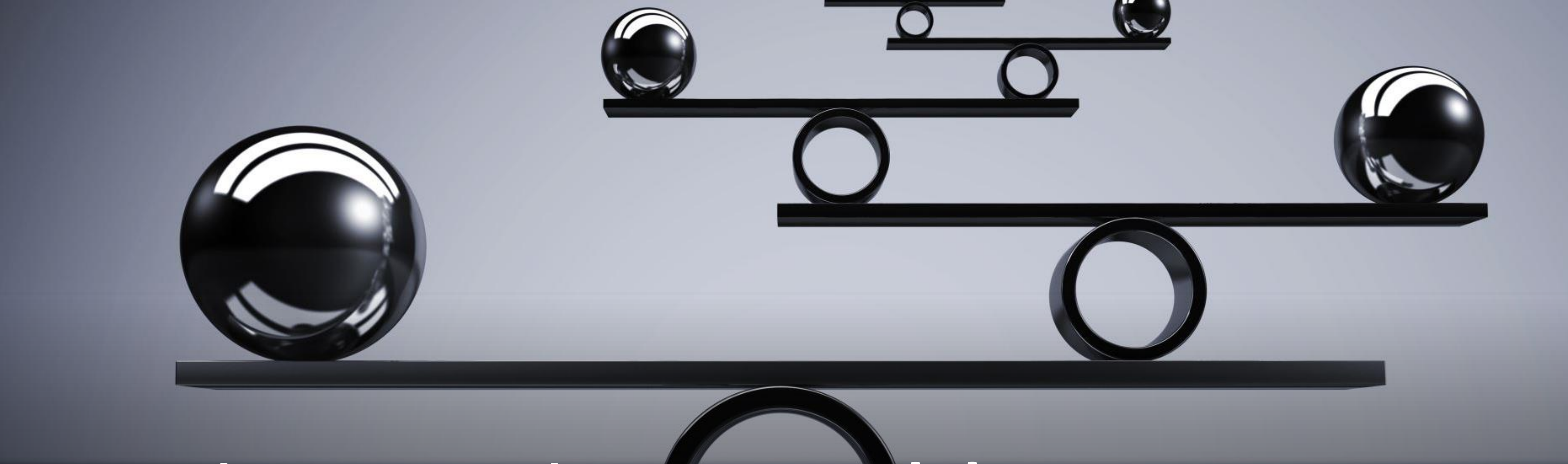
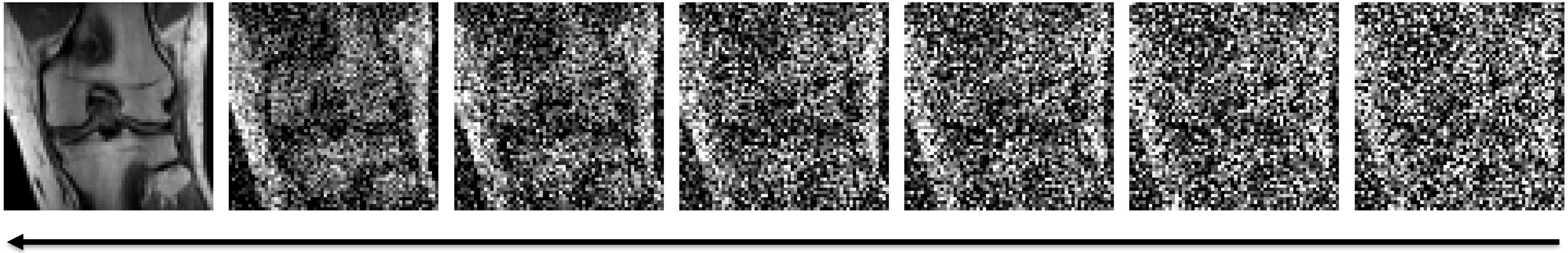


Consistency Trajectory Models:
Learning Probability Flow ODE Trajectory of Diffusion

