Do Heavy Tails Help Diffusion?

On the Subtle Trade-off Between Initialization and Training

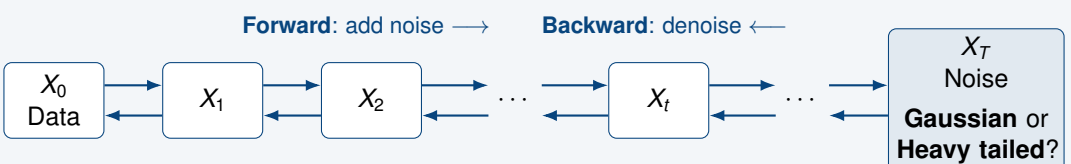
Hamza Cherkaoui, Hélène Halconruy, Antonio Ocello

GENERATING RARE EXTREMES: WHY CHANGING THE NOISE?

The challenge: realistic extremes from limited data

Climate and risk applications need generative models of precipitation and wildfire scenarios.

Diffusion models: add noise, then learn to reverse it



Does better tail alignment guarantee better generation?

FASTER INITIALIZATION, HARDER REVERSE LEARNING

VE-GD: Gaussian noise

Under polynomial tails and smoothness,

$$\mathrm{TV}\Big(\pi_{\mathrm{data}}, \mathcal{L}(\widehat{Y}_{T-t_0})\Big) \lesssim \underbrace{t_0^a + \mathrm{polylog}(n) n^{-c} t_0^{-b}}_{\text{early-stopping bias + estimation}} + \underbrace{T^{-1/2}}_{\text{initialization}} .$$

DLPM: α-stable noise

Under reverse-object regularity and exponential terminal SNR decay, with probability $\geq 1-\delta$,

$$\mathbb{E}_A\Big[\mathrm{TV}\Big(k_{0|A}^{(\alpha)}, \overleftarrow{q}_{0|A}^{\widehat{\theta}_n}\Big)\Big] \lesssim \underbrace{T m^{-\beta(\alpha)/d} + T\sqrt{\frac{\mathrm{Comp} + \log(1/\delta)}{n}}}_{\text{approximation + estimation over the horizon}} + \underbrace{e^{-c_{\mathrm{mix}} T}}_{\text{initialization}} .$$

n: sample size; m: model size; T: horizon; t_0: early stopping.
$a, b, c > 0$ depend on tail index, smoothness and dimension; Comp measures model complexity.

The bounds expose a trade-off: initialization gains can be offset by learning error.

BENCHMARK: MEASURE BOTH FIT AND COVERAGE

Targets

- Isotropic stable law and imbalanced stable mixture ($d = 150$).
- HRRR hourly precipitation: East and West ($d = 150$).
- US wildfire size ($d = 1$).
- CIFAR-100-LT: rare-class coverage.

Models and protocol

Gaussian: DDPM-V, DDIM-V, linear Gaussian flow.

Heavy-tailed: DLPM (α-stable), t-EDM (Student-t).

20–25k training / 1k validation samples for the non-image targets; validation-selected models.

Complementary diagnostics

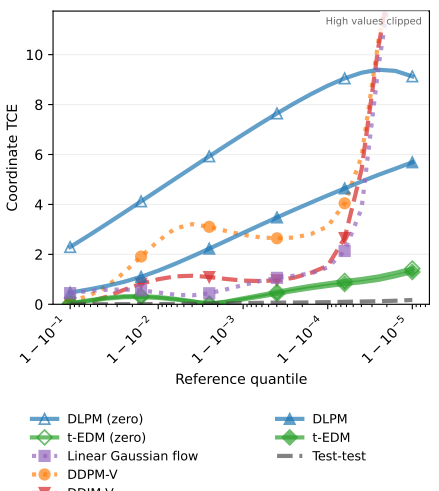
MMD^2 $\downarrow$: global distributional fit. **TCE $\downarrow$:** tail coverage error.

$$\mathrm{TCE}(q) = \frac{1}{d}\sum_{j=1}^{d}\left|\log\frac{r_{\mathrm{gen},j}(q)+\varepsilon}{r_{\mathrm{ref},j}(q)+\varepsilon}\right|.$$

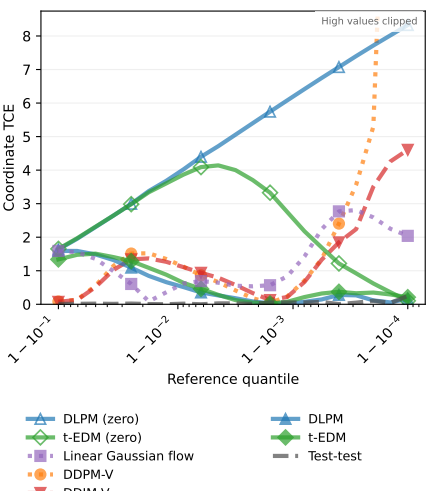
$r_{s,j}(q)$: exceedance rate above the reference q-quantile in coordinate j
Open markers / unfilled bars: zero-network controls; test–test: reference baseline.

GOOD TAIL COVERAGE CAN COME FROM SAMPLING DYNAMICS

Isotropic α-stable ($d = 150$)



US wildfire size ($d = 1$)



Remark. • On a challenging task, DLPM shows limited learning, while t-EDM shows none.
- Light-tailed models remain competitive, even in the far tails.
- On a simpler task, both DLPM and t-EDM show clear learning effects.

TAKE-HOME MESSAGE

Better tail alignment at initialization
does not guarantee better generation.

- t-EDM and DLPM require statistically less demanding regimes to learn the reverse dynamics effectively.
- Gaussian baselines remain competitive, often outperforming in both bulk fidelity and tail recovery.

Join us at our poster!

NeurIPS 2026 — Paris Poster Session 4

Paris Poster Hall | Thursday, Dec. 10, 17:30–19:30 CET