Learning to **D**iscretize **D**enoising **D**iffusion (**LD3**) ODEs

Vinh Tong
and
Dung Hoang, Anji Liu, Guy Van den Broeck, Mathias Niepert

Flux 70 default steps
7 seconds

Prompt: 'A cloud formation in the sky spelling "Hope" in soft, fluffy letters'

Flux 30 default steps
3 seconds

Flux 70 default steps
7 seconds

Prompt: 'A cloud formation in the sky spelling "Hope" in soft, fluffy letters'

Flux 30 default steps
3 seconds

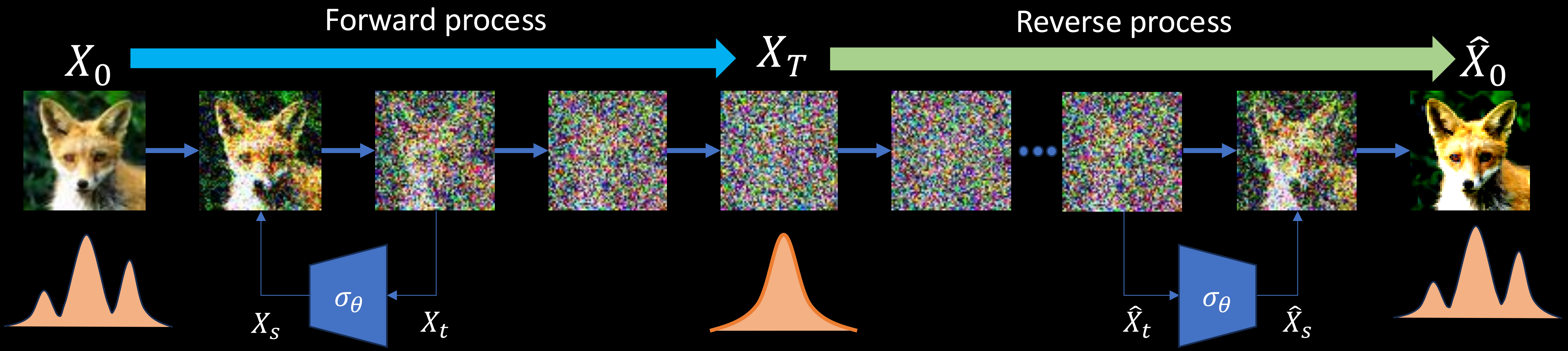
Flux 70 default steps
7 seconds

Flux 30 optimized steps
3 seconds

Prompt: ‘A cloud formation in the sky spelling “Hope” in soft, fluffy letters’

How Diffusion Models Work

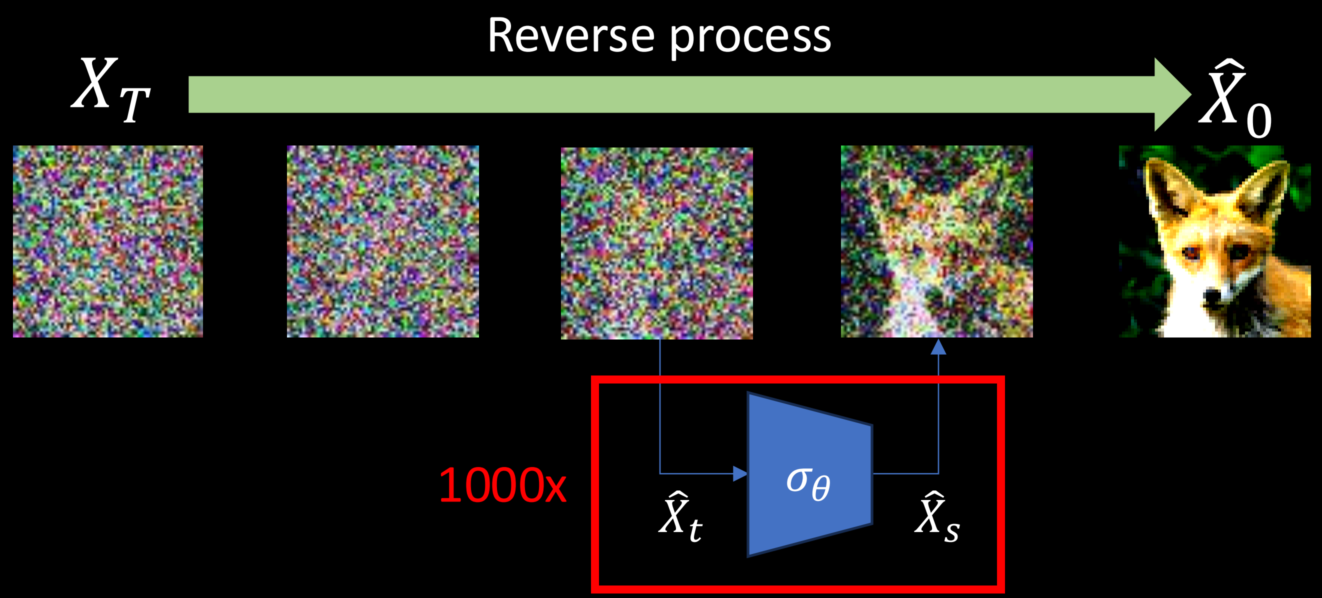
"Diffusion models transform noise into images by learning to reverse a step-by-step noising process."



When $t \rightarrow s$:

