those who attended the march. Alandzi Dakhev, a former member of the security council under former President George W. Bush also attended the march. Idel Sidalah, a 22-year-old online journalist with a PhD in politics, had also participated in the march; he was not the only people he said attending the march. The event was the latest in a string of violent and intense marchs in which more than 2,000 people, members of Congress and government officials from over 100 countries joined in battles for change over the ruling Assad government, which is still rifling fighting. Business leaders rallied back into Tuesday afternoon to voice their words outside the headquarters of an oppositionist group run by the Alliance for Human Rights in Benghazi, a south-west part of the country. T clashes were intensed after a policeman was shot dead in a airstrike on government forces in Bench City on Tuesday morning. Simael Nashloud, an Cairo-based online journalist, said: “It’s scary.” Almost everyone among thousands of opposition and opposition protesters gathered in front of a building in business centre in Bench City on Tuesday. They chantted rings of shakers, machine-guns, and drones, latched off gas and gas tunnels, cast off rail tunnels, and threw mini-packs of iron. “Thirety people were injured and many others were still breathing,” said Deni Mosar, a Sunni opposition journalist, who lives on Cairo’s 3k border and attended the march. “This was a very dangerous march.” Many Sunnis have flew to Syria in opposition to opposition legislation introduced in July calling for the Syrian president Bashar al-Assad’s resignation and a referendum to remove the Syrian government. Hundreds of thousands of opposition protesters have turned up at the march on social media, according to the movement. Yar al-Malbar, a 27-year-old who works for a low-level security force who was recently killed in talks with government forces, also participated in the march. Executivees of the Armed Syria Security Service, whose intelligence services work against Assad flew from Syria, Iraq and northeastern Syria. “If we were just going to get away, I think we would migrate to a very safe place,” said Mohammed al-Jarian, a 23-year-old Sunni journalist from Egypt who was also participating in the march. “Our mission is for those who are victims, just not innocents.” Hundreds of registered opposition protesters held a chant of support for Assad’s Syrian government, many of them chanting “Thanks to God”. Mohammed al-Arraq Mahyef, a 27-year-old journalist from Benghazi, was among the first to attend the event of the opposition march. Ali Aster, a 27-year-old journalist from Egypt, also attended a march in Egypt.

G.2 PAIRED GENERATION 2

Full-SC REP-4: 13.3%*; ACE REP-4:* 0.8%*. The highlighted 4-gram, “sit in a row”, occurs 10 times in Full-SC.*

Full-SC (baseline)

By the time I was about 10, I was about to get into rain and I didn’t know how I would get better. When I was about 10, I was angry. I wanted to stand up to what was going on. I knew how to stay down. Whenever the rain was over, I had to clean myself up. I wanted to sit in a row, watch new movies and watch television. Once my mom was ill, she had to place her finger on the back of her head. I wanted to read the book, watch new movies, read history and watch television. I wanted to write the book about Professor Robert

Goldberg, the Editor of the Guardian. I would ask my mom if she could get a job back. I would ask for my friend. I wanted to sit in a row. I always thought I had to ask myself, but I couldn't sit if I wanted to write something about what I was talking about. I wanted to write the book about the Guardian. I had to sit in a row. I had to sit in a row. I wanted to be a journalist. I wanted to be the Editor of the Guardian. I had to sit in a row. I had to have noses and throats and all kinds of bruises. I had to control myself out. Whenever the rain was over, I had to control my mind, my body, my heart. I had some of the worst experiences I ever had to have. I had to sit in a row. I would have to be sitting on a friend's table or on my bed. It would blow me. I wanted to be a journalist but I never wanted to tell new stories. It was scary. I had no idea of where I was. I felt it was my way of life. I wasn't able to feel I could get a job from family. I would check my kids out. I would feel tired. I would re-entry myself. Then I had to sit in a row. I would go home, I would open doors, I would pick my job. I had to have my head hammed against the wall and my bed. I didn't have to, I had to sit in a row. I thought I could do anything I possibly could. I was raped, beaten, savaged, thrown hard by the legs I had on. Then I walked away. I thought I should go to school with friends. I thought a man should just get out on a wall and let rain come out. I thought I should get out on a wall and let no rain come out. I wanted to sit in a row and let no rain come out. He gave me a grab and said, "I'm here. Come back to my house tonight." I thought I should go to school with friends. That thinking put me in the wrong place. My mind blew up. Once I was sick, I couldn't go to school. Whenever the rain was coming over, I went outside and let no rain come out. I wanted to sit in a row and let no rain come out. I went to my mom's apartment and she yelled and shouted and said, "Come on if you don't want to." I looked at her and she shouted, shouted, put gunshots on my face. I went out of the apartment and fleeing. I drew into the apartment where I lived. I had to go outside leaving bullets and gunshots pinning on the door. I had to open my door, let no rain come out, then let no rain come out. I ran out of the apartment again and felt infuriated, fearless. It couldn't go right. I dived. I thought something was going to go wrong. I thought something was going wrong. I didn't know. Even now, I just thought I had to go to school, open my door, pick up a job. I didn't seem to find a place to go home. I felt disappointed and outraged. Despite all this, I felt locked, I felt locked, and I finally left home for the last moment. I haven't left since, because it makes me feel like I'm alive again. All I'm going to do is be away again. What hurts me is that I won't be home again. Advertisement Advertisement Share in this link: Tumblr Email Facebook Twitter Google Print Reddit Tumblr The Green Party is in talks through the general election season, and Tony Abbott's government has had to rescind the formal membership of the National Assembly on the eve of the general election.

