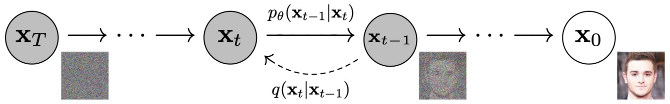
# Beyond Diffusion: Efficient Generative Modeling with Flow Matching

Jia-Wei Liao

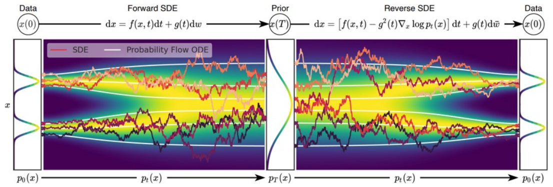
Ph.D. Candidate in Computer Science

National Taiwan University

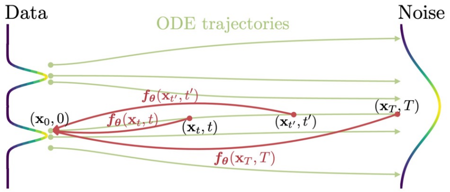
# Evolution of Diffusion Models



**DDPM [NeurIPS'20]**
2020 / 6 (UCB)



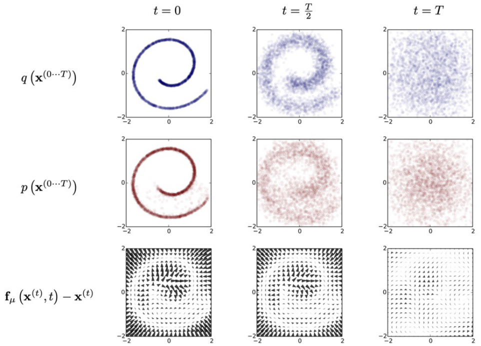

**Score-based SDE [ICLR'21]**
2020 / 11 (Ermon Group)



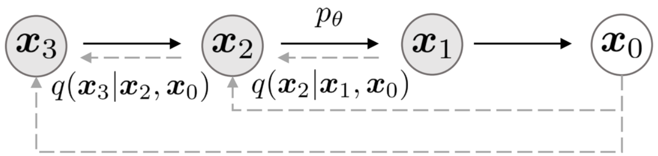

**CM [ICML'23]**
2023 / 3 (OpenAI)

**Flow Matching**

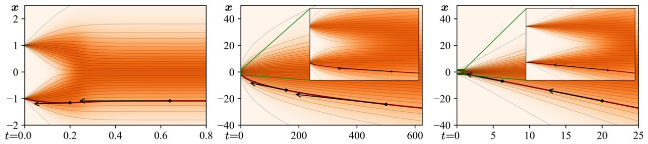
**DPM [ICML'15]**
2015 / 3 (Stanford)



**DDIM [ICLR'21]**
2020 / 10 (Ermon)

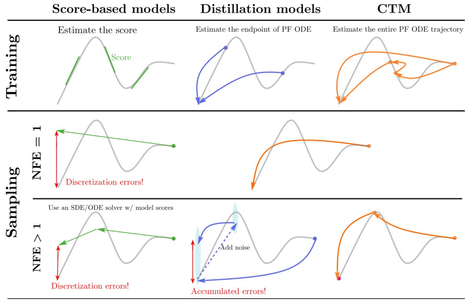


**EDM [NeurIPS'22]**
2022 / 6 (NVIDIA)

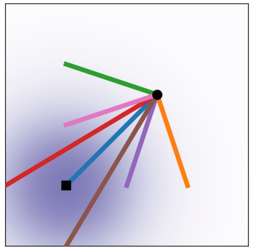


(a) Variance preserving ODE [49] (b) Variance exploding ODE [49] (c) DDIM [47] / Our ODE

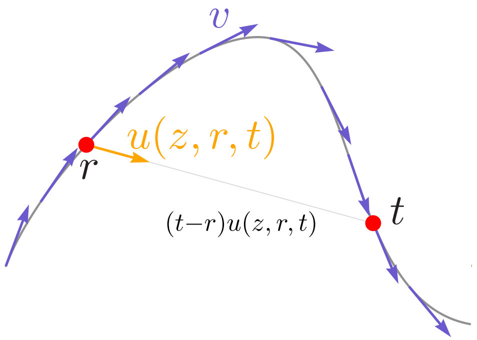
**CTM [ICLR'24]**
2023 / 10 (Sony AI)