- **Training** = Learning an ODE
- **Sampling** = Solving the ODE from a prior

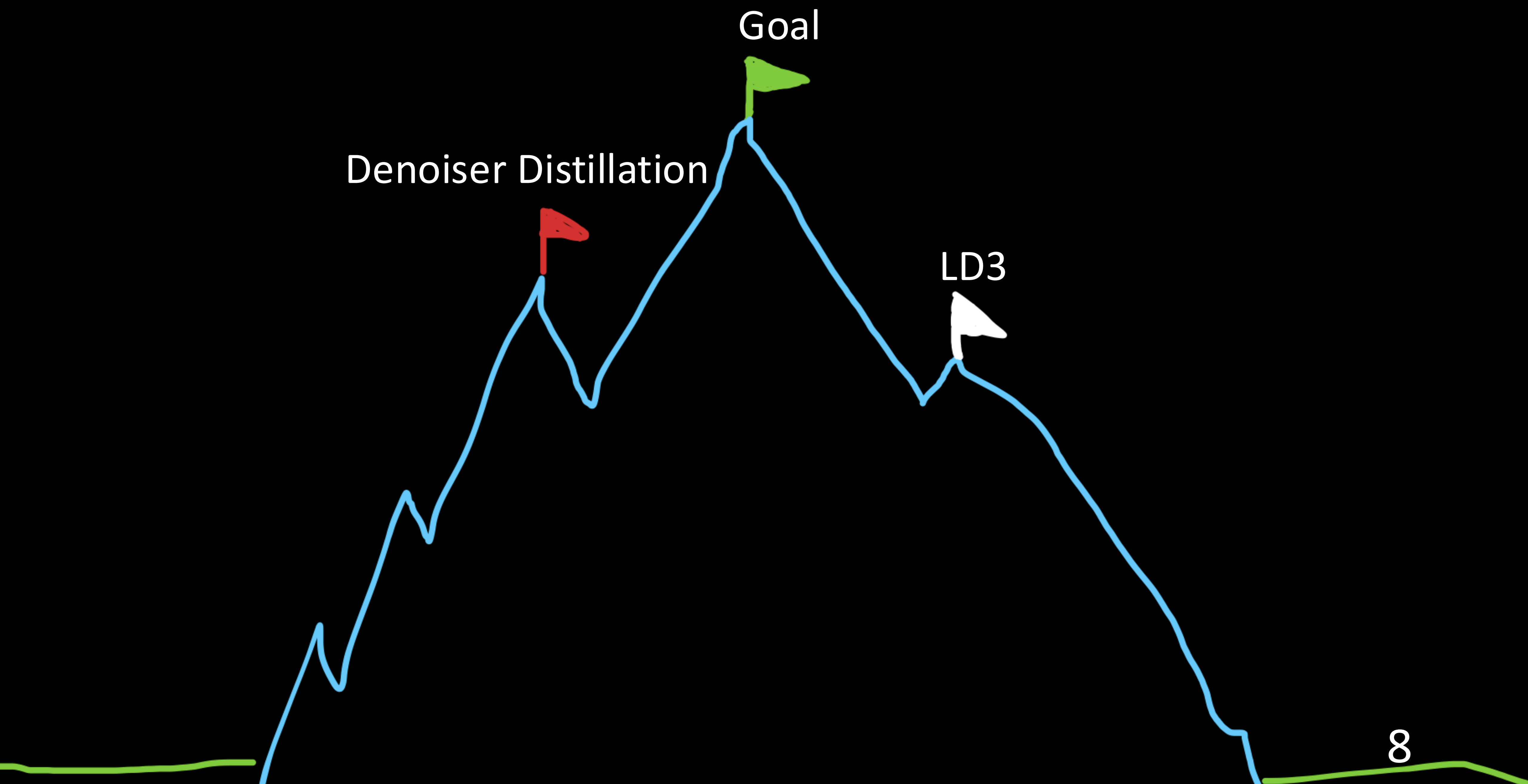
Diffusion Models are slow



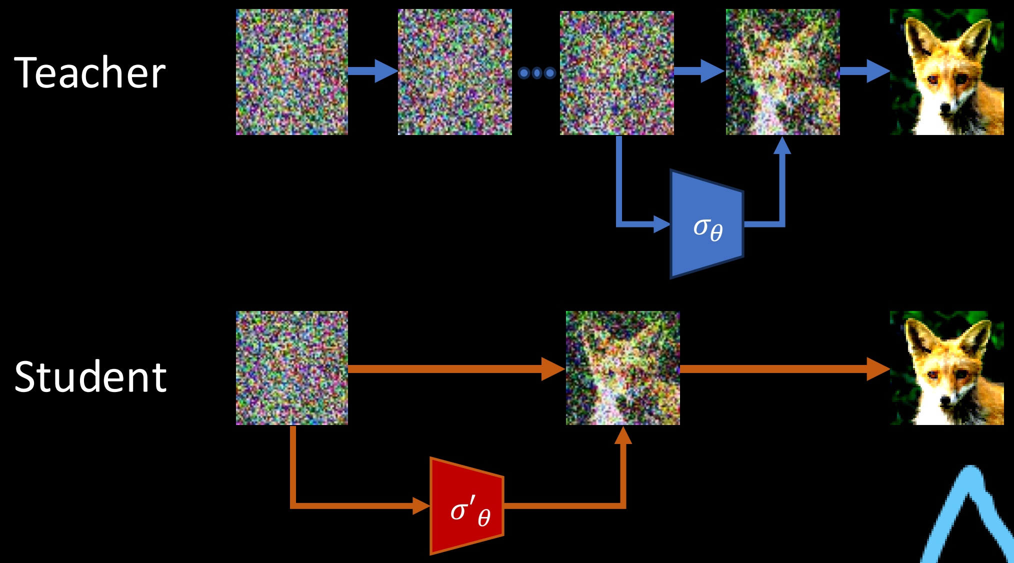
The denoiser is called multiple times

Goal

Fewer sampling steps while preserving quality


Goal
Denoiser Distillation
LD3

Denoiser Distillation



Pros: High quality in few/one step generation

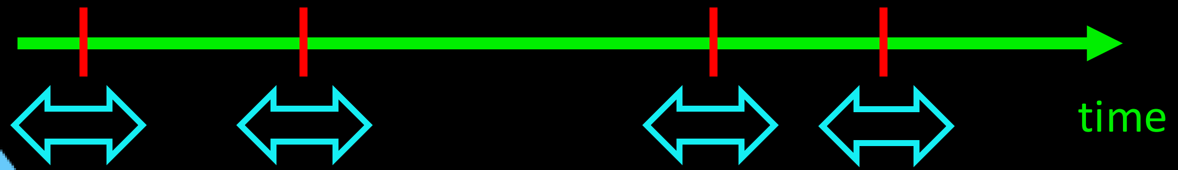
Cons: Costly training

Consistency Distillation training: 8 days (A100)

