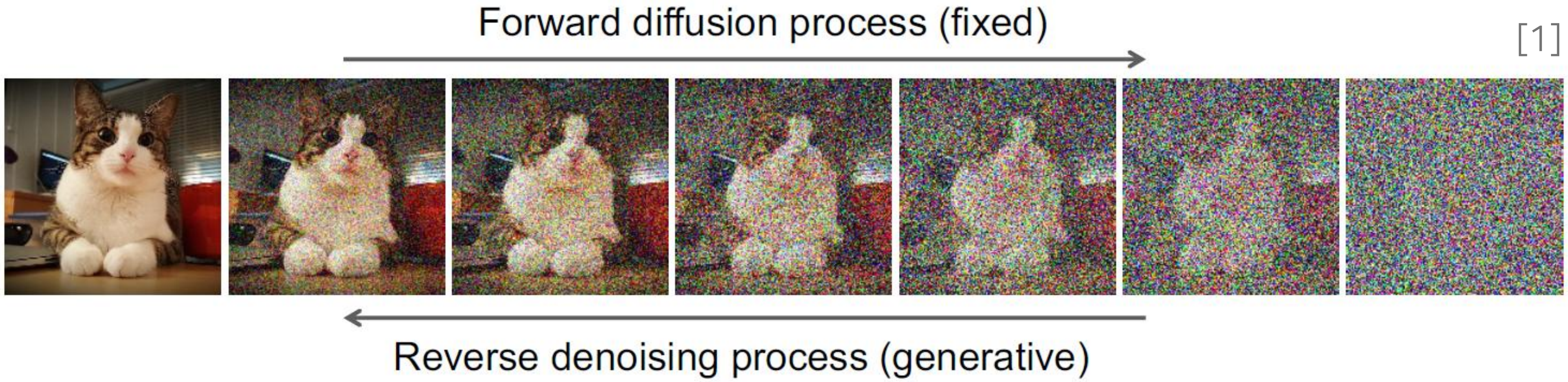
Tortoise and Hare Guidance: Accelerating Diffusion Model Inference with Multirate Integration

Yunghee Lee, Byeonghyun Pak, Junwha Hong, Hoseong Kim

Agency for Defense Development

Diffusion model inference

data
x_0



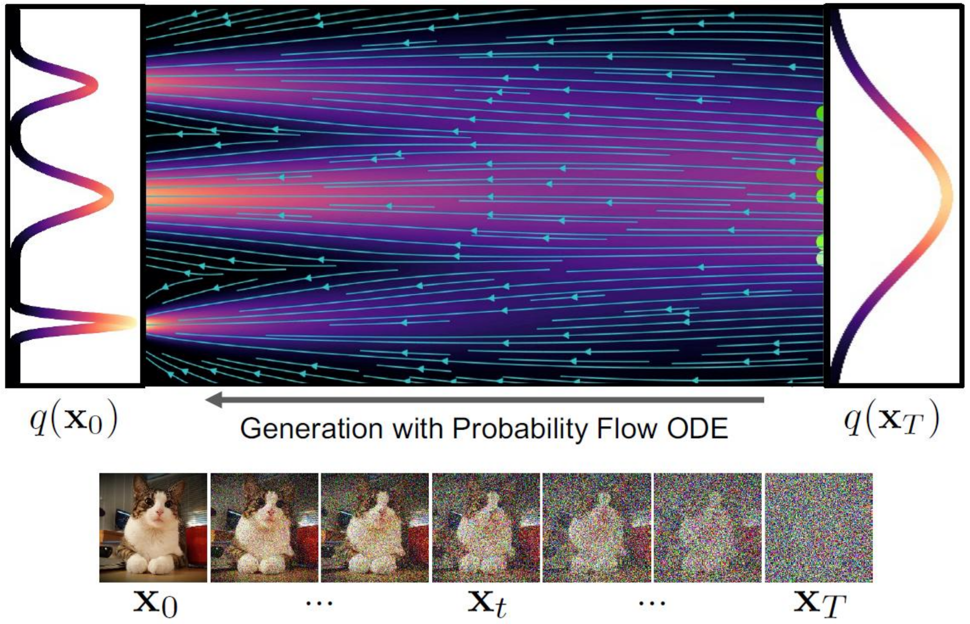
[1]

noise
x_T

Repeatedly remove noise from pure noise x_T to sample x_0

[1] https://cvpr2022-tutorial-diffusion-models.github.io

Diffusion ODE



[1]