# From Diffusion Model to Flow Matching

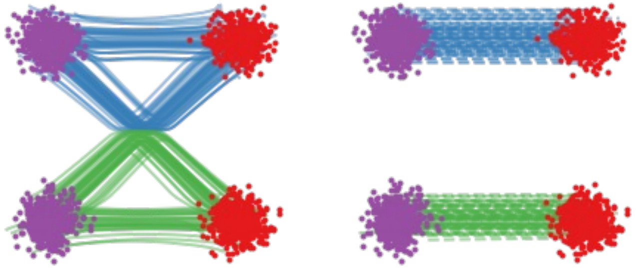


**FM [ICLR’23]**
2022 / 10 (Meta)

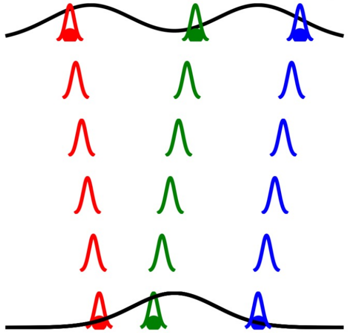



**MeanFlow [NeurIPS’25]**
2025 / 5 (Kaiming Group)

**RectFlow [ICLR’23]**
2022 / 9 (UT-Austin)



**Batch-OT FM [TMLR’24]**
2023 / 2 (UdeM)

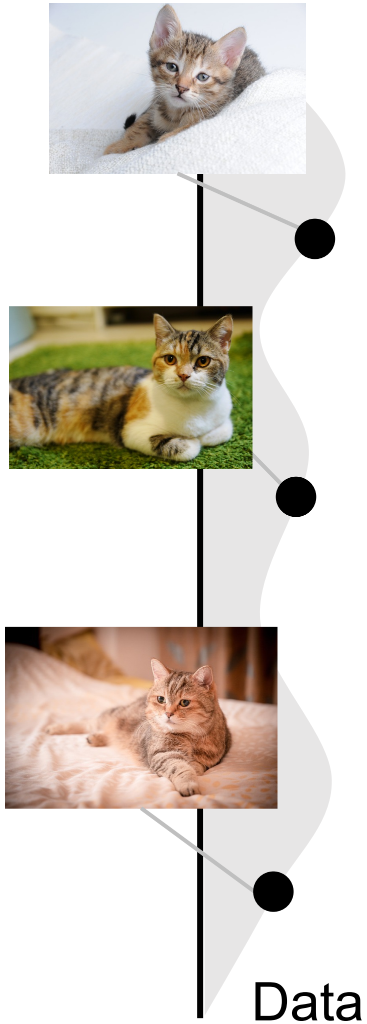


# What Will We Cover Today?

- Flow Matching
- Rectified Flow
- Batch-OT Flow Matching
- MeanFlow

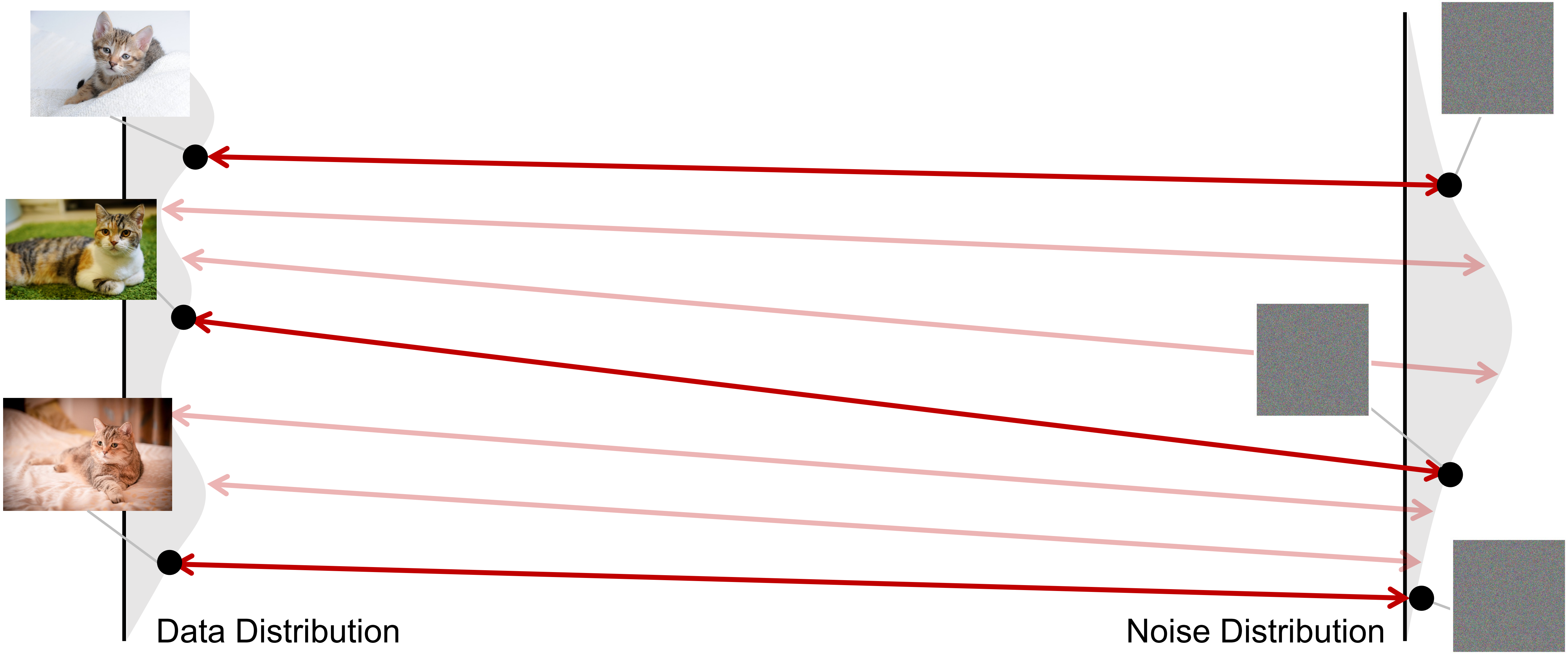
# What is Generative Model Learning?