CIFAR10

LD3

Learning to Discretize Denoising Diffusion ODEs



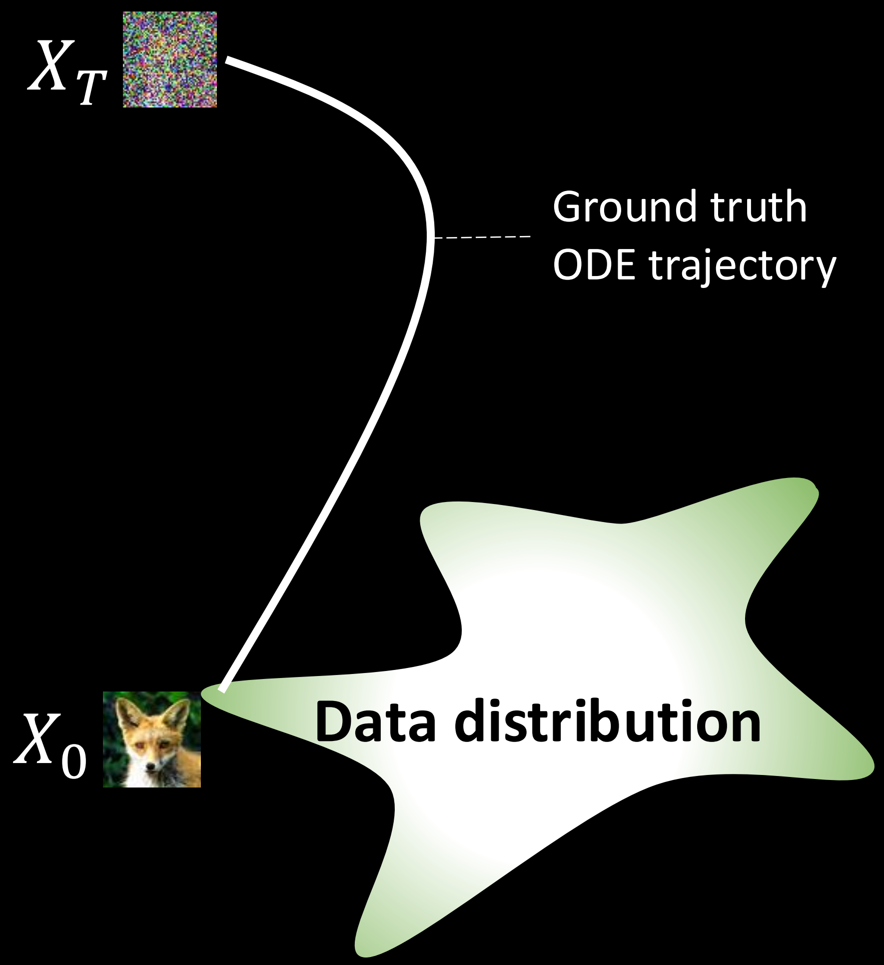
LD3 training: 5 minutes (A100)