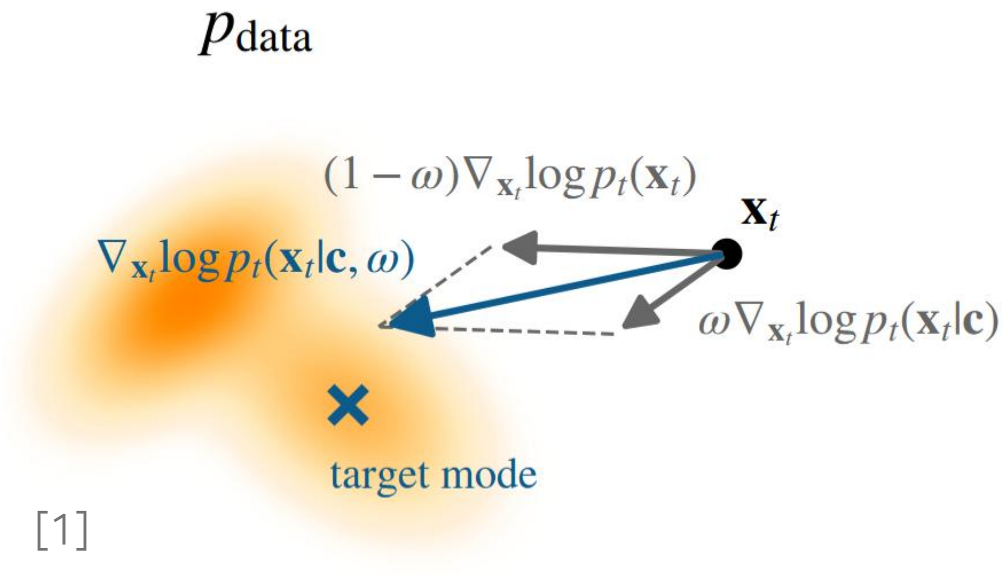
$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = f(t)x_t + \frac{g^2(t)}{2\sigma_t}\hat{\epsilon}_\theta(x_t)$$

$$x_T \sim \mathcal{N}(0, I)$$

Inference is solving an **ODE initial-value problem**

[1] https://cvpr2022-tutorial-diffusion-models.github.io

Classifier-Free Guidance (CFG)



[1]

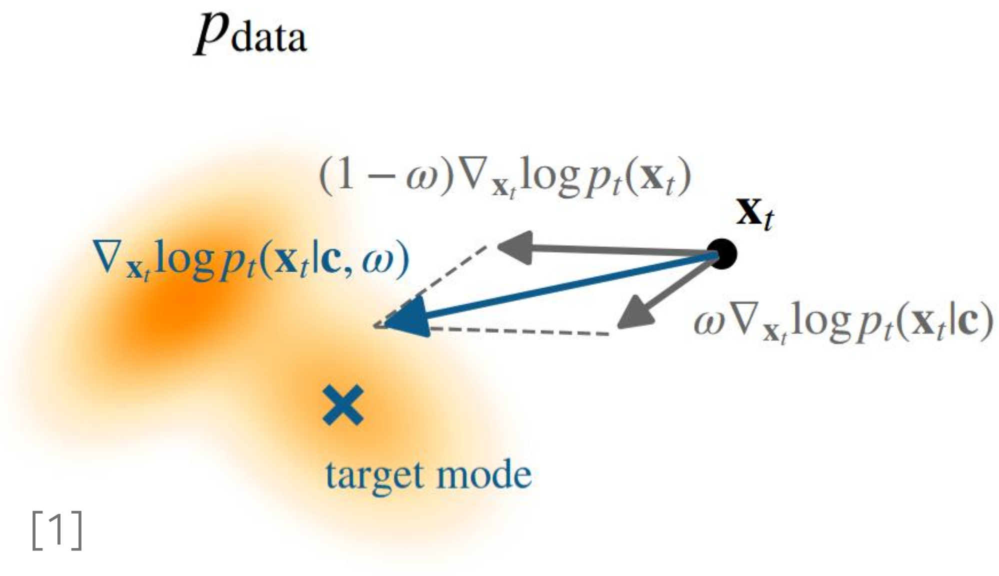
Conditional output $\hat{\epsilon}_c(x_t)$
Unconditional output $\hat{\epsilon}_\emptyset(x_t)$

$$\hat{\epsilon}_c^\omega(x_t) = \hat{\epsilon}_\emptyset(x_t) + \omega \cdot \left(\hat{\epsilon}_c(x_t) - \hat{\epsilon}_\emptyset(x_t)\right)$$