# What is Generative Model Learning?

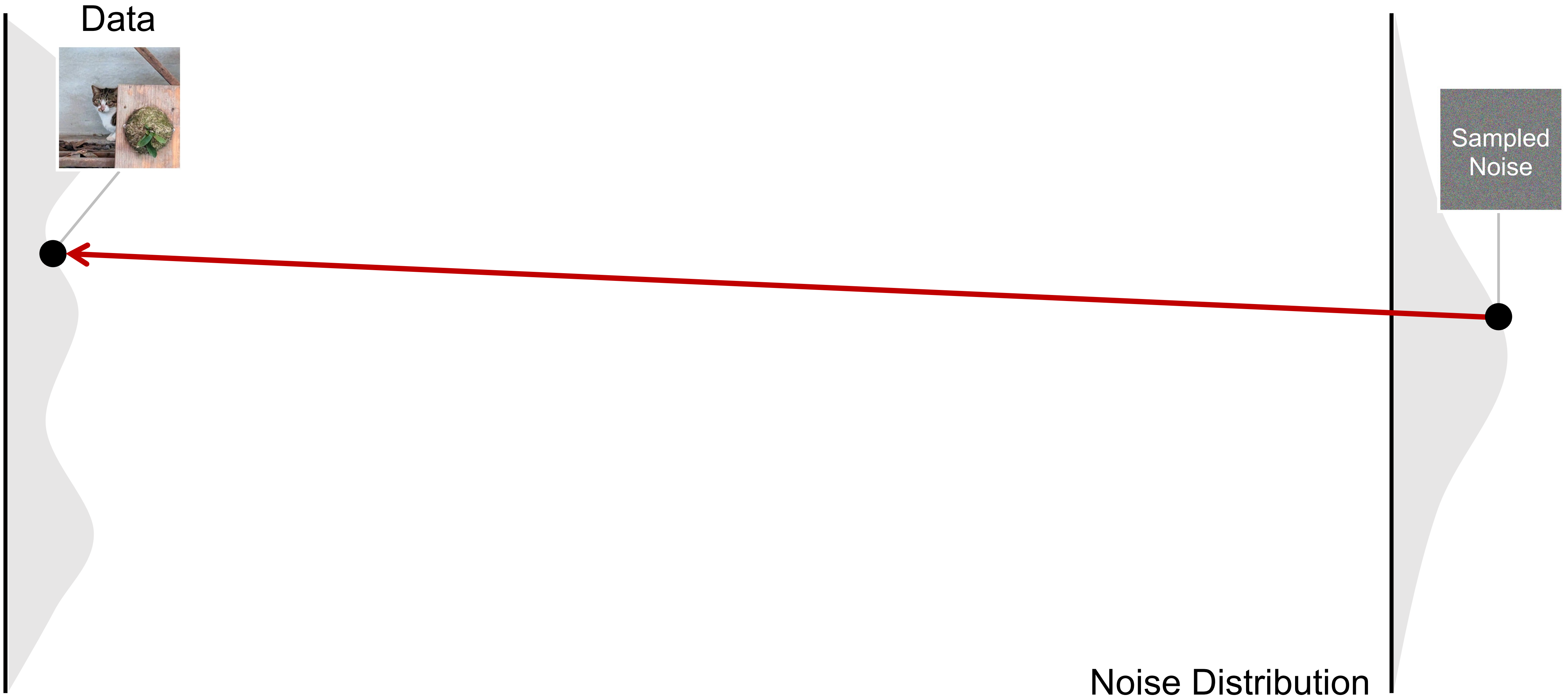


Data Distribution

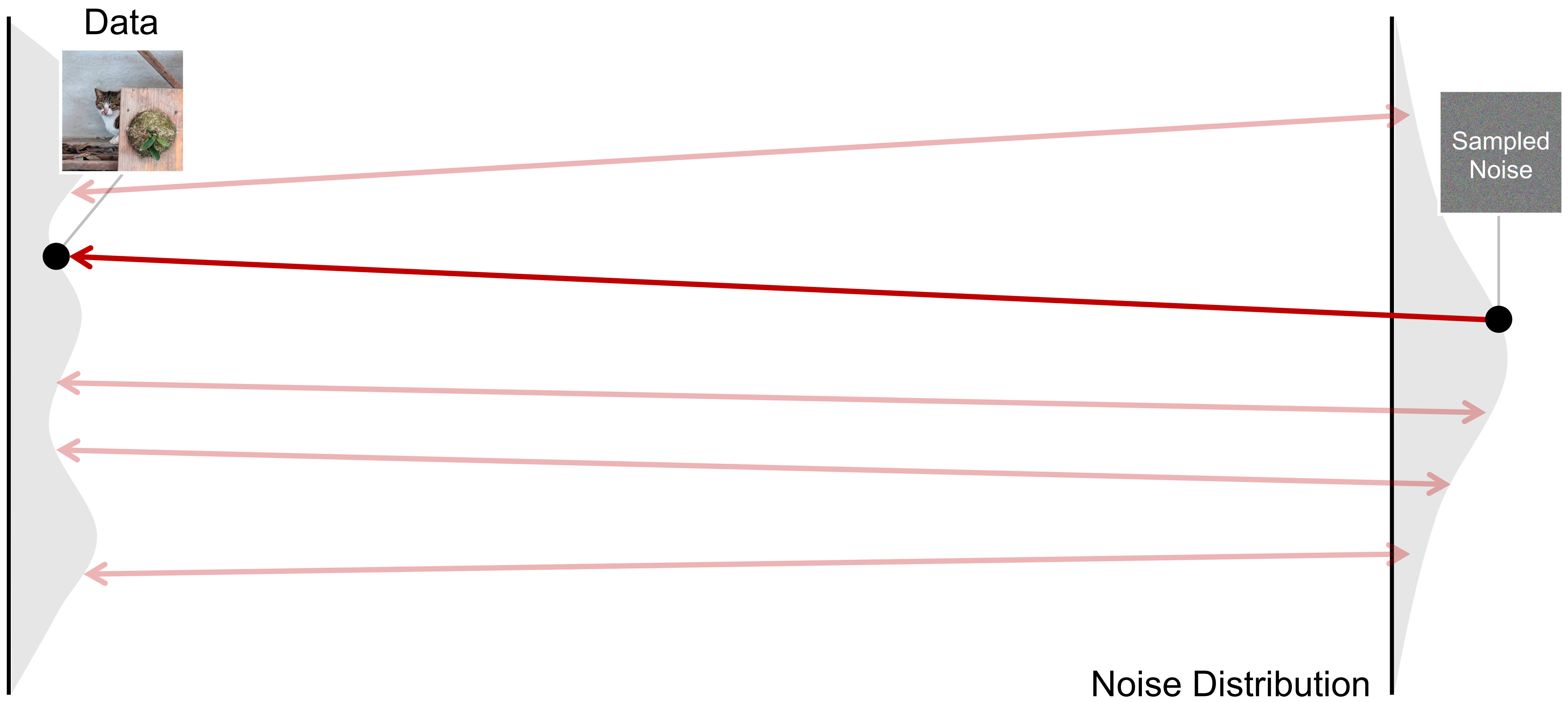
# What is Generative Model Learning?



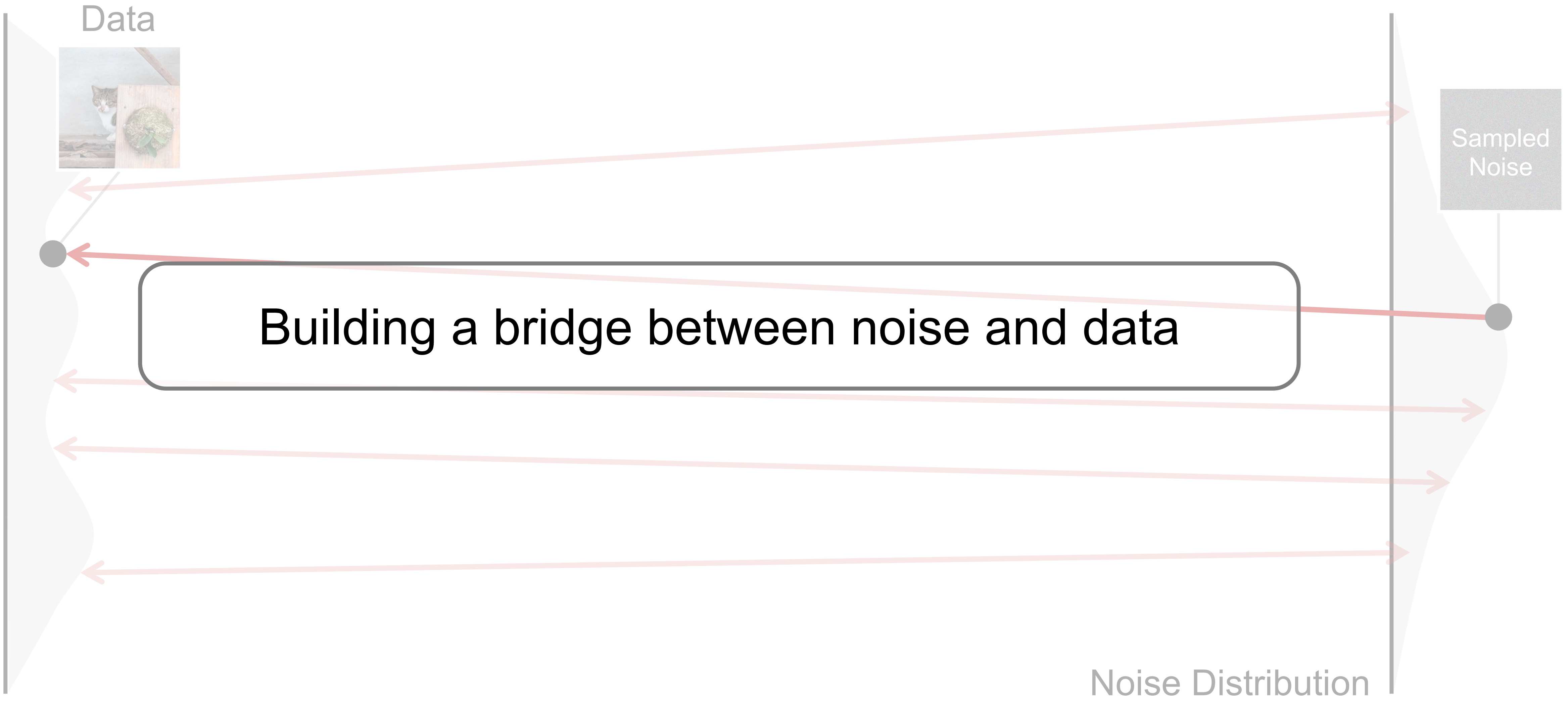
# What is Generative Model Learning?