ICLR 2024 Paper by Sony AI | Presented by Mary-Brenda Akoda

Introduction



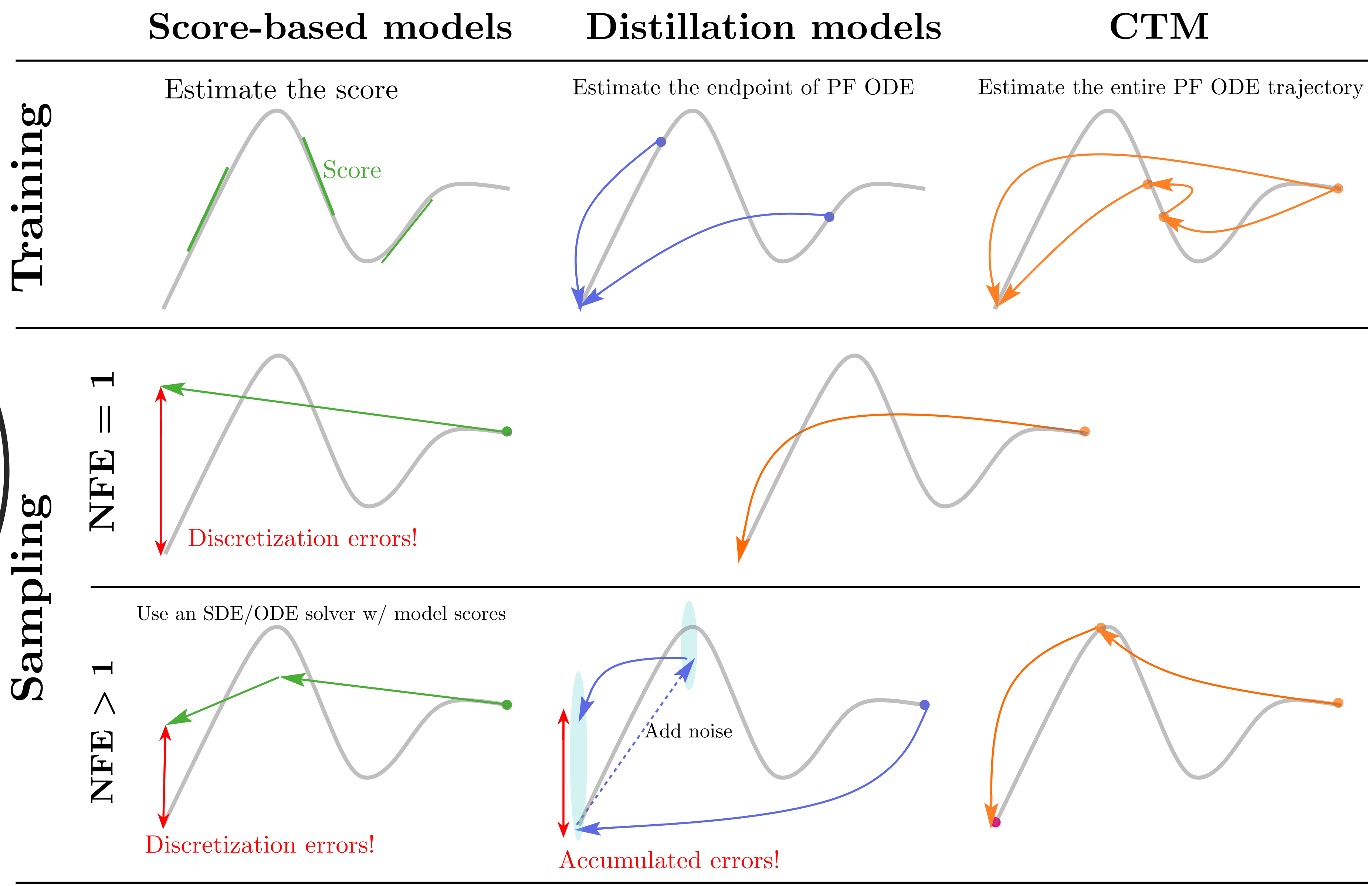
Probability Flow ODE (Reverse Diffusion):

- Same marginal distributions as original SDE, enabling deterministic transformations.

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = -t\boldsymbol{s}_{\boldsymbol{\phi}}(\mathbf{x}_t, t).$$