CFG uses **2 NFE per step** (twice the work)

[1] Chieh-Hsin Lai, Yang Song, Dongjun Kim, Yuki Mitsufuji, and Stefano Ermon. The Principles of Diffusion Models. arXiv preprint arXiv:2510.21890, 2025.

Classifier-Free Guidance (CFG)



[1]

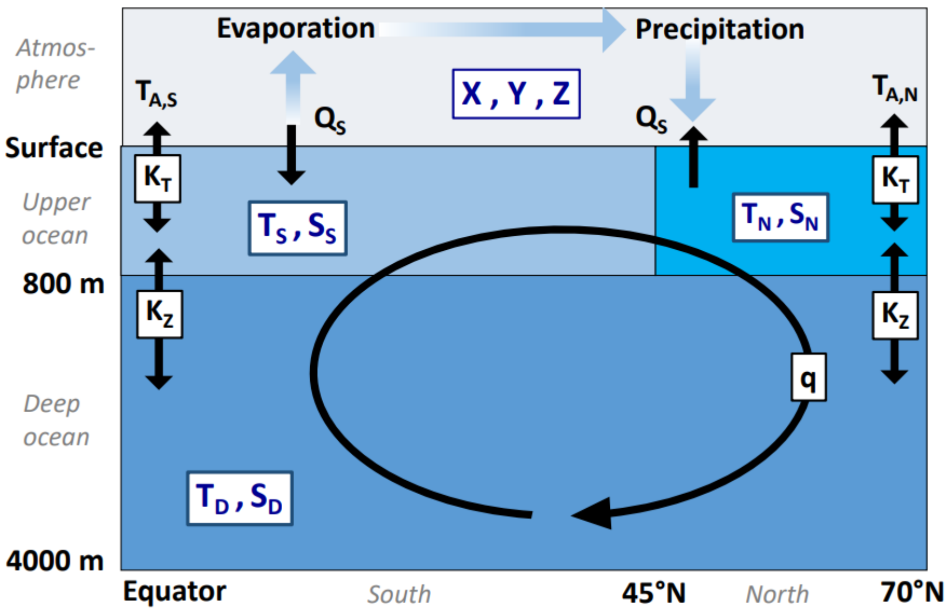
$$\begin{aligned}\hat{\epsilon}_c^{\omega}(x_t) &= \hat{\epsilon}_\emptyset(x_t) + \omega \cdot \left(\hat{\epsilon}_c(x_t) - \hat{\epsilon}_\emptyset(x_t)\right) \\ &= \hat{\epsilon}_c(x_t) + (\omega - 1) \cdot \Delta\hat{\epsilon}_c(x_t)\end{aligned}$$

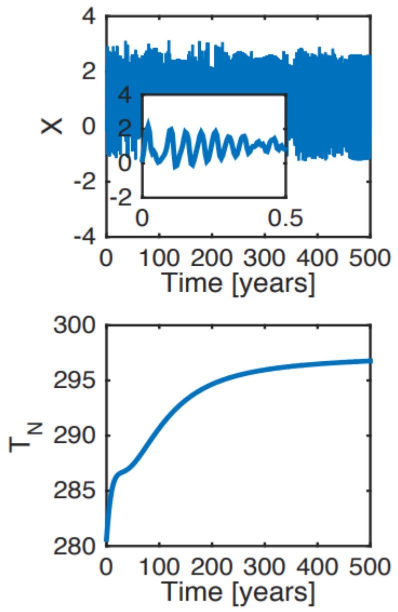
$$(\, \Delta\hat{\epsilon}_c(x_t) \coloneqq \hat{\epsilon}_c(x_t) - \hat{\epsilon}_\emptyset(x_t) \,)$$

CFG uses noise estimate and additional guidance, can we leverage this?

[1] Chieh-Hsin Lai, Yang Song, Dongjun Kim, Yuki Mitsufuji, and Stefano Ermon. The Principles of Diffusion Models. arXiv preprint arXiv:2510.21890, 2025.