# The Goal of Generative Model



# The Goal of Generative Model



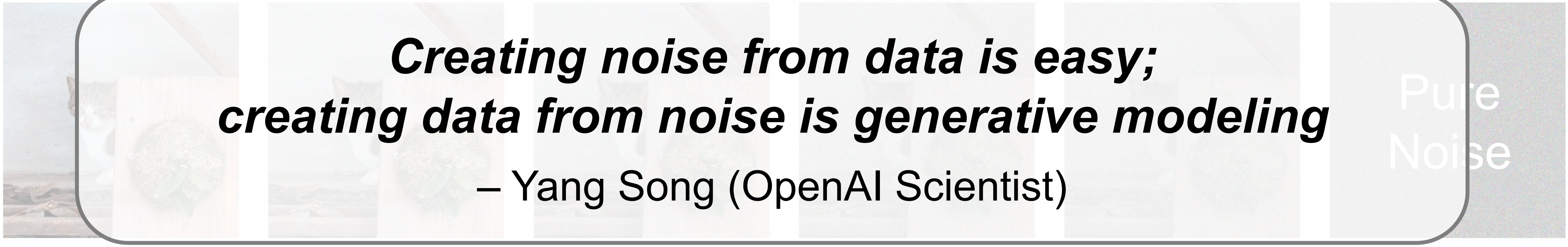
# What is Diffusion Model?

**Forward Process:** add noise step by step, from data to pure noise


**Reverse Process:** generate data from pure noise by denoising

# What is Diffusion Model?

**Forward Process:** add noise step by step, from data to pure noise

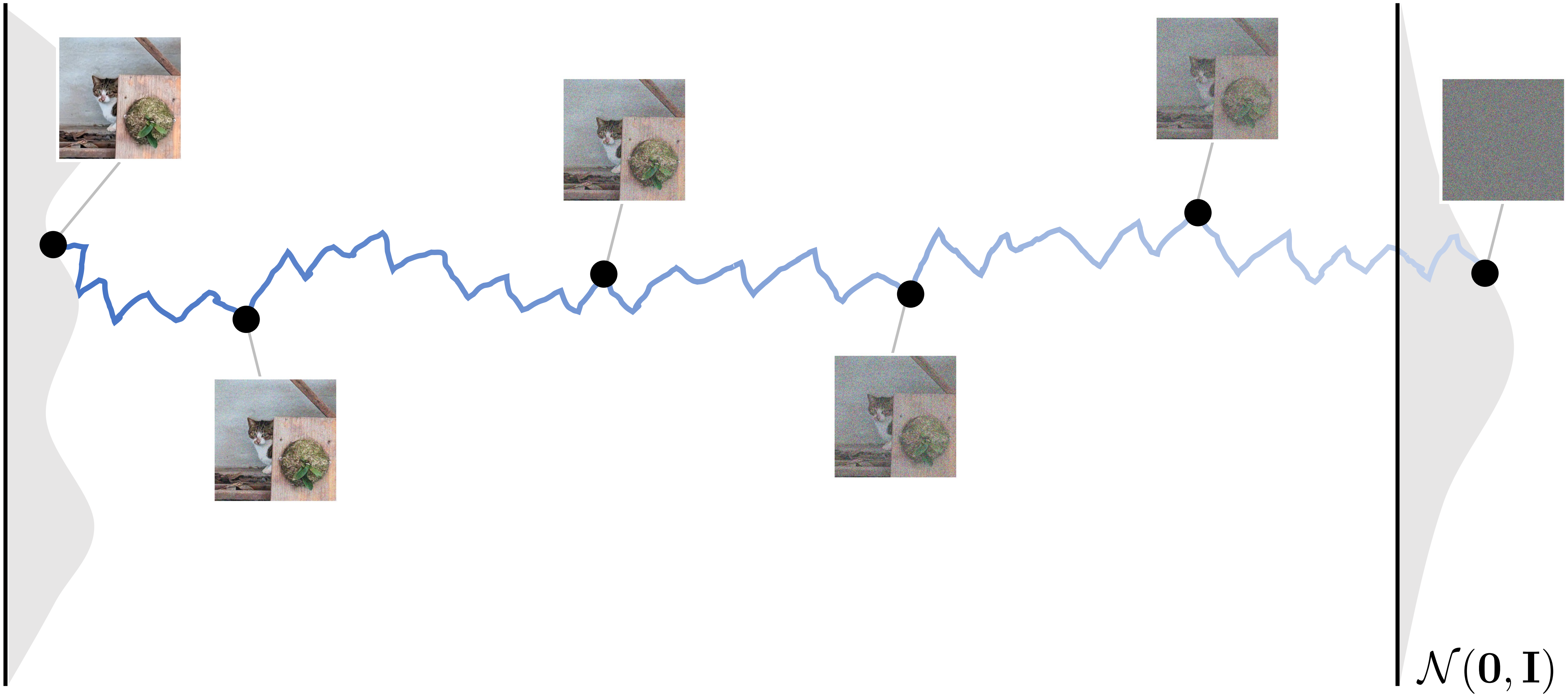



***Creating noise from data is easy;***
***creating data from noise is generative modeling***

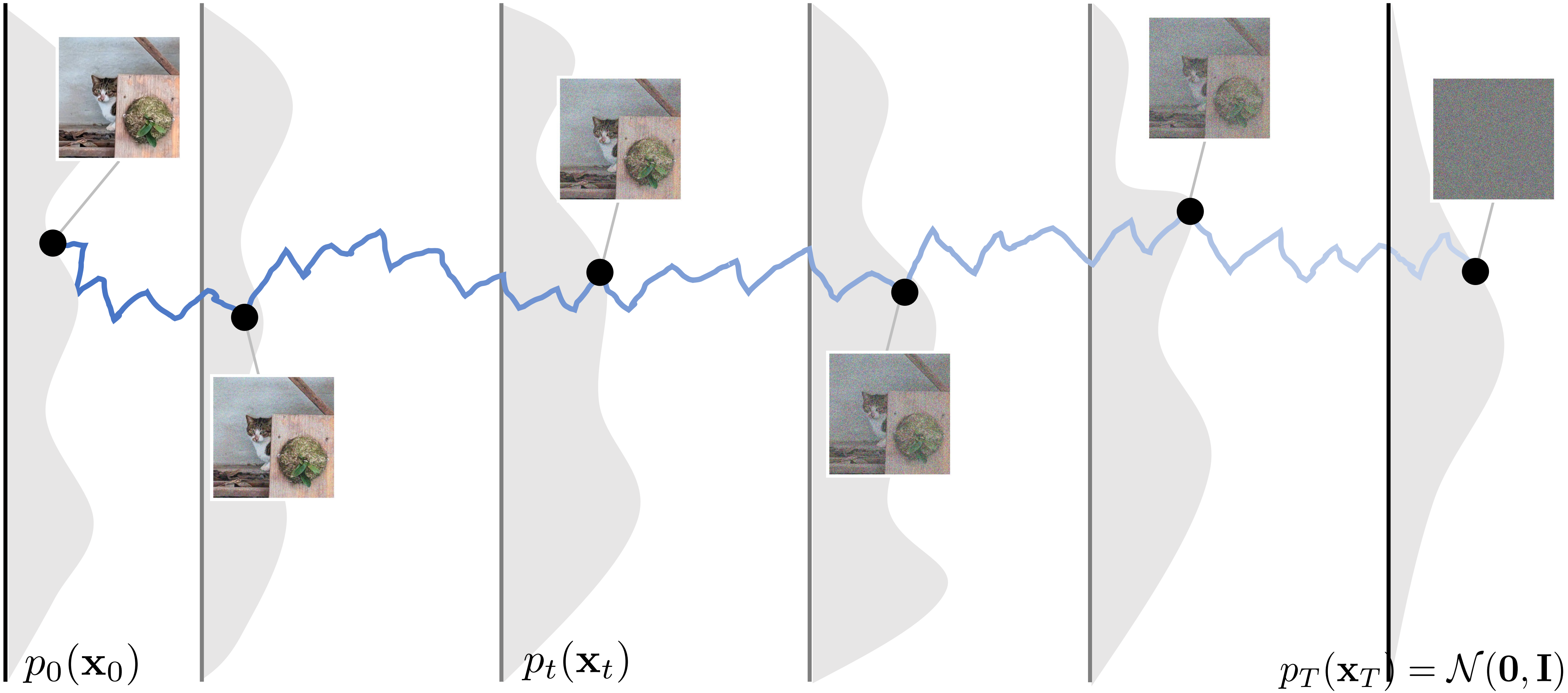
– Yang Song (OpenAI Scientist)