CIFAR10

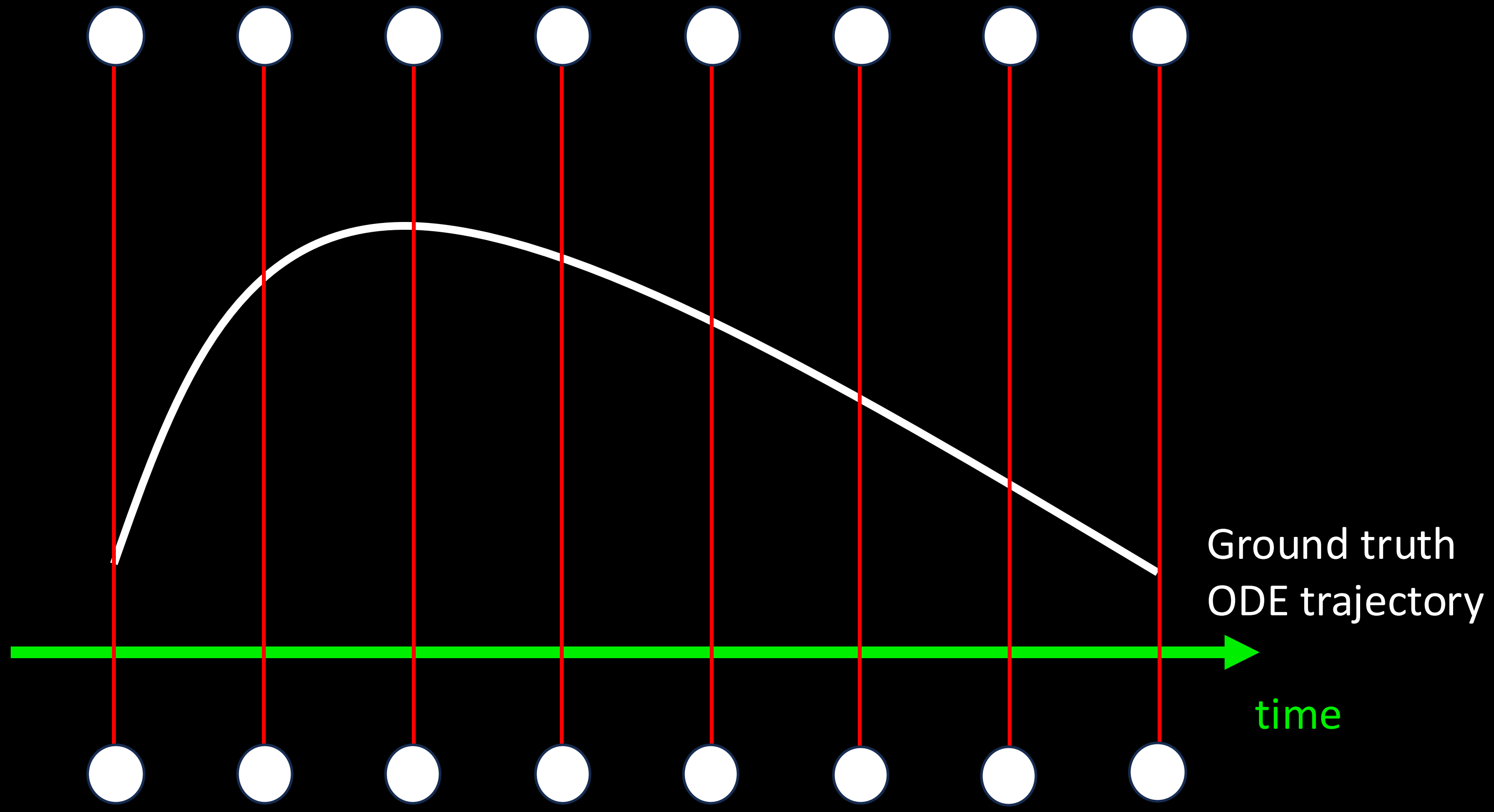
High quality few-step generation

Fast training

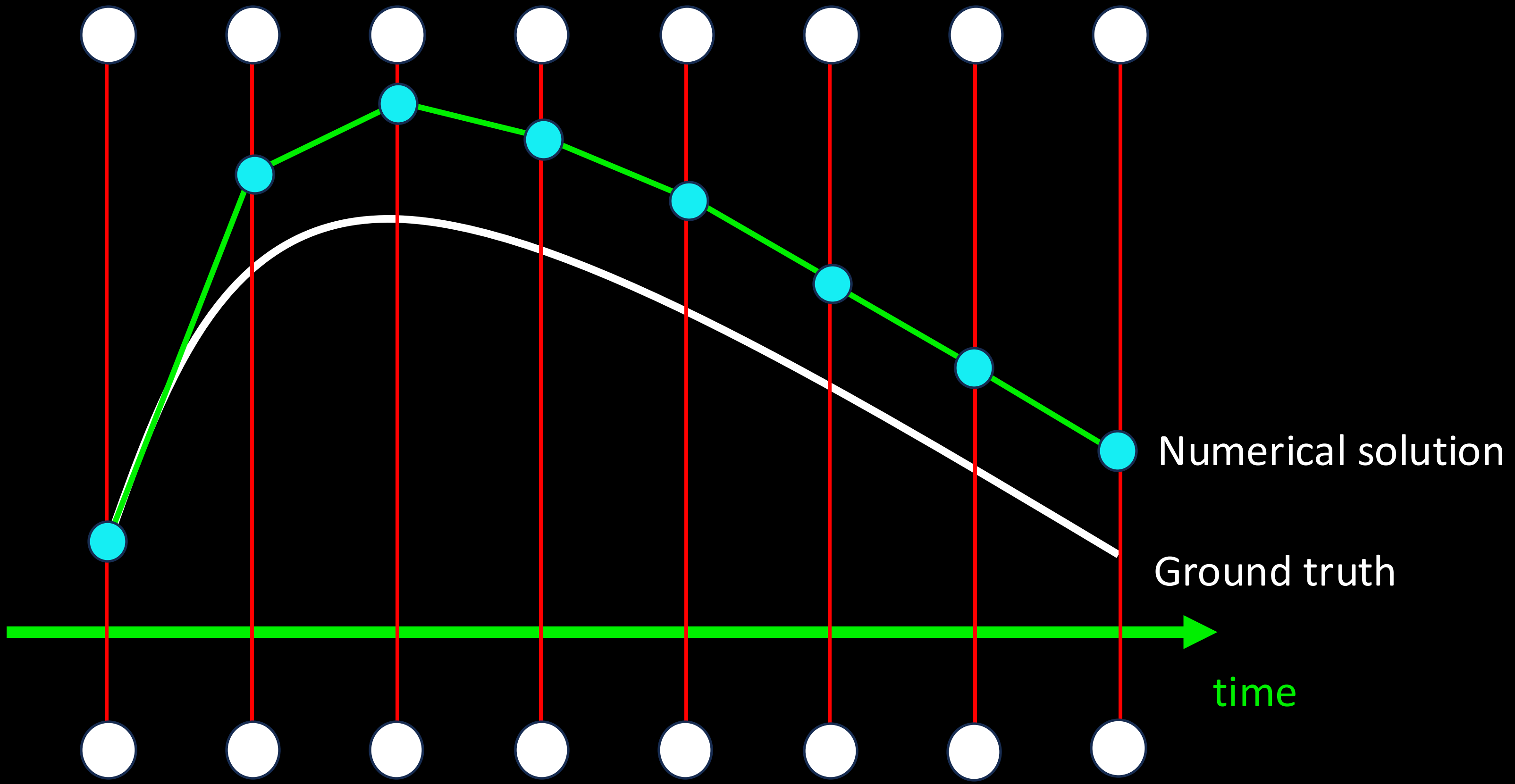
The Importance of Timesteps