How to tackle systems

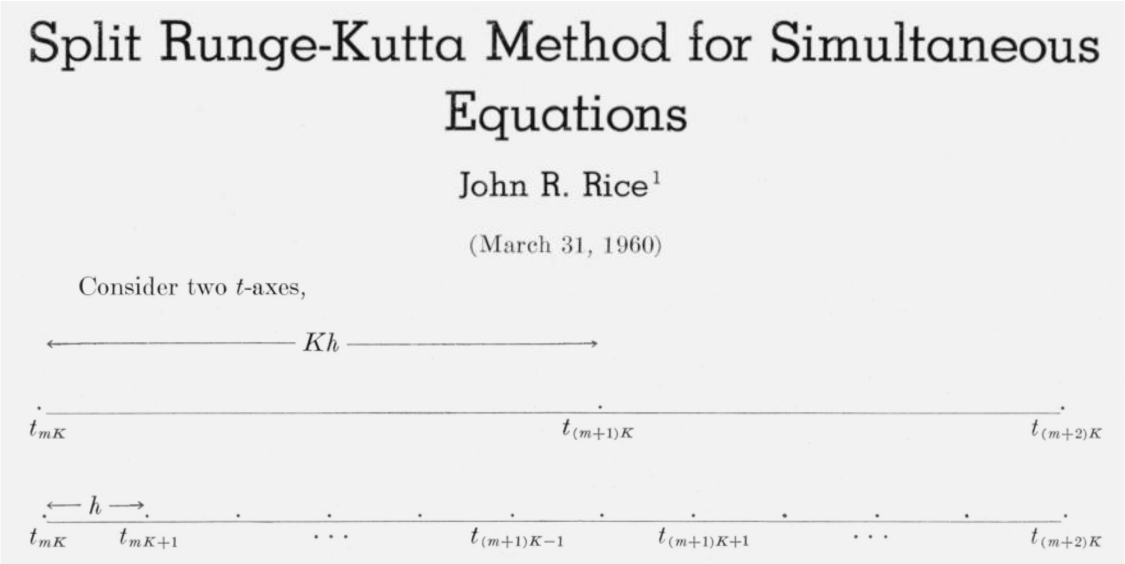




Systems can have **multiple time scales**

[1] https://cnls.lanl.gov/tim2023/talks/Sandu_tim2023.pdf

How to tackle systems


Split Runge-Kutta Method for Simultaneous Equations

John R. Rice[1]

(March 31, 1960)

Consider two t-axes,

[1]

Multirate Integration helps model such systems

[1] John R Rice. Split Runge-Kutta methods for simultaneous equations. Journal of Research of the National Institute of Standards and Technology, 60, 1960.