**Reverse Process:** generate data from pure noise by denoising

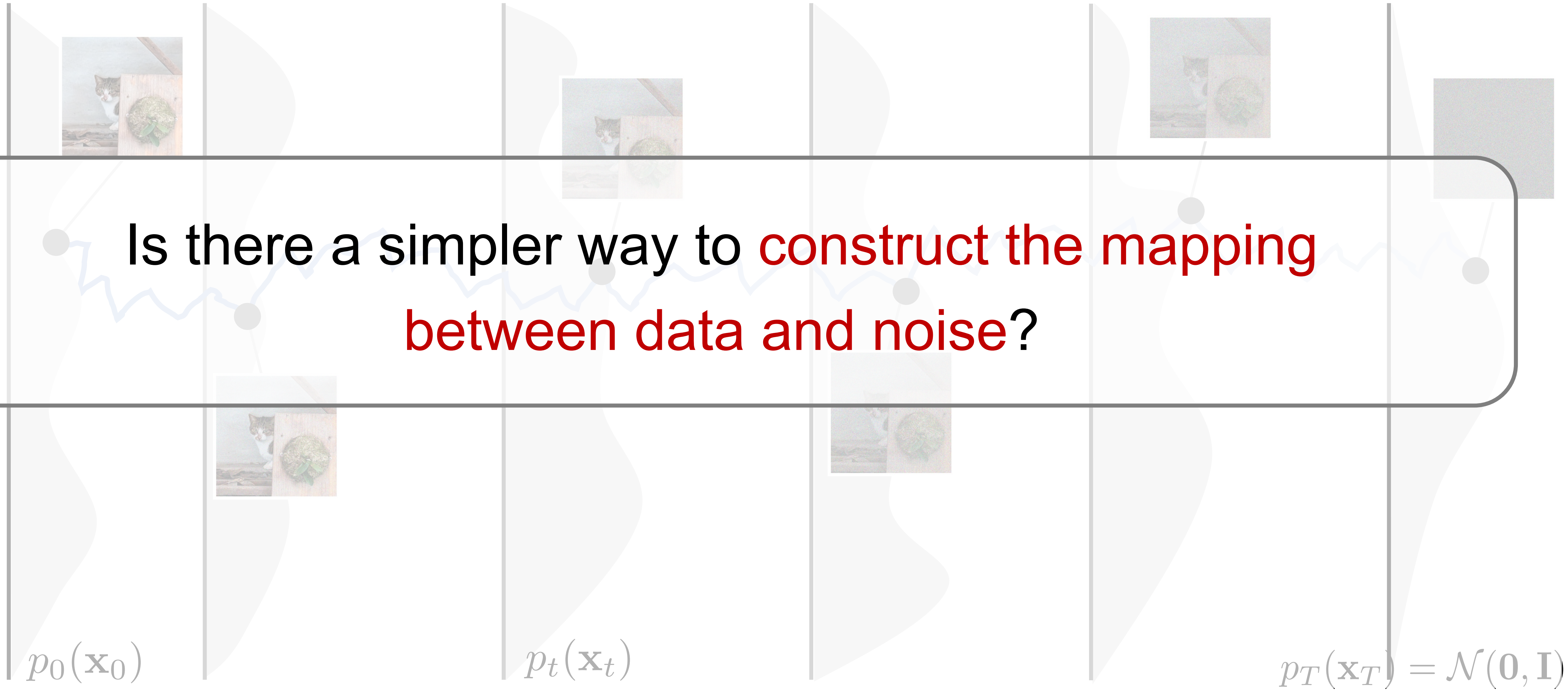
# Diffusion Models



# Diffusion Models



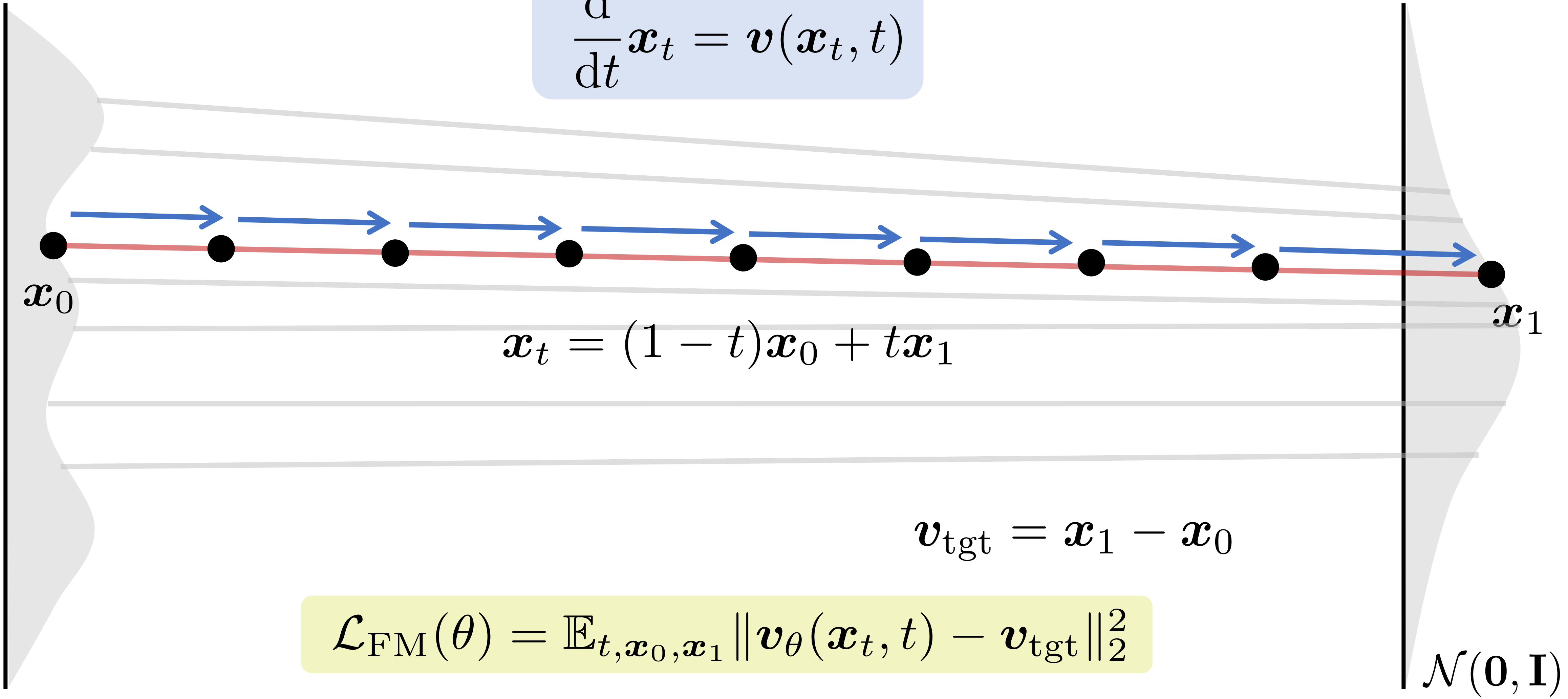
# Diffusion Models




Is there a simpler way to construct the mapping between data and noise?

# Flow Matching [Lipman+ ICLR’23]

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}_t = \boldsymbol{v}(\boldsymbol{x}_t, t)$$