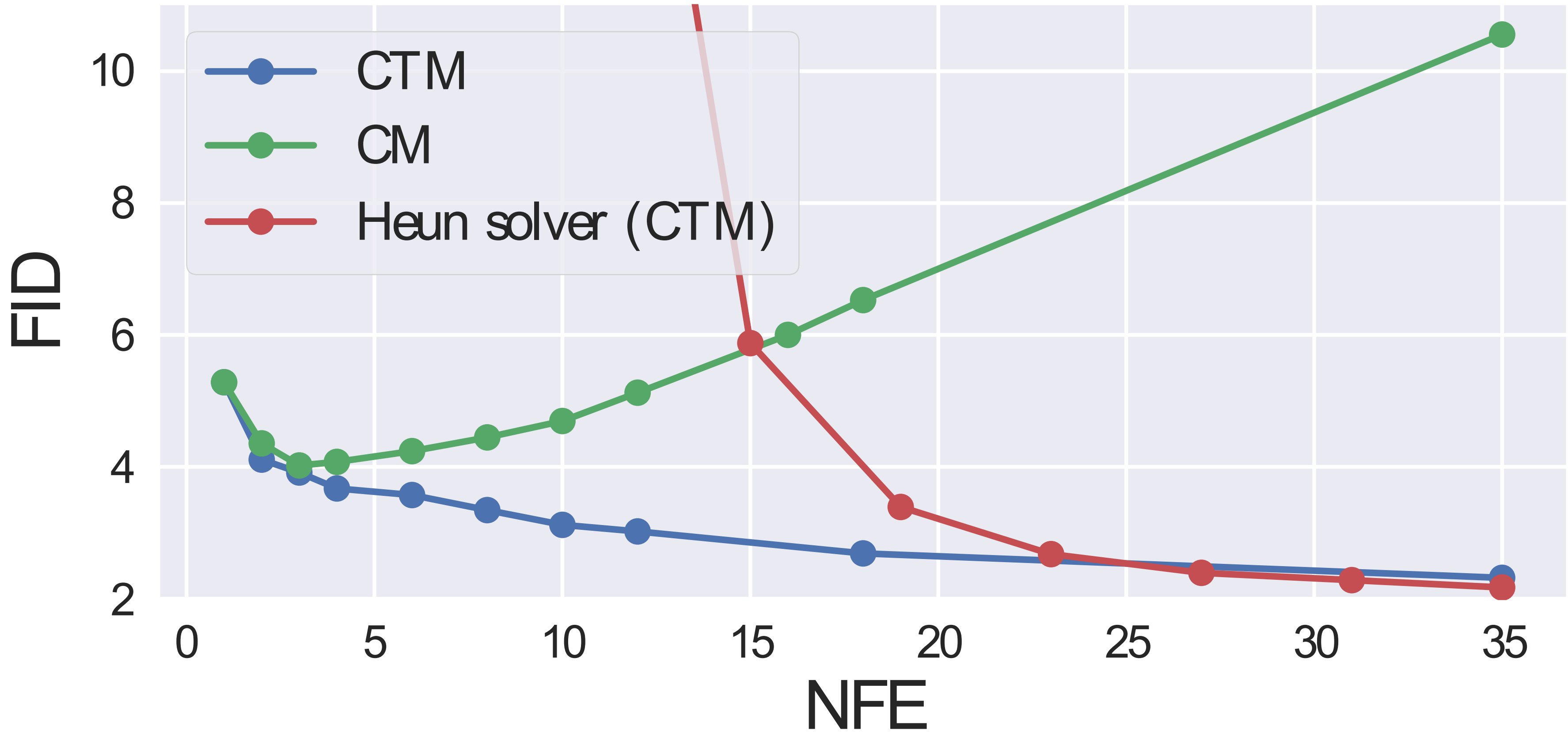
$s_{\emptyset}(x_t, t)$ is neural network approximation of $\nabla \log p_t(x_t)$ *(the score function of $p_t(x_t)$)*

Challenges of DMs and CMs



Score-based models exhibit discretization errors during SDE/ODE solving, while distillation models can accumulate errors in multistep sampling.

Challenges of CMs



CM's multistep sampler significantly degrades in quality with higher NFE. CTM’s deterministic sampling matches Heun's solver as NFE increases (i.e., CTM has a clear trade-off between sample quality and speed).