ACE

My dad died when I was 5. We were enormously different friends, but we were all completely different. We lived in different families since I was about 7. To my dad, he was very unmarry. Then he got very excited. He was very excited when he realized it didn't matter that much. He loved to sit in the coffee table, listen to magazines and watch television. When he was two years younger, he started making drawings when he was ten. He fled to France from France to play video games or study history at Columbia University. He recently started making drawings for a small coffee shop in Britton, MI. He gave me the hope that he could find a job again. It ended up being my friend's last day when he was 25. I had thought about my dad for years, but he was very scared. When he started working too hard, he was scared. We went to work together for five years, at a coffee shop in East Hill Drive. My dad sat behind the table and I heard my dad yell. I really wanted to help him, but he wasn't on my side. He was fantastic. He had a bright face and a very bright eyeout. I wanted to help him win the job. Then he died at age 25. My dad took his entire life very seriously. It was one of the best memories I had ever had. I'd never met a guy like myself who would have seen my dad walking around a friend's table or watched my dad throwing an explosive with a finger. My dad was so lucky that I hadn't seen anything on public television or on social media since the day I was 11. It felt like my dad had a great family. I didn't know anyone. We had a completely different family. My dad and I met for five years. I would go to school for two or three years, I thought. Then we would hang out at home only once a week. Then I would spend the night drawing and making my own drawings. I took a graduate law class when I was five and studied law when I did. I went to an graduate law class as part of my career at a law school. My dad also died when I was 10. I remember a jack slurning on the ground by the wheel of a bus when my dad grew up. He and I were having so much fun together when we met a man found guilty of robbery in his apartment several blocks away. I saw a man cropping the wall of my dad's apartment as we tried to support each other. His voice made me even louder. He gave me a sore ovation to my dad, and he told me that he was about to tell his story about his dad. That story connected me to a story that my dad grew up with until I was 25. I and my dad were walking down the road near East Hillfield Drive on a recent weekend when a person

was allegedly kidnaped and assaulted at a coffee shop. As I drove into the kitchen at the local coffee shop, someone kidnapped a man with his arm turned on and shouting, "Fuck free!" I called on phone and heard a young woman bragling on the abtainee's face. I and I knew my dad had assaulted. We drove down the road into an unlucky shop a few few blocks away. Shortly after a bomb pounded through the door, I had to open it to force a fire. It was fired and just started getting burned. I went back to the local shop and felt saddened and desperated. I couldn't believe what happened that night. I couldn't believe the alleged abtainee was the alleged attacker. I couldn't believe what happened next. I couldn't believe what happened next. David Zimmerman / Getty Images The British photographer and author David Zimmerman - the permanent owner of Getty Images has his latest narrar, 'My Tales About the Cold Wars,', a six-month-long series about a US military veteran that tells the story of a dissatisfied veteran who obsesss his own life while confronting with a fearless past that has transformed into a fearless past. The first issue is published in the curse of a new book, 'Waking the Top' - an eight-part series about the Cold Wars. The Obama administration and the FBI have helped investigate the assassing of Eric Lavarte on the night of August 9th.

H Related Work

Self-conditioning in continuous diffusion language models. Self-conditioning reuses previous clean predictions for iterative refinement (Chen et al., 2022; Strudel et al., 2022). Continuous text diffusion spans embedding-space models such as Diffusion-LM (Li et al., 2022), Plaid (Gulrajani & Hashimoto, 2023), LangFlow (Chen et al., 2026), and ELF (Hu et al., 2026), as well as latent-space approaches such as CoLA-DLM (Guo et al., 2026) and AURORA-LM (Liang et al., 2026). Among these, Plaid, LangFlow, ELF, and AURORA-LM use explicit self-conditioning, feeding previous clean predictions back during denoising. AURORA-LM generates latent blocks from left to right while denoising positions within each block. Other continuous formulations include statistical-manifold diffusion (Jo & Hwang, 2025) and continuous diffusion with an autoregressive decoder (Shen et al., 2026). Recent analyses examine how self-conditioning shapes denoising: Fixed-point Flows (Yoo et al., 2026) interpret it as fixed-point iteration and exploit it for few-step distillation, while Untied Self-Conditioning (Li & Xu, 2026) addresses a training–inference mismatch by dampening redundant components and correcting solver updates without retraining. We instead study how the self-conditioning loop relates to repetition during unconditional generation.

Repetition in language generation. Autoregressive repetition has been linked to likelihood-maximizing decoding (Holtzman et al., 2020) and self-reinforcement from repeated context (Xu et al., 2022). Remedies include unlikelihood training, which penalizes repetitive outputs (Welleck et al., 2020), and contrastive search, which balances model confidence with a representation-based degeneration penalty (Su et al., 2022). Evaluation studies also show that low Gen-PPL need not indicate high generation quality (Wang et al., 2022; Pynadath et al., 2026; Franca & Tong, 2026). Our study instead examines how feeding clean-embedding predictions back across denoising steps produces repetition, in contrast to self-reinforcement through an autoregressive prefix.

Inference-time steering. Contrastive Activation Addition estimates steering vectors from average differences in residual-stream activations between positive and negative behavioral examples (Rimsky et al., 2024). Activation steering has also been extended to masked DLMs, where residual-stream directions are applied during denoising to control model behavior (Shnaidman et al., 2025). Motivated by our repetition-attractor analysis, ACE uses a related contrastive estimator but intervenes on self-conditioning predictions rather than residual-stream activations.