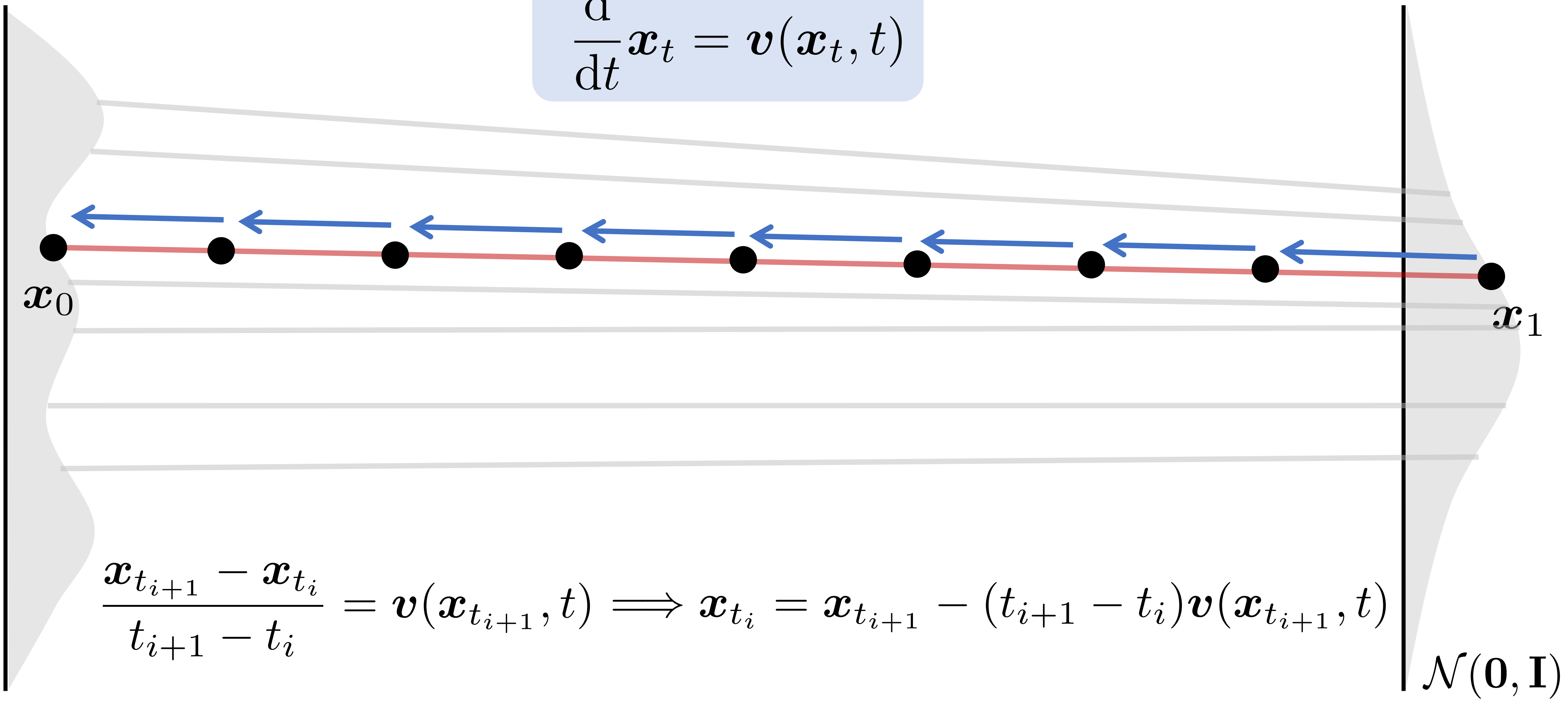

$$\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$$

$$\boldsymbol{v}_{\mathrm{tgt}} = \boldsymbol{x}_1 - \boldsymbol{x}_0$$

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{x}_1} \|\boldsymbol{v}_\theta(\boldsymbol{x}_t, t) - \boldsymbol{v}_{\mathrm{tgt}}\|_2^2$$