Key Contributions

CTM models **anytime-to-anytime** jumps along PF ODE

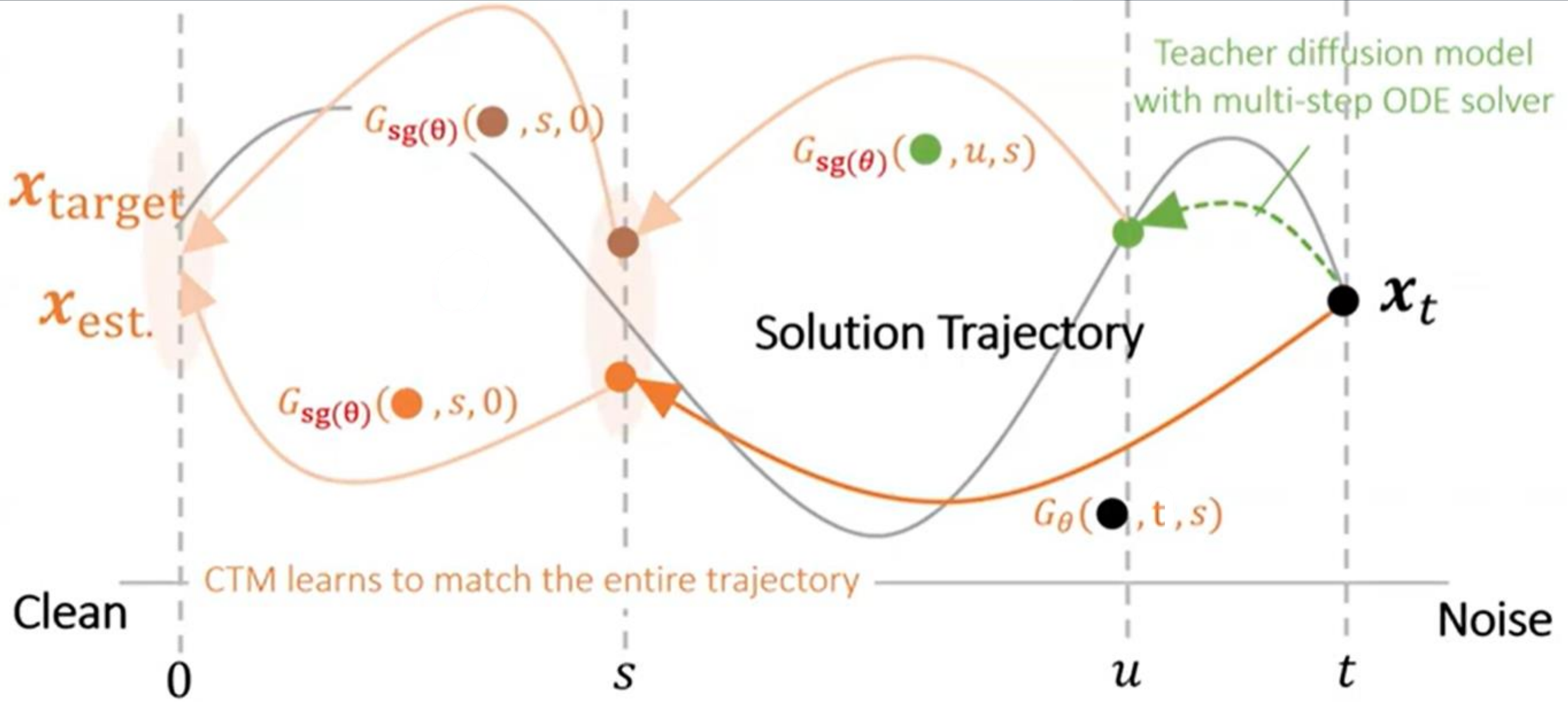
Three Losses: Soft Consistency Loss with Denoising Score Matching (DSM) and GAN losses.

Unifies DMs and CMs by predicting:

Long jumps (integral over time)	Infinitesimal jumps (score function)