A multirate formulation

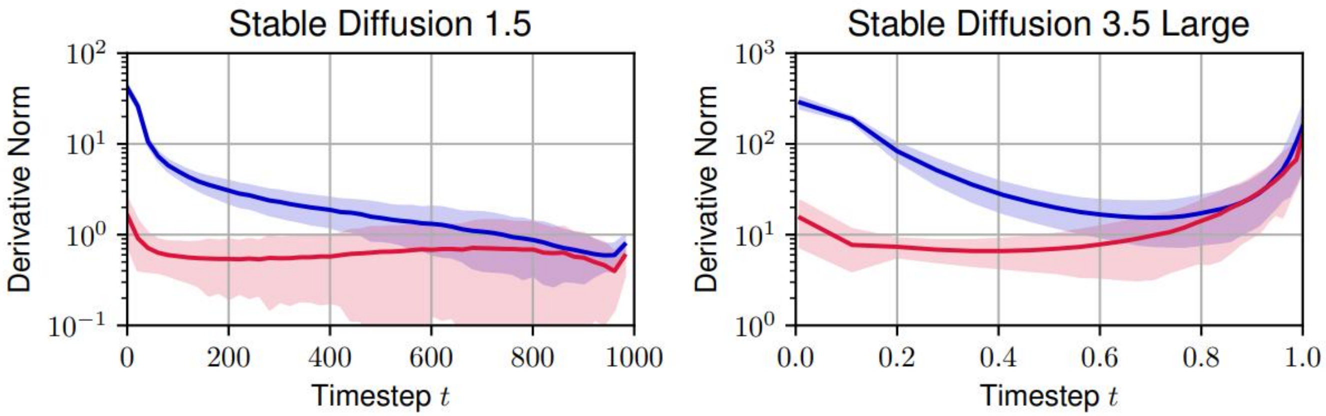


Figure 2: **Time-derivative norms of the noise estimate $\hat{\epsilon}_c(x_t)$ and additional guidance $\delta\hat{\epsilon}_c(x_t)$.** We plot the L2 norms of the time derivatives $\frac{\mathrm{d}}{\mathrm{d}t}\hat{\epsilon}_c(x_t)$ and $\frac{\mathrm{d}}{\mathrm{d}t}\delta\hat{\epsilon}_c(x_t)$ across diffusion timesteps for Stable Diffusion 1.5 and Stable Diffusion 3.5 Large. The results confirm that the noise estimate exhibits greater temporal sensitivity compared to the guidance term. Shaded areas denote two standard deviations over multiple prompts.

We find additional guidance has **less temporal sensitivity** than noise estimate

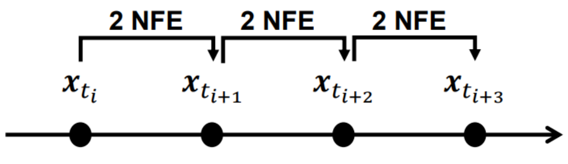
∴ multiple time scales

A multirate formulation

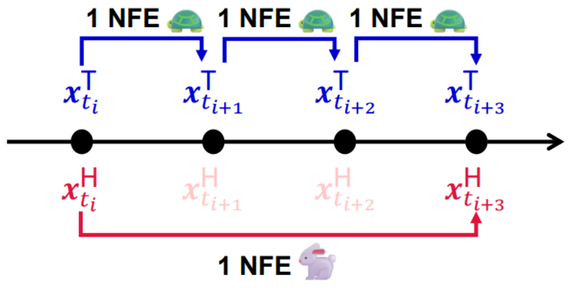
$$\frac{\mathrm{d}}{\mathrm{d}t}x_t = f(t)x_t + \frac{g^2(t)}{2\sigma_t}\hat{\epsilon}_\theta(x_t, t) \quad \Rightarrow \quad \begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}x_t^{\mathrm{T}} &= f(t)x_t^{\mathrm{T}} + \frac{g^2(t)}{2\sigma_t}\hat{\epsilon}_c\left(x_t^{\mathrm{T}} + x_t^{\mathrm{H}}\right) \\ \frac{\mathrm{d}}{\mathrm{d}t}x_t^{\mathrm{H}} &= f(t)x_t^{\mathrm{H}} + \frac{g^2(t)}{2\sigma_t}(\omega - 1)\cdot\Delta\hat{\epsilon}_c\left(x_t^{\mathrm{T}} + x_t^{\mathrm{H}}\right) \end{aligned}$$

We split the diffusion ODE into tortoise equation (x_t^{T}) and hare equation (x_t^{H})

THG: Tortoise and Hare Guidance



(a) Standard diffusion ODE (CFG)



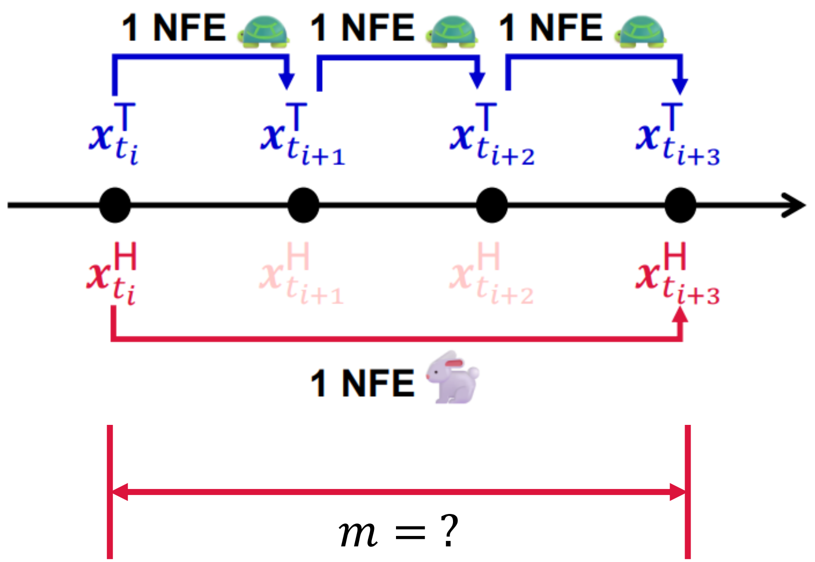
(b) Multirate system of ODEs (Ours)