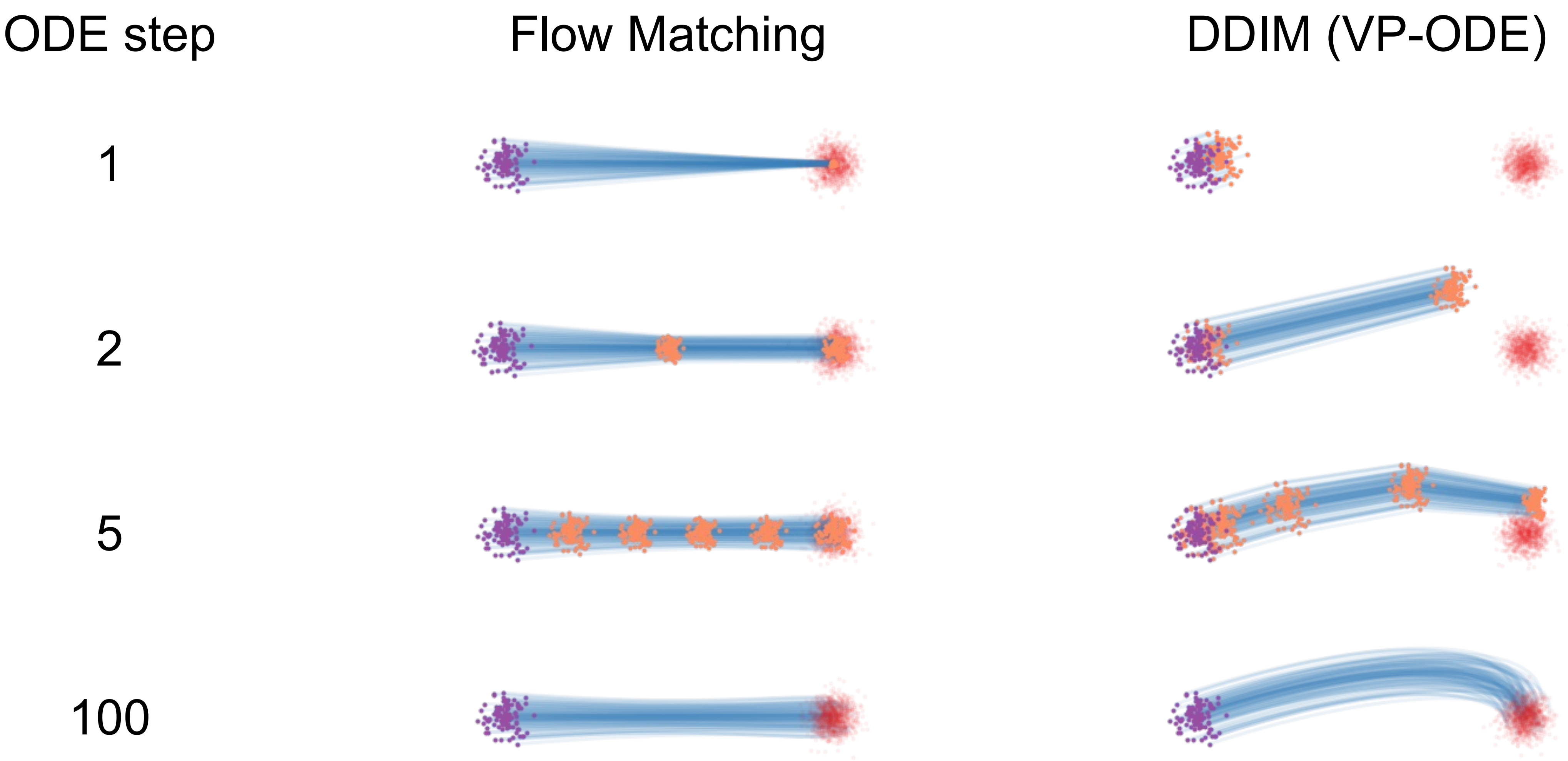
# Flow Matching: Sampling

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}_t = \boldsymbol{v}(\boldsymbol{x}_t, t)$$

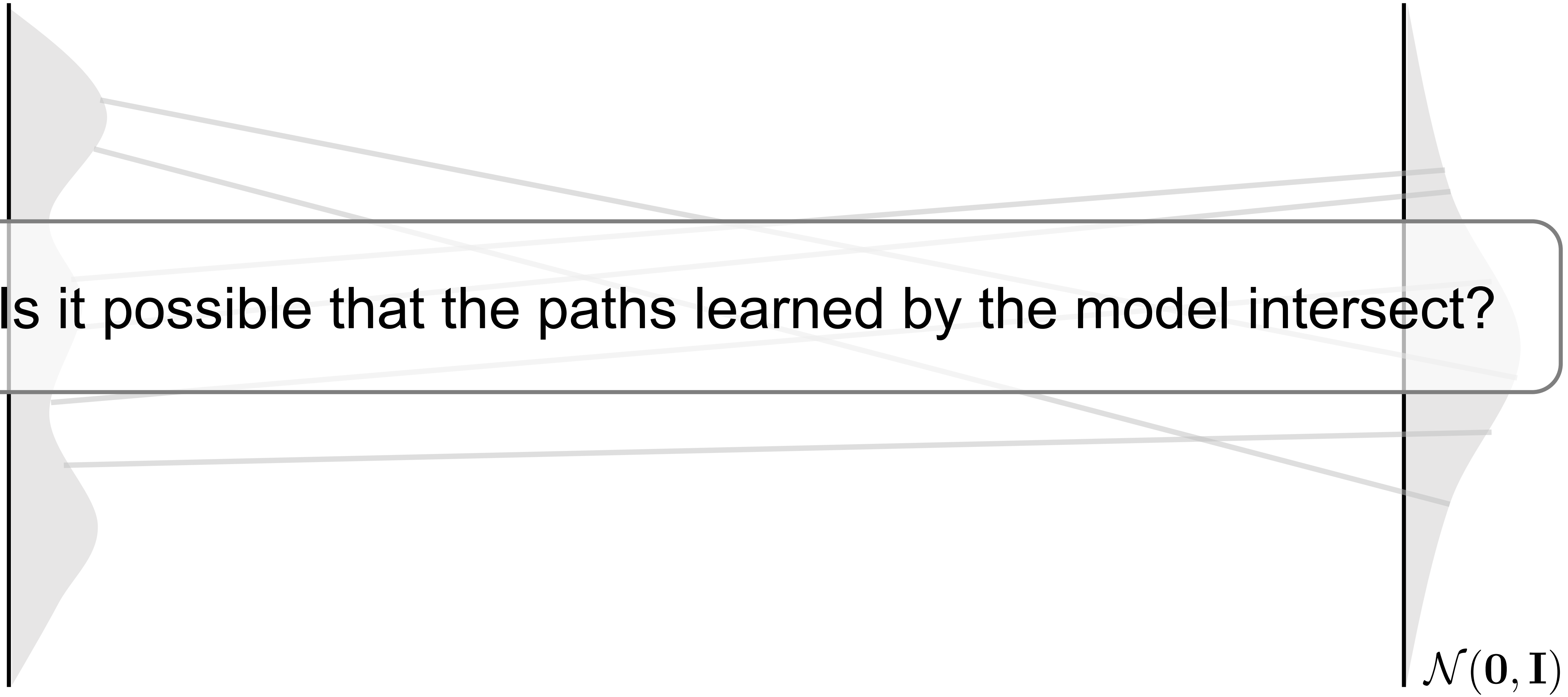



$$\frac{\boldsymbol{x}_{t_{i+1}} - \boldsymbol{x}_{t_i}}{t_{i+1} - t_i} = \boldsymbol{v}(\boldsymbol{x}_{t_{i+1}}, t) \Longrightarrow \boldsymbol{x}_{t_i} = \boldsymbol{x}_{t_{i+1}} - (t_{i+1} - t_i)\boldsymbol{v}(\boldsymbol{x}_{t_{i+1}}, t)$$

# Flow Matching vs DDIM

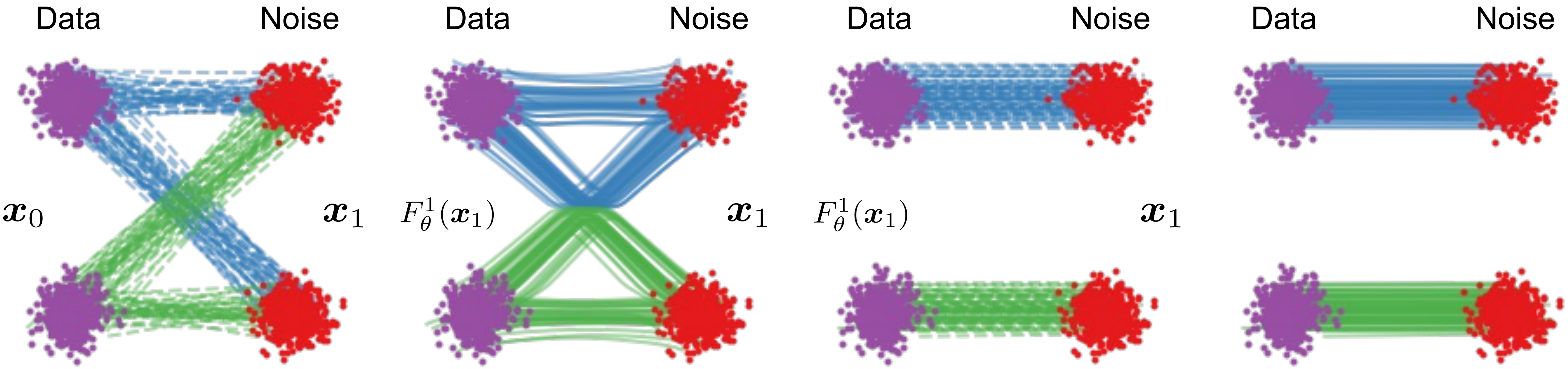


# Flow Matching




Is it possible that the paths learned by the model intersect?