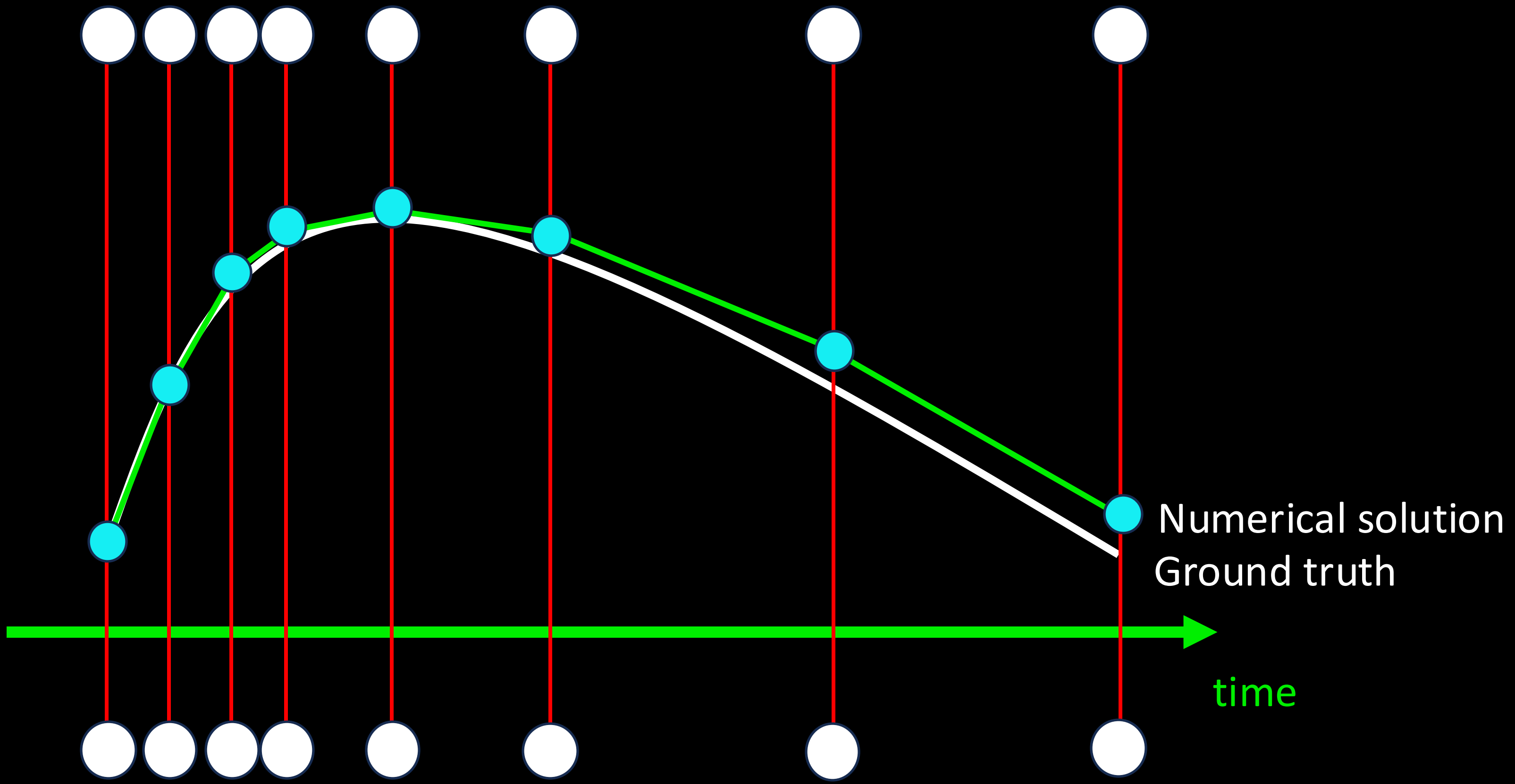
The Importance of Timesteps



The Importance of Timesteps



The Importance of Timesteps



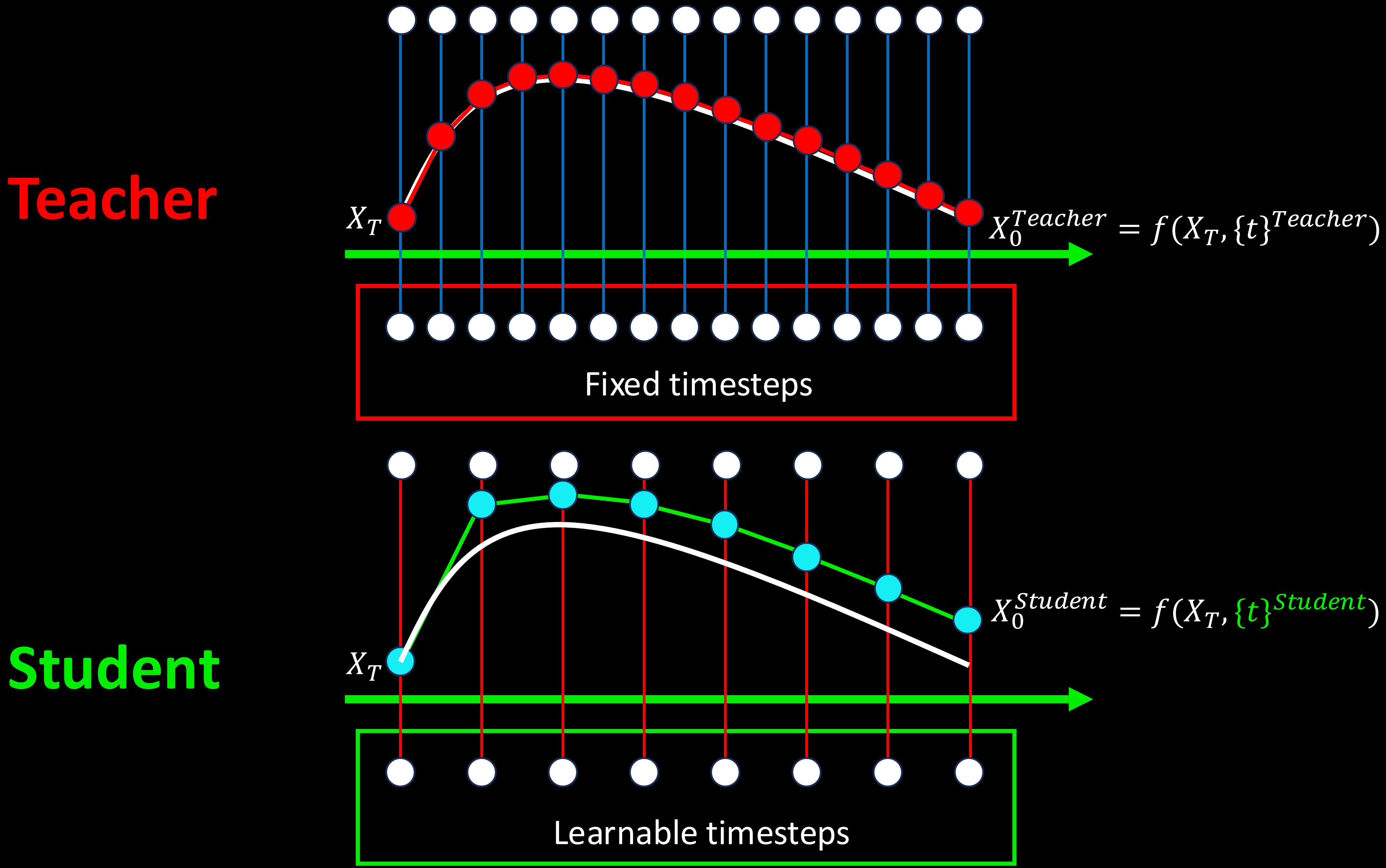