Algorithm 1 Tortoise and Hare Guidance Algorithm

Require: $x_T \sim \mathcal{N}(0, \sigma_T^2 I)$ ▷ Initial noise
Require: $\omega \geq 0$ ▷ Guidance scale
Require: $\{t_i\}_{0 \leq i \leq N}, t_0 = T, t_N = 0$ ▷ Fine-grained timestep grid
Require: $C \subset \{t_i | 0 \leq i \leq N\}, 0 \in C, T \in C$ ▷ Coarse timestep grid
$x^{\mathsf{T}}_T \leftarrow x_T$
$x^{\mathsf{H}}_T \leftarrow 0$
for $i = 0$ **to** $N - 1$ **do**
  $\hat{\epsilon}_c \leftarrow \hat{\epsilon}_\theta(x^{\mathsf{T}}_{t_i} + x^{\mathsf{H}}_{t_i}, c)$ ▷ 1 NFE
  $x^{\mathsf{T}}_{t_{i+1}} \leftarrow \mathrm{Solver}(x^{\mathsf{T}}_{t_i}, \hat{\epsilon}_c, t_i, t_{i+1})$ ▷ Compute $x^{\mathsf{T}}_{t_{i+1}}$ given $x^{\mathsf{T}}_{t_i}$
  if $t_i \in C$ **then**
    $\hat{\epsilon}_\varnothing \leftarrow \hat{\epsilon}_\theta(x^{\mathsf{T}}_{t_i} + x^{\mathsf{H}}_{t_i}, \varnothing)$ ▷ 1 NFE (only if $t_i \in C$)
    $\delta\hat{\epsilon}_c \leftarrow \hat{\epsilon}_c - \hat{\epsilon}_\varnothing$
    $j \leftarrow i$
    repeat ▷ Compute x^{H} up to the next coarse timestep
      $j \leftarrow j + 1$
      $x^{\mathsf{H}}_{t_j} \leftarrow \mathrm{Solver}(x^{\mathsf{H}}_{t_i}, (\omega - 1) \cdot \delta\hat{\epsilon}_c, t_i, t_j)$ ▷ Compute $x^{\mathsf{H}}_{t_j}$ given $x^{\mathsf{H}}_{t_i}$
    until $t_j \in C$ ▷ t_j equals the next coarse timestep at inner loop exit
  end if
end for
$x_0 \leftarrow x^{\mathsf{T}}_0 + x^{\mathsf{H}}_0$
return x_0

THG **reduces total NFE** by integrating the hare in a **coarser timestep grid** C

Challenge



How do we decide C?

Approximation error bound analysis

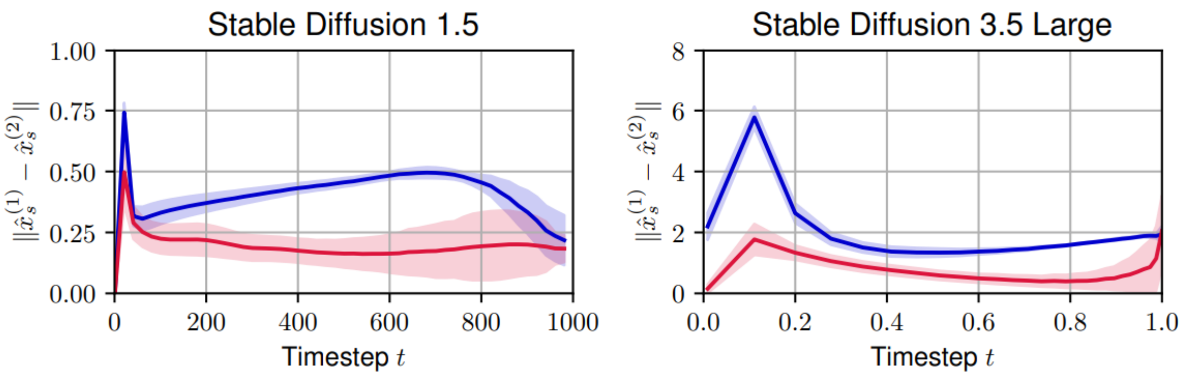


Figure 3: **Approximation error bounds of the tortoise x_t^{T} and the hare x_t^{H}.** We show the per-timestep error bound of the tortoise and the hare terms across sampling steps. The consistently higher bounds for the tortoise curve indicate that the noise estimate is more sensitive to timestep resolution than the additional guidance. Shaded areas denote two standard deviations over multiple prompts.