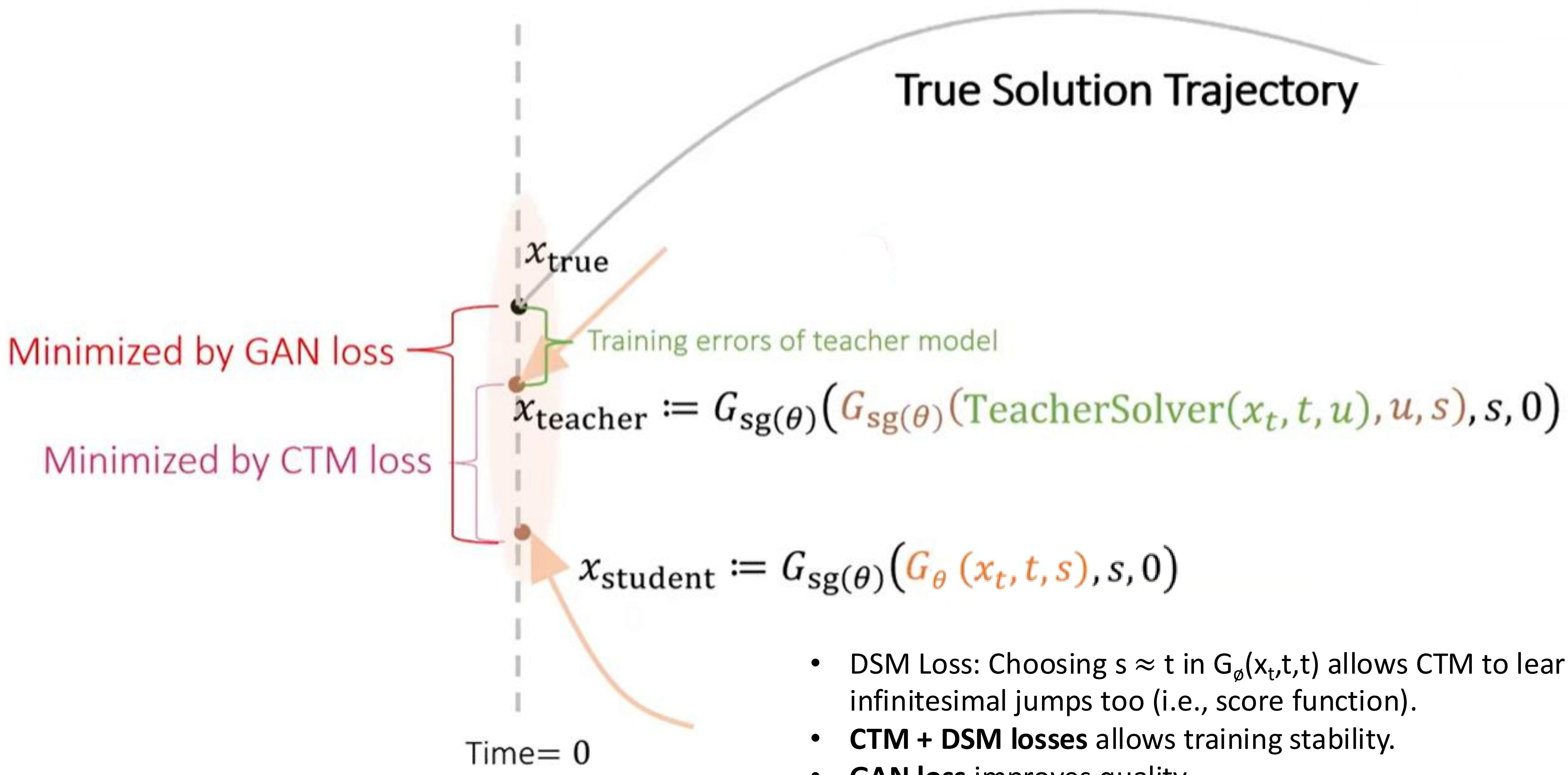
Novel γ-Sampling ensures **better sampling** quality, flexibility & diversity.

CTM Training Mechanism



Student Network (CTM): $G_{ø}$(data start pt., t_{start} , t_{end})

Student Beats Teacher Mechanism



- DSM Loss: Choosing $s \approx t$ in $G_{\phi}(x_t,t,t)$ allows CTM to learn infinitesimal jumps too (i.e., score function).
- **CTM + DSM losses** allows training stability.
- **GAN loss** improves quality.

Student Beats Teacher Results

Teacher: EDM (Sampling steps = 79)

Student: CTM (Sampling steps = 1)