LD3 - Overview

LD3: A teacher–student approach

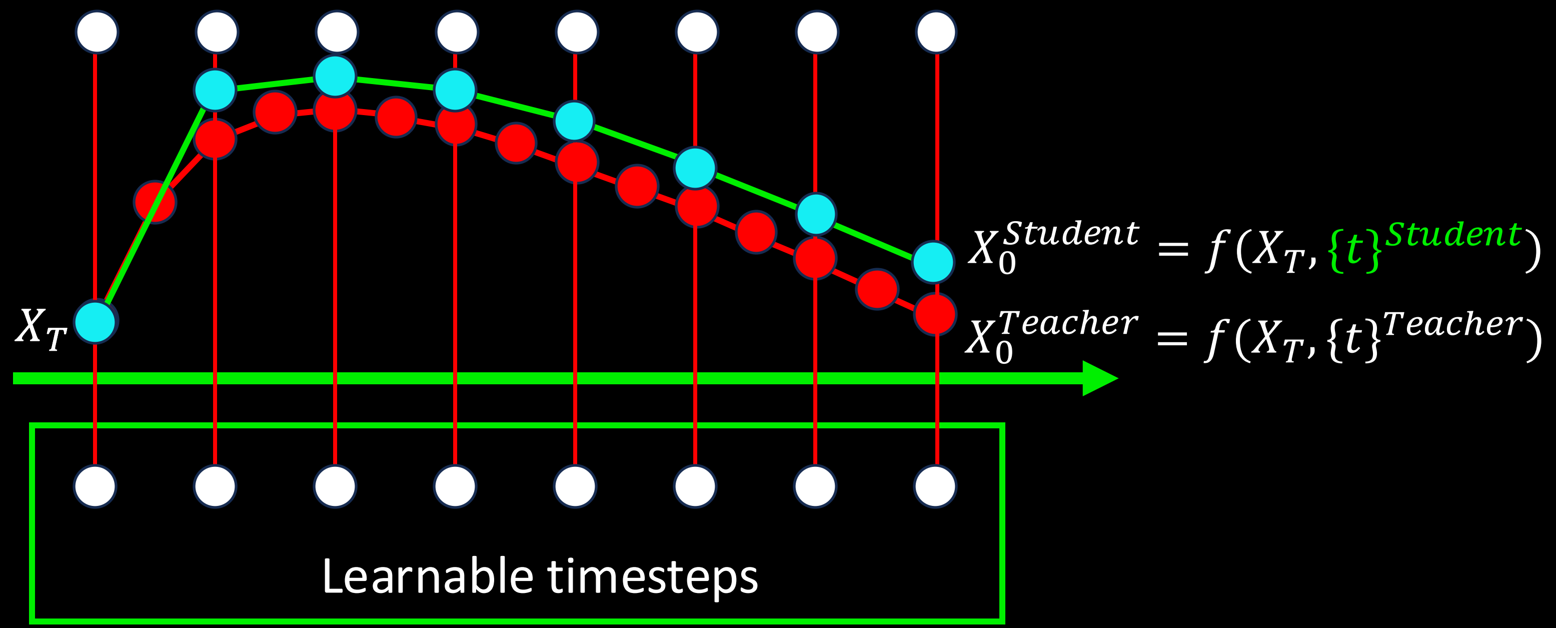
- Teacher: Many-step solver
- Student: Few-step solver

Use the same denoiser network!