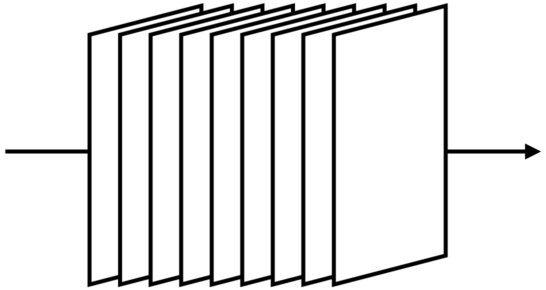
$$\frac{\|\hat{x}^{\mathrm{H}} - x^{\mathrm{H}}\|}{\|\hat{x}^{\mathrm{T}} - x^{\mathrm{T}}\|} \leq \rho$$

$$\Rightarrow m \leq \left(\rho \|c^{\mathrm{T}}\| / \|c^{\mathrm{H}}\|\right)^{1/p}$$

C is decided to keep hare error smaller than tortoise error

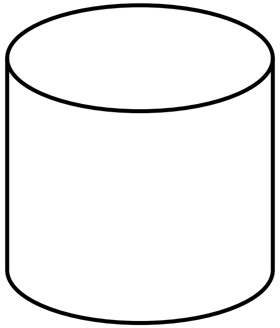
Experiments

Diffusion model



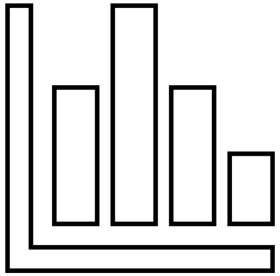
SD 1.5, SD 3.5 Large
AudioLDM 2

Dataset



MS COCO 2014
AudioCaps

Metrics



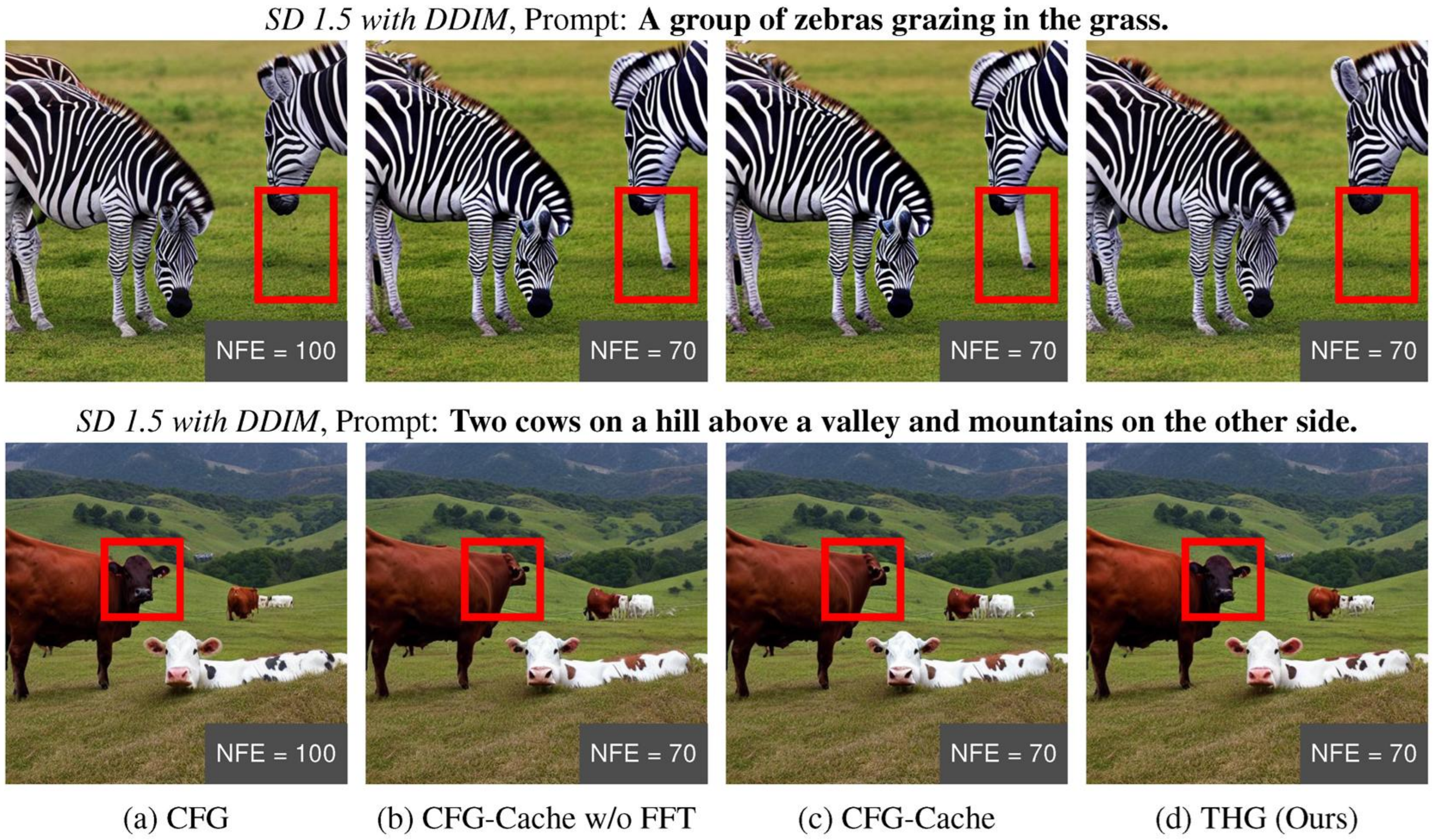
FID·FAD, CMMD
CLIP·CLAP Score, IR

Experiments

Method	NFE ↓	Distributional similarity FID ↓	Distributional similarity CMMD ↓	Prompt fidelity CS ↑	Prompt fidelity IR ↑
Stable Diffusion 1.5 with DDIM					
CFG [18]	100	14.057	0.58885	26.294	0.14765
CFG-Cache w/o FFT [31]	70	14.240	**0.59187**	26.141	0.08757
CFG-Cache [31]	70	14.367	0.59556	26.180	0.09735
THG (Ours)	70	**14.165**	0.59223	**26.189**	**0.11499**
Stable Diffusion 1.5 with DPM-Solver-2 [30]					
CFG [18]	100	13.255	0.60379	26.254	0.16148
CFG-Cache w/o FFT [31]	70	13.387	**0.60665**	26.107	0.10513