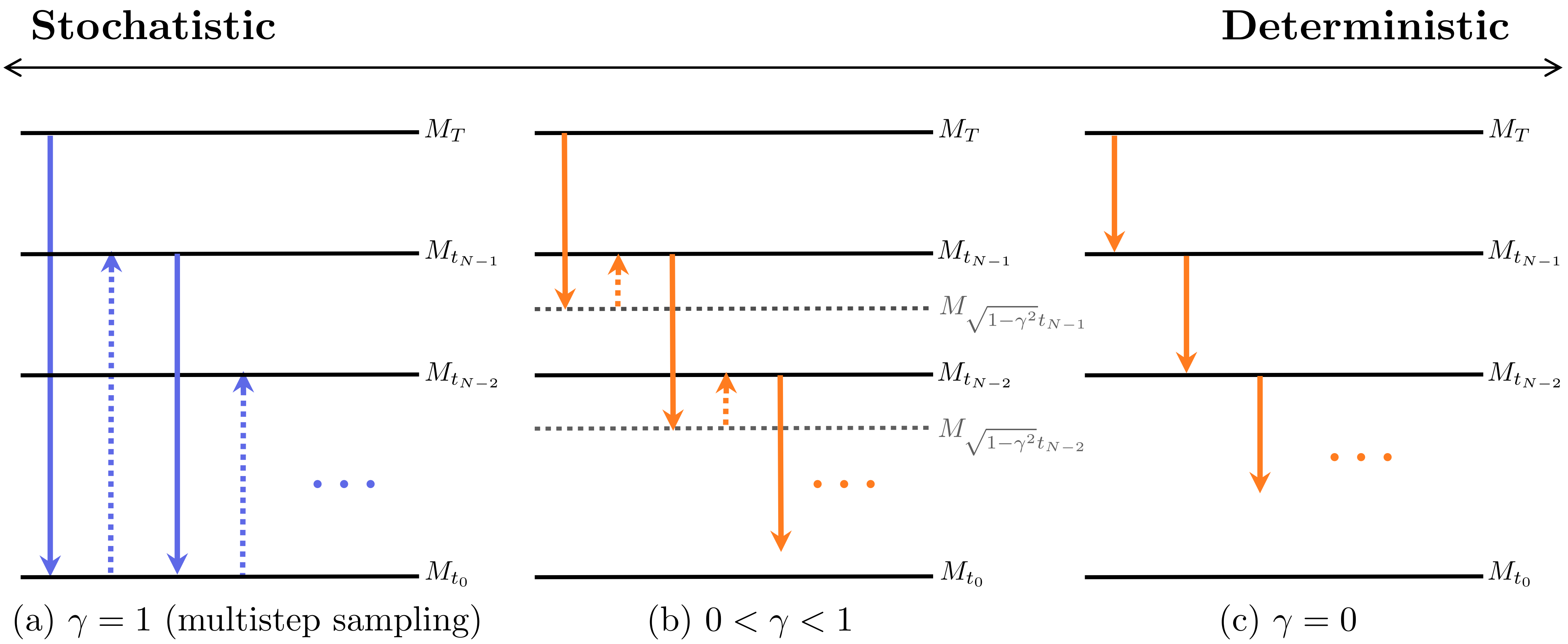
Table 1: Performance comparisons on CIFAR-10.

Model	NFE	Unconditional		Conditional
		FID↓	NLL↓	FID↓
GAN Models				
BigGAN (Brock et al., 2018)	1	8.51	✗	-
StyleGAN-Ada (Karras et al., 2020)	1	2.92	✗	2.42
StyleGAN-D2D (Kang et al., 2021)	1	-	✗	2.26
StyleGAN-XL (Sauer et al., 2022)	1	-	✗	1.85
Diffusion Models – Score-based Sampling				
DDPM (Ho et al., 2020)	1000	3.17	3.75	-
DDIM (Song et al., 2020)	100	4.16	-	-
	10	13.36	-	-
Score SDE (Song et al., 2020)	2000	2.20	3.45	-
VDM (Kingma et al., 2021)	1000	7.41	2.49	-
LSGM (Vahdat et al., 2021)	138	2.10	3.43	-
EDM (Karras et al., 2022)	35	2.01	2.56	1.82
Diffusion Models – Distillation Sampling				
KD (Luhman & Luhman, 2021)	1	9.36	✗	-
DFNO (Zheng et al., 2023)	1	5.92	✗	-
Rectified Flow (Liu et al., 2022)	1	4.85	✗	-
PD (Salimans & Ho, 2021)	1	9.12	✗	-
CD (official report) (Song et al., 2023)	1	3.55	✗	-
CD (retrained)	1	10.53	✗	-
CD + GAN (Lu et al., 2023)	1	2.65	✗	-
CTM (ours)	1	1.98	**2.43**	1.73
PD (Salimans & Ho, 2021)	2	4.51	-	-
CD (Song et al., 2023)	2	2.93	-	-
CTM (ours)	2	**1.87**	**2.43**	**1.63**

Table 2: Performance comparisons on ImageNet 64 × 64.