- **Objective:** Learn discretization of **timesteps** to align student with teacher output


Teacher
X_T
$X_0^{Teacher} = f(X_T, \{t\}^{Teacher})$
Fixed timesteps
Student
X_T
$X_0^{Student} = f(X_T, \{t\}^{Student})$
Learnable timesteps

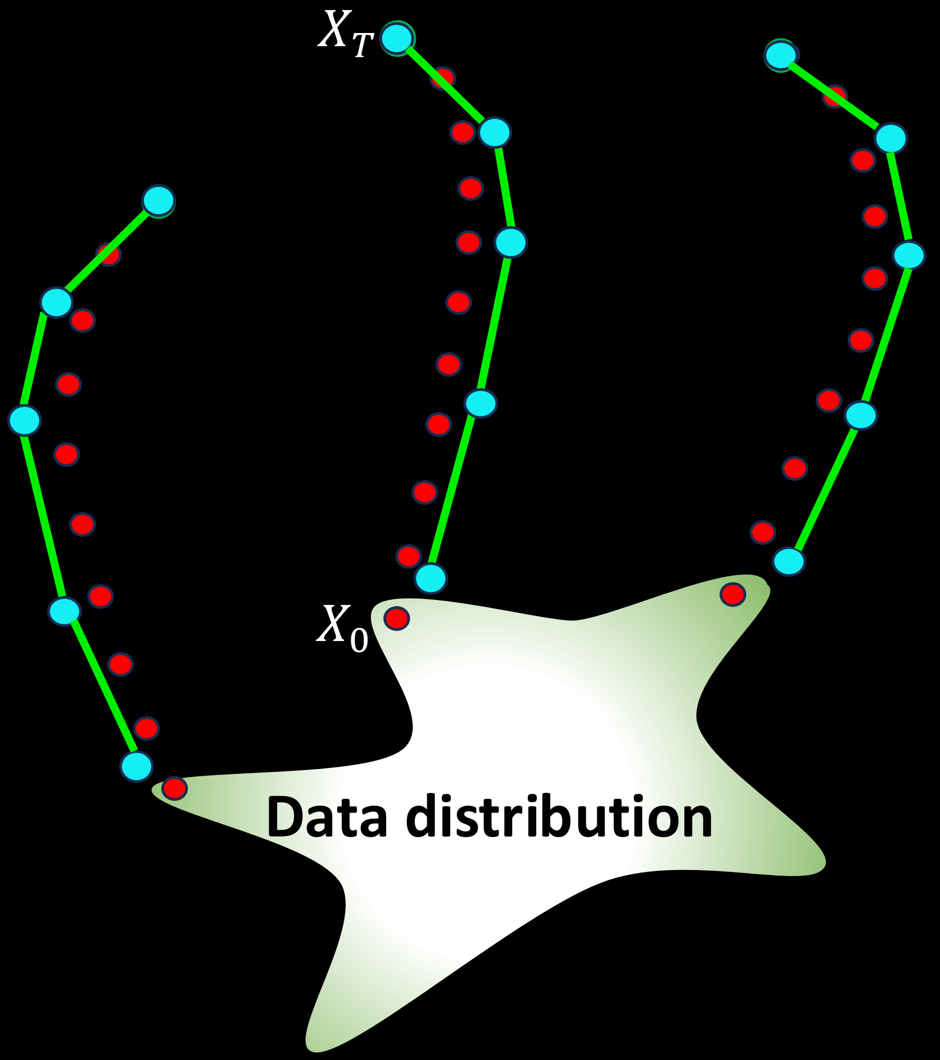
LD3 Minimizes Global Discretization Error



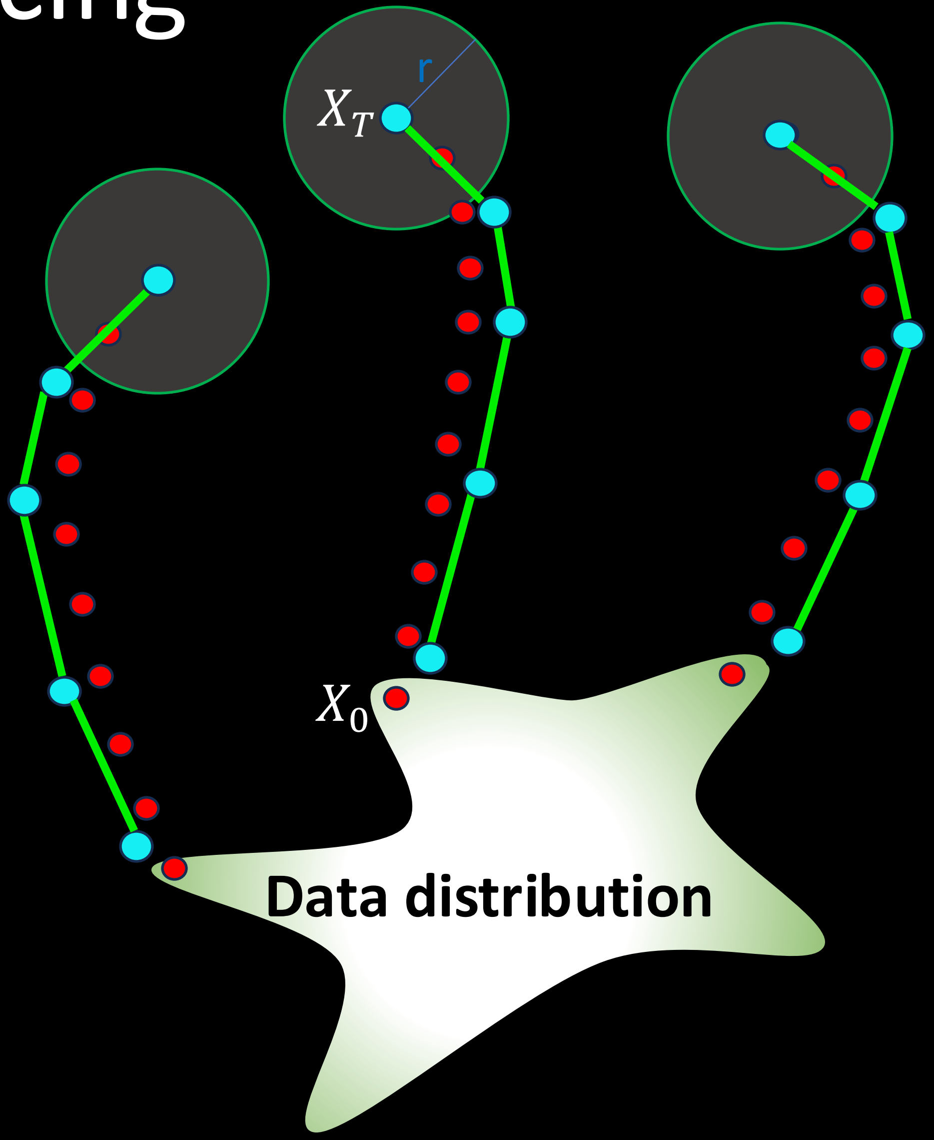
Objective: Align trajectory endpoints via backpropagating gradients for the timesteps through the student ODE solver

