# A. Appendix

## A.1. Proof of Proposition 1

Consider the (per-dimension) linear Flow-Matching trajectory

$$X(t) = (1-t)\,x_0 + t\,\varepsilon, \quad t \in [0,1],$$

with $x_0 \in \mathbb{R}^d$ and a fixed noise realization $\varepsilon \sim \mathcal{N}(0, I_d)$. Given $t_s < t_e$, write $X(t_s) = (1-t_s)x_0 + t_s\varepsilon$ and $X(t_e) = (1-t_e)x_0 + t_e\varepsilon$. We compare two linear re-noise updates:

$$\widehat{X}_e = a\,X(t_s) + b\,\varepsilon \quad \text{(unified/shared noise)},$$
$$\widetilde{X}_e = a\,X(t_s) + b\,\varepsilon' \quad \text{(independent-stage noise)},$$

where $\varepsilon' \sim \mathcal{N}(0, I_d)$ and $\varepsilon' \perp \varepsilon$. Coefficients $a, b \in \mathbb{R}$ may depend on $(t_s, t_e)$ but not on the noise.

**Theorem 1 (Unified re-noise equals a time reparameterization).** There exist

$$a^\star = \frac{1-t_e}{1-t_s}, \qquad b^\star = t_e - a^\star t_s$$

such that $\widehat{X}_e = a^\star X(t_s) + b^\star\varepsilon = X(t_e)$ and hence $\mathbb{E}\Big[\|\widehat{X}_e - X(t_e)\|^2\Big] = 0$.
Define

$$A := a(1-t_s) - (1-t_e), \qquad B := at_s + b - t_e.$$

Then

$$\Delta_{\text{uni}} := \widehat{X}_e - X(t_e) = A\,x_0 + B\,\varepsilon.$$

Choosing $(a^\star, b^\star)$ as above gives $A = B = 0$, hence $\Delta_{\text{uni}} = 0$ and the claim follows.

**Theorem 2 (Independent-stage re-noise incurs irreducible deviation).** For any $a, b \in \mathbb{R}$,

$$\mathbb{E}\Big[\|\widetilde{X}_e - X(t_e)\|^2\Big] \;\geq\; b^2 d.$$

Let

$$A := a(1-t_s) - (1-t_e), \qquad B' := at_s - t_e.$$

Then

$$\Delta_{\text{ind}} := \widetilde{X}_e - X(t_e) = A\,x_0 + B'\varepsilon + b\,\varepsilon'.$$

Conditioning on $(x_0, \varepsilon)$ and using independence/zero-mean of $\varepsilon'$,

$$\mathbb{E}\big[\|\Delta_{\text{ind}}\|^2 \mid x_0, \varepsilon\big] = \|Ax_0 + B'\varepsilon\|^2 + b^2\,\mathbb{E}\|\varepsilon'\|^2 \;\geq\; b^2 d,$$

since $\mathbb{E}\|\varepsilon'\|^2 = d$. Taking expectation in $(x_0, \varepsilon)$ yields the result.

**Corollary 1 (Unified re-noise strictly dominates independent re-noise).** Let $\widehat{X}_e$ and $\widetilde{X}_e$ be defined as above. Then

$$\mathbb{E}\Big[\|\widehat{X}_e - X(t_e)\|^2\Big] \;\leq\; \mathbb{E}\Big[\|\widetilde{X}_e - X(t_e)\|^2\Big],$$

with equality only in the trivial case $b = 0$ and the deterministic term vanishing.

By Theorem 1, $\min_{a,b}\mathbb{E}\|\widehat{X}_e - X(t_e)\|^2 = 0$. By Theorem 2, for any $a, b$, $\mathbb{E}\|\widetilde{X}_e - X(t_e)\|^2 \geq b^2 d \geq 0$. Hence the stated ordering holds.

**Corollary 2 (Multi-stage lower bound).** For multiple stage transitions with coefficients $\{b_i\}_{i=1}^m$ in the independent scheme,

$$\mathbb{E}\Big[\big\|\text{path}_{\text{ind}} - \text{path}_{\text{FM}}\big\|^2\Big] \;\geq\; d\sum_{i=1}^m b_i^2.$$

Each transition contributes an independent term $b_i\,\varepsilon_i'$ to the deviation; the expected squared norm adds up across stages.

**Remark (Link to implementation).** Our *Unified Noise Field* deterministically maps token coordinates $(y, x, d)$ to a fixed sample $\epsilon = \mathcal{E}_s(y, x, d)$, ensuring that all stages reuse the same noise realization. The practical re-noise $z \leftarrow \beta z + \alpha\,\epsilon$ instantiates $(a, b)$ in Theorem 1, behaving as a time reparameterization and avoiding the irreducible drift identified in Theorem 2.

## A.2. Detailed Experiment Settings

**Model Configurations and Evaluation Protocols.** We evaluate Fresco across five representative generative models covering text-to-image and text-to-video generation. For text-to-image tasks, we include FLUX.1-dev, FLUX.1-Lite-8B, FLUX-Schnell, and FLUX.1-dev-int8. For text-to-video generation, we evaluate HunyuanVideo. All models are tested under their native resolutions and standard evaluation protocols.

### A.2.1. Text-to-Image Generation

**FLUX.1-dev.** We use the DrawBench benchmark with 200 prompts spanning objects, scenes, fine-grained descriptions, and compositional reasoning. Images are generated at $1024 \times 1024$ resolution. We further evaluate images of $1440 \times 1440$ and $2048 \times 2048$ resolution in discussion section. Following prior work, we report ImageReward and CLIP Score as indicators of visual quality and text–image alignment.