Model	NFE	FID↓	IS↑
ADM (Dhariwal & Nichol, 2021)	250	2.07	-
EDM (Karras et al., 2022)	79	2.44	48.88
BigGAN-deep (Brock et al., 2018)	1	4.06	-
StyleGAN-XL (Sauer et al., 2022)	1	2.09	**82.35**
Diffusion Models – Distillation Sampling			
PD (Salimans & Ho, 2021)	1	15.39	-
BOOT (Gu et al., 2023)	1	16.3	-
CD (Song et al., 2023)	1	6.20	40.08
CTM (ours)	1	1.98	70.86
PD (Salimans & Ho, 2021)	2	8.95	-
CD (Song et al., 2023)	2	4.70	-
CTM (ours)	2	**1.79**	64.14

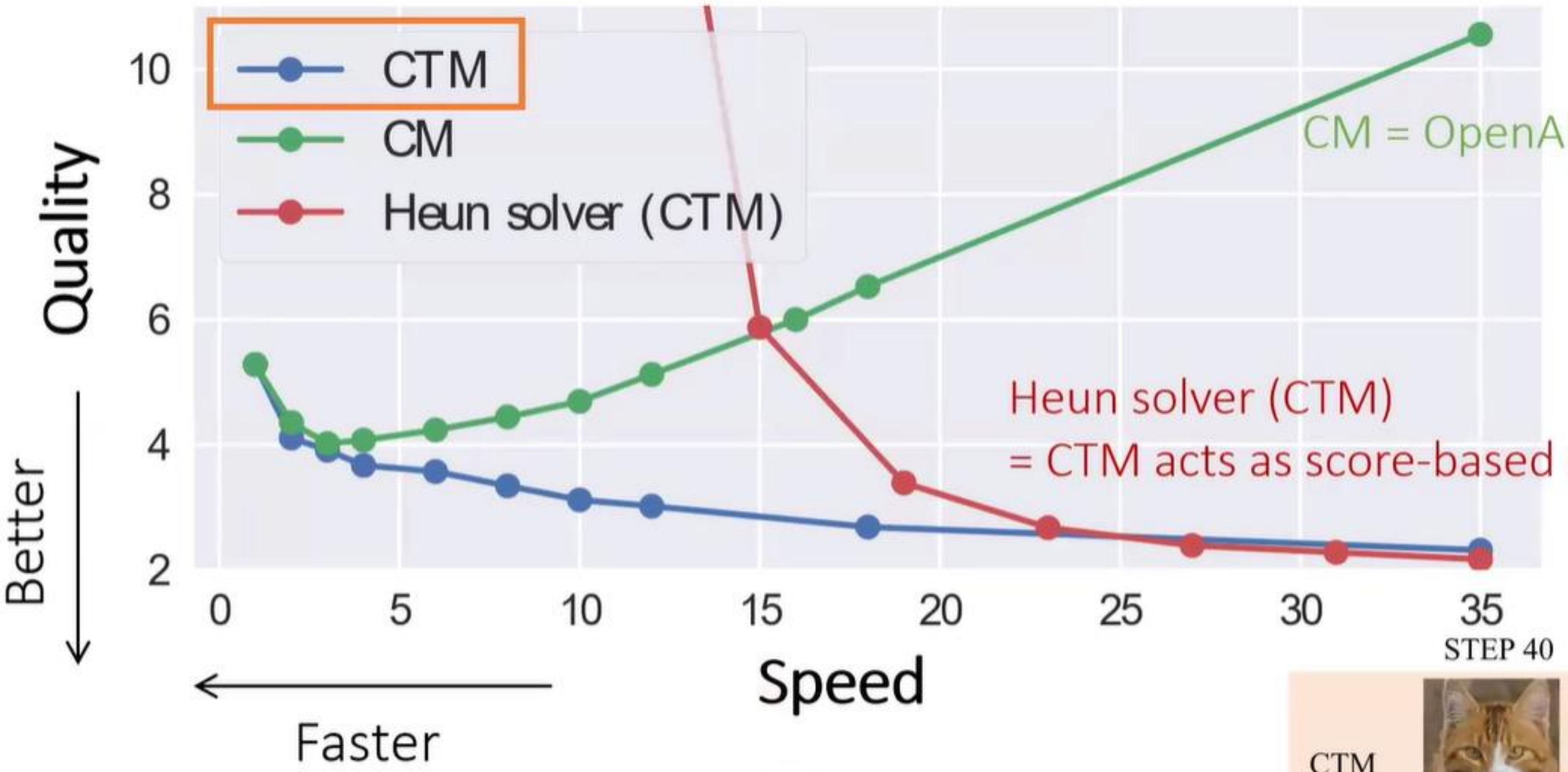
CTM's Novel and Flexible γ-Sampling



Same as CM multistep: Accumulation errors due to overlapping regions

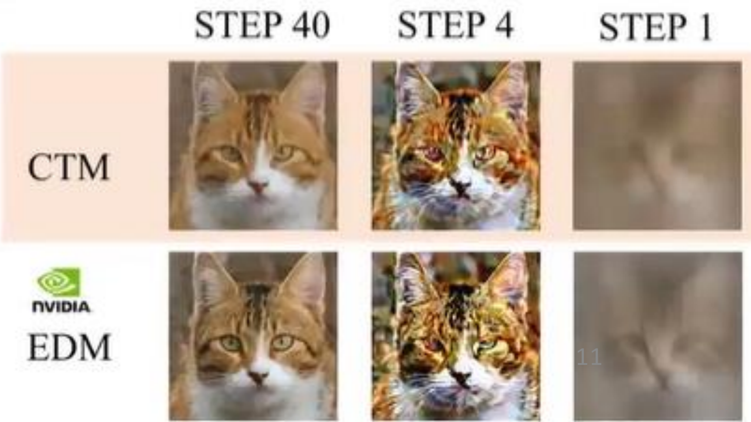
- ✓ No accumulation errors
- ✓ No discretization errors

Multi-steps: Clear Trade-off b/w Speed and Quality