Underfitting risk: only a few parameters are optimized

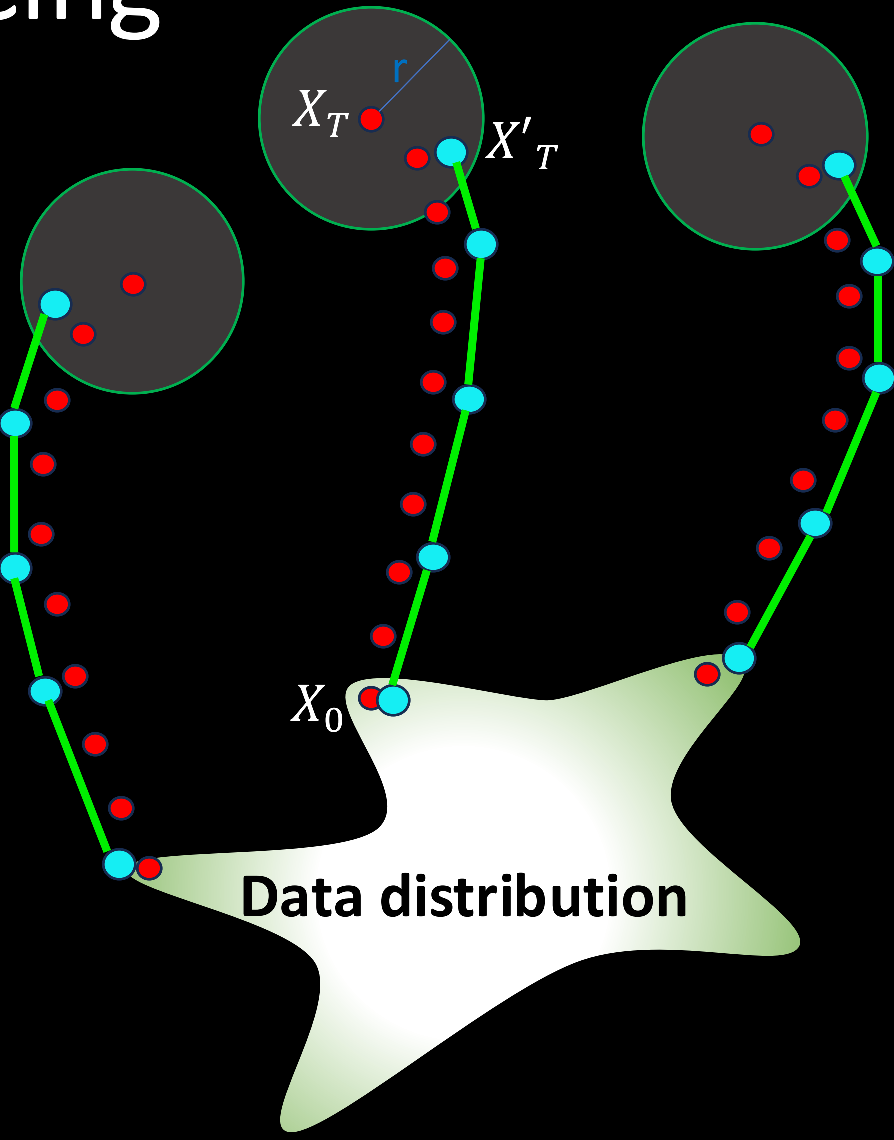
Hard Teacher Forcing



Soft Teacher Forcing



Soft Teacher Forcing



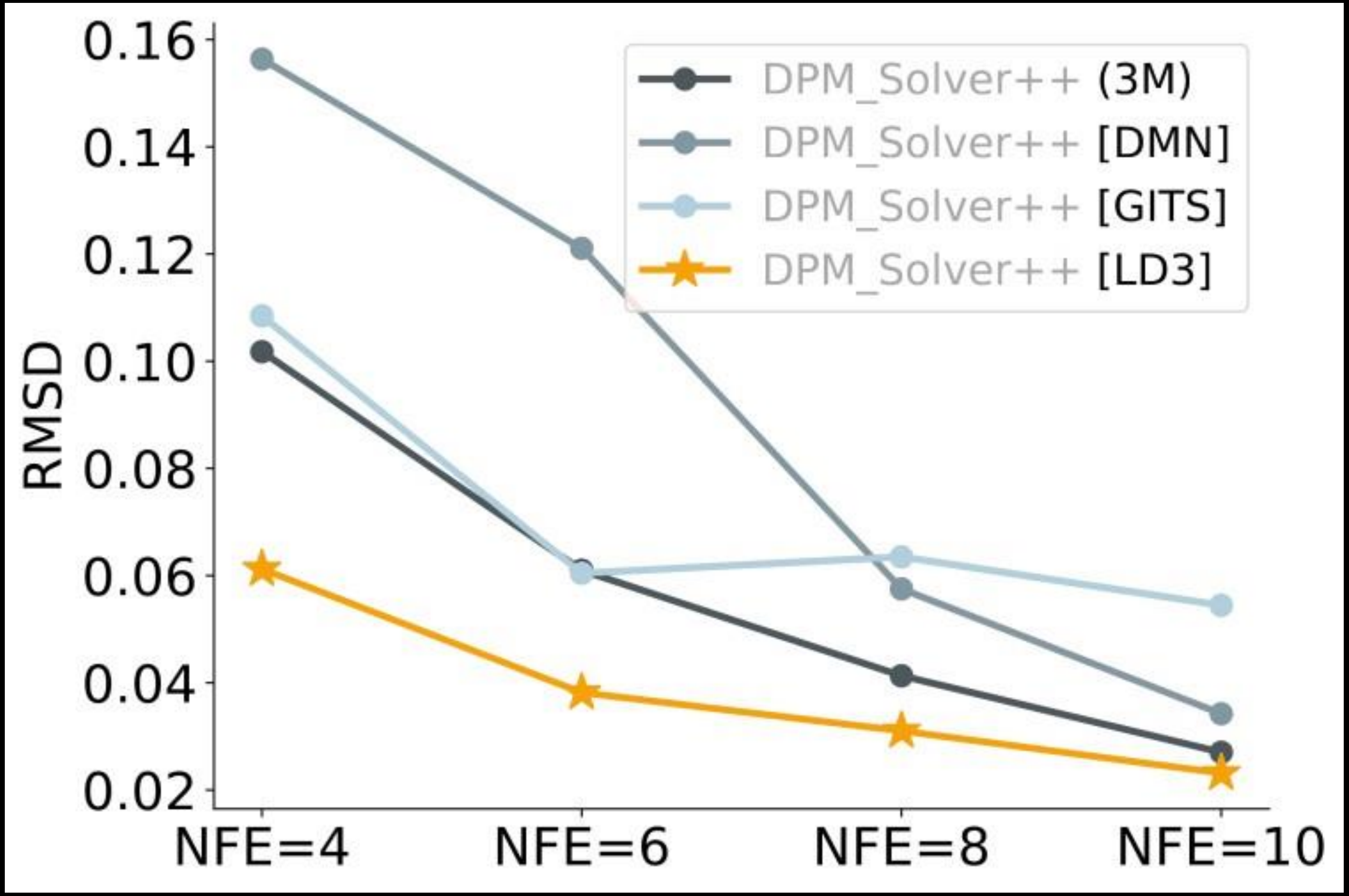
Experimental Results

Image Synthesis
FID score (↓ lower is better)

CIFAR10 (pixel space) – Uni_PC				
Method	*NFE=4*	*NFE=6*	*NFE=8*	*NFE=10*
3M	43.9	13.1	4.4	3.2
GITS	25.3	11.2	5.7	3.7
DMN	26.4	8.1	5.9	**2.5**
LD3	**13.7**	**5.9**	**3.4**	2.9

ImageNet-256 (latent space) – Uni_PC				
Method	*NFE=4*	*NFE=5*	*NFE=6*	*NFE=7*
3M	20.0	8.5	5.9	5.2
GITS	54.9	34.9	14.6	9.0
DMN	16.7	8.0	7.5	7.8
LD3	**9.9**	**5.0**	**4.5**	**4.3**

LD3 Is Closer to The Teacher Solver



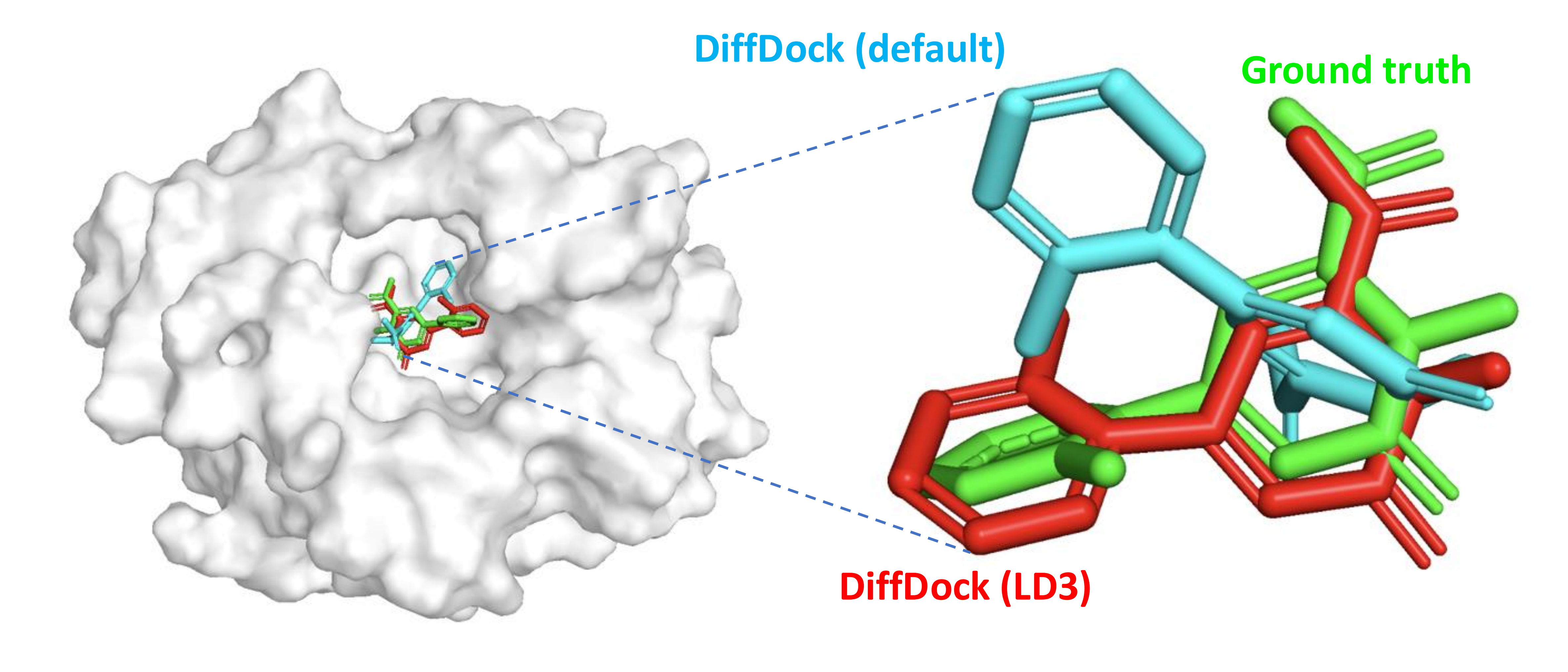

FLUX & LD3

“A biomechanical dragon, part machine, part organic, soaring above a dystopian wasteland, its wings made of pure energy.”

Default discretization (30 steps)

LD3 optimized discretization (30 steps)

Beyond Image Generation



Conclusion

LD3: A lightweight framework

- Reduces computational cost in sampling from pre-trained DPMs
- Learns time discretization for ODE-based sampling
- Minimal training overhead
- High-quality output with few-step sampling

Thanks to the Collaborators

Dung Hoang

Anji Liu

Guy Van den Broeck

Mathias Niepert

University of Stuttgart
Germany

imprs-is

Ucla

UNIVERSITÄT
BERN