**FLUX.1-lite-8B.** FLUX.1-Lite-8B is a distilled variant of FLUX.1-dev. We evaluated it on the same DrawBench prompts at its native resolution $1024 \times 1024$ with 28 sampling steps. We quantify the generation quality using CLIP-IQA and CLIP Score.

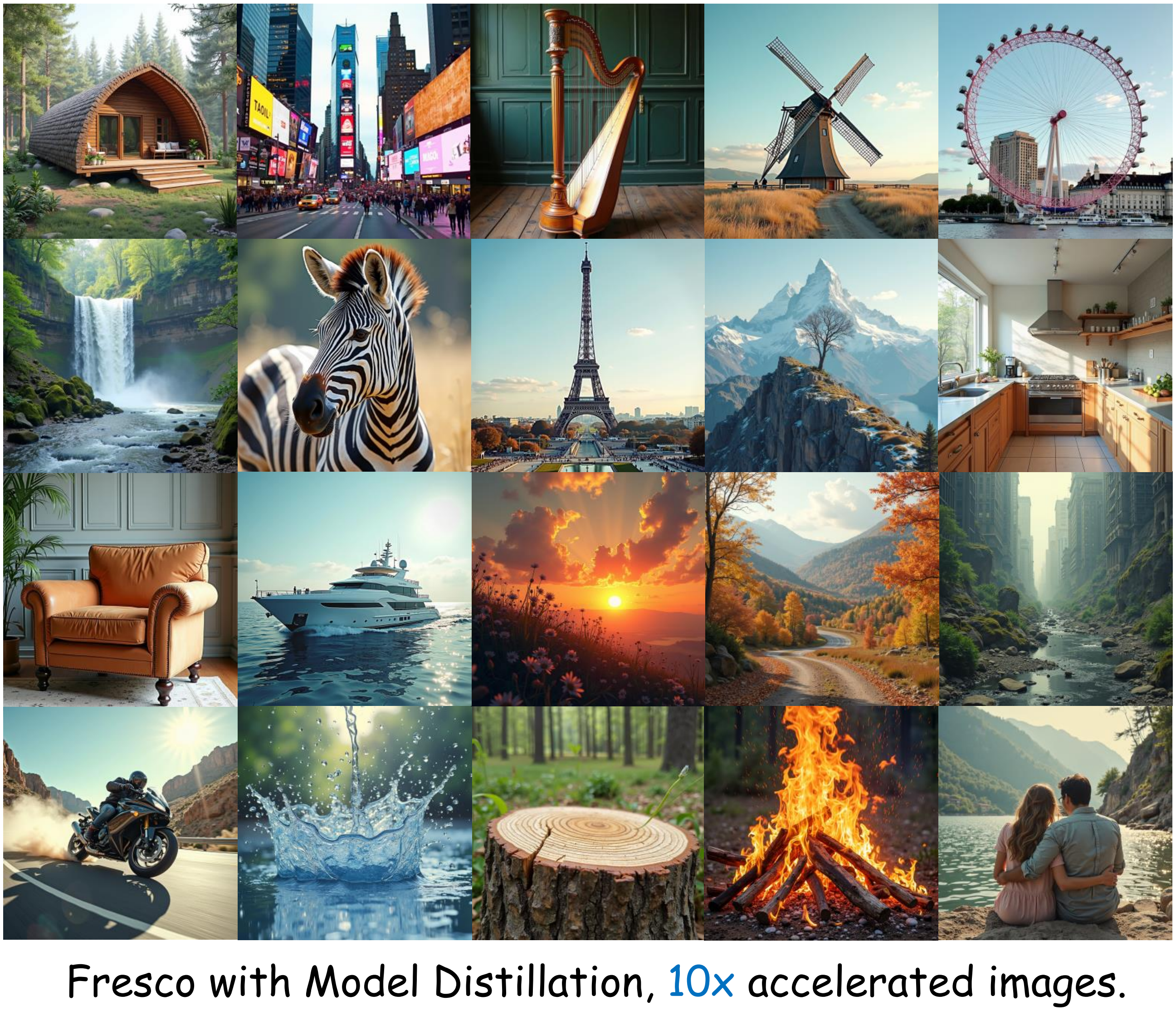


Figure 7. **Visualization of Fresco on FLUX.1-lite-8B.** Fresco combined with model-distillation produces 10× accelerated high-quality images. Distilled models gain additional speedups when paired with CoT's progressive resolution sampling.

**FLUX-Schnell.** As a step-distilled variant of FLUX.1-dev, FLUX-Schnell is evaluated with 200 DrawBench prompts at $1024 \times 1024$ resolution. Quality assessment follows the same metrics as Lite-8B: CLIP-IQA and CLIP Score.

**FLUX.1-dev-int8.** To validate compatibility with quantization, we evaluate the torchao INT8 version of FLUX.1-dev. All settings follow the DrawBench protocol at $1024 \times 1024$ resolution, reporting CLIP-IQA and CLIP Score to measure fidelity and semantic alignment under quantized weights.

**TaylorSeer.** To examine orthogonality between Fresco and feature caching, we integrate the official TaylorSeer configuration with $N$=3 and order-2 Taylor prediction. We evaluate it with CLIP-IQA and CLIP Score as well.

### A.2.2. Text-to-Video Generation

**HunyuanVideo.** For text-to-video generation, we adopt the VBench evaluation suite using 946 prompts from the official metadata file. For each prompt in VBench, we generate one video at $720 \times 1280$ with 125 frames. VBench reports 18 human-aligned metrics covering motion quality, appearance consistency, temporal coherence, semantic alignment, and overall visual fidelity. We report the aggregated VBench Total Score, Quality Score and Semantic Score along with latency and FLOPs to quantify acceleration.


Figure 8. **Visualization of Fresco on FLUX.1-dev-INT8.** Fresco combined with quantization enables 9× faster text-to-image generation while preserving visual quality.