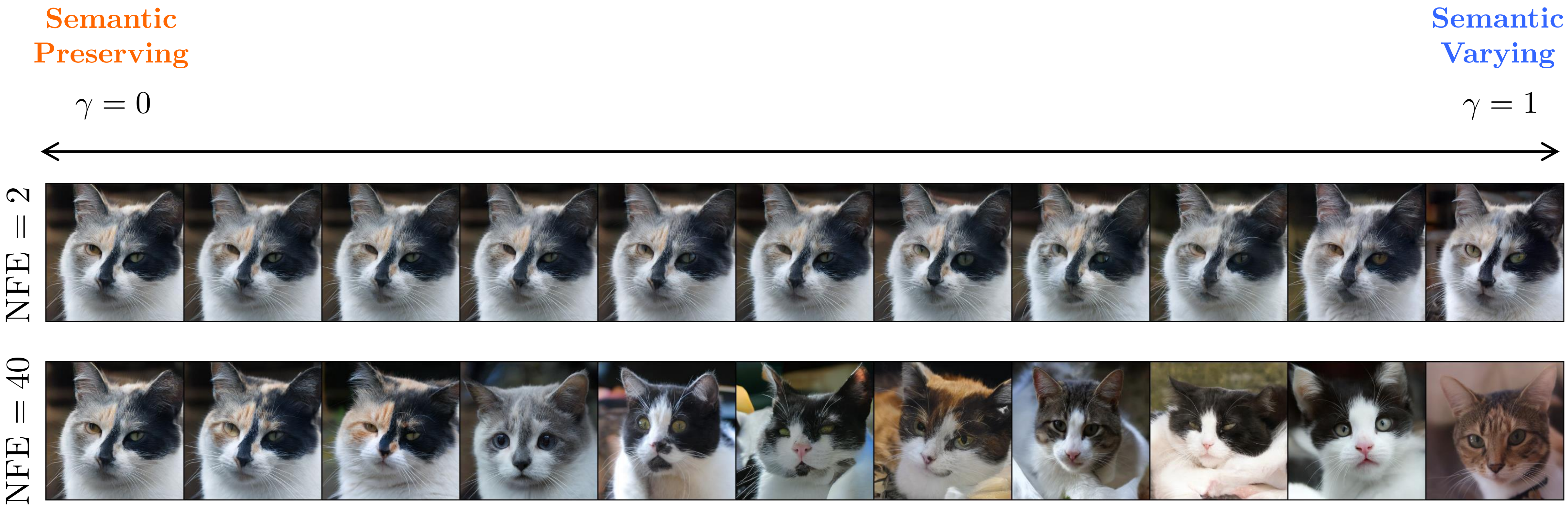
CM = OpenAI's model

Heun solver (CTM)
= CTM acts as score-based models



CTM has a clear trade-off between sample quality and speed

Multi-steps: Diversity of CTM's γ-Sampling



CTM Summary

Unified framework of score-based and distillation models

Flexible in training

Supports both distillation and training from scratch.	Faster convergence and stability due to soft matching and DSM losses	Integration with DSM and adversarial training boosts performance.

Flexible and Powerful in sampling

Supports all diffusion inference techniques	Clear trade-off between sample quality and speed	Student beats teacher as distillation due to GAN loss	Controllable on stochasticity and semantics