CFG-Cache [31]	70	13.468	0.60880	26.147	0.11474
THG (Ours)	70	**12.909**	0.60868	**26.205**	**0.14926**
Stable Diffusion 1.5 with 2nd-order Linear Multistep Method [29]					
CFG [18]	100	13.540	0.60653	26.260	0.15966
CFG-Cache w/o FFT [31]	70	**13.686**	**0.60844**	26.107	0.09881
CFG-Cache [31]	70	13.798	0.61142	26.144	0.10805
THG (Ours)	70	**13.686**	0.61094	**26.204**	**0.15184**
Stable Diffusion 3.5 Large with Euler method					
CFG [18]	56	68.158	0.81106	26.624	1.03569
CFG-Cache w/o FFT [31]	38	67.931	0.76448	26.643	1.00715
CFG-Cache [31]	38	**67.914**	**0.75324**	26.668	1.00745
THG (Ours)	38	68.252	0.80092	**26.672**	**1.02365**

Method	NFE ↓	Distributional similarity FAD ↓	Prompt fidelity CLAP Score ↑
AudioLDM 2 with DDIM			
CFG [18]	100	2.596	0.2409
CFG-Cache w/o FFT [31]	70	2.901	0.2251
THG (Ours)	70	**2.764**	**0.2342**

Experiments



NEURAL INFORMATION PROCESSING SYSTEMS

Tortoise and Hare Guidance: Accelerating Diffusion Model Inference with Multirate Integration

Yunghee Lee, Byeonghyun Pak, Junwha Hong, Hoseong Kim

yhl@add.re.kr / yhlee.add@gmail.com

ADD