# *k*-Rectified Flow [Liu+ ICLR'23]




$\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$

(Random coupling)

Flow Matching $F_\theta^1$

(1-Rectified Flow)

Not straight but non-crossing

$\boldsymbol{x}_t^{(1)} = (1-t)F_\theta^1(\boldsymbol{x}_1) + t\boldsymbol{x}_1$

(Generate new coupling from 1-Rectified Flow)

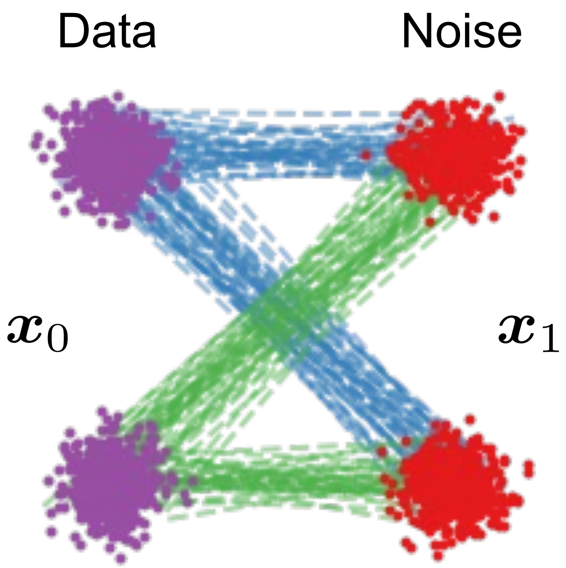
2-Rectified Flow

Straight

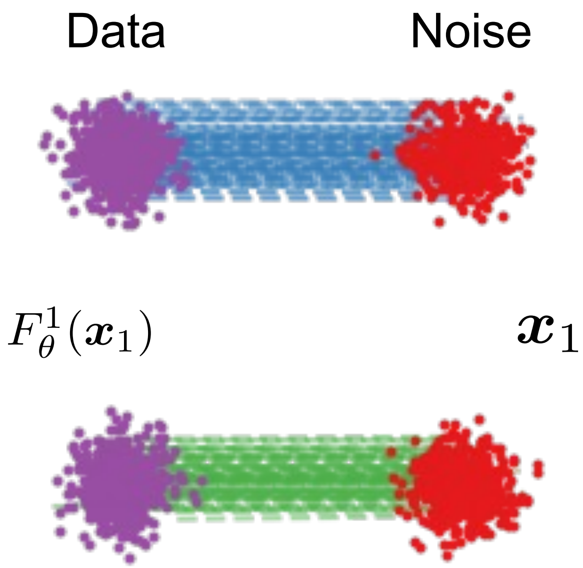
# *k*-Rectified Flow

A reflow step reduces the distance between data and noise coupling

$$\mathbb{E}[\|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2^2] \geq \mathbb{E}[\|F_\theta^1(\boldsymbol{x}_1) - \boldsymbol{x}_1\|_2^2] \geq \mathbb{E}[\|F_\theta^2(\boldsymbol{x}_1) - \boldsymbol{x}_1\|_2^2] \geq \cdots$$

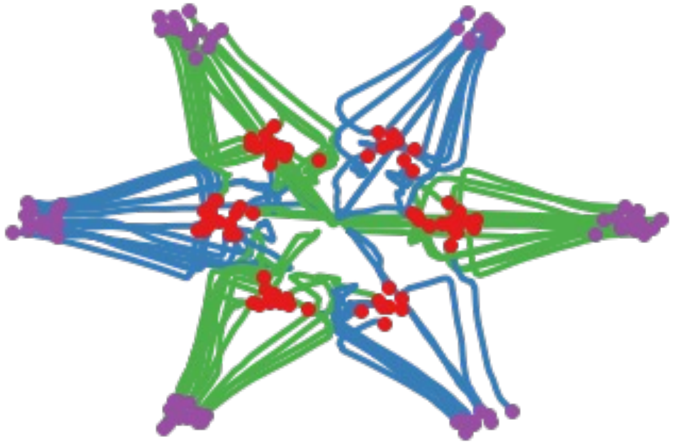



Random Coupling

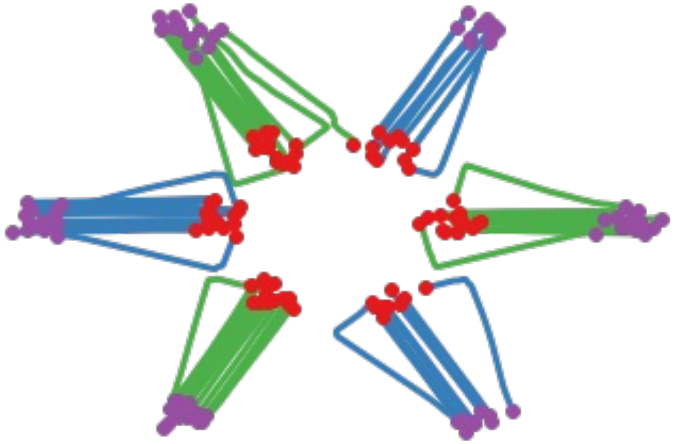



Generate New Coupling
from 1-Rectified Flow

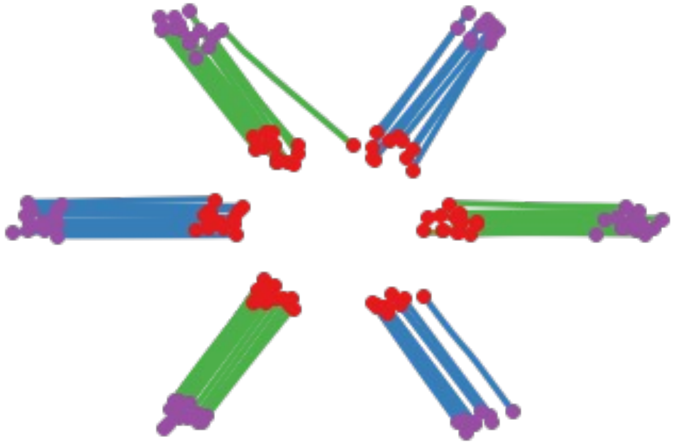
# *k*-Rectified Flow



1-Rectified Flow



2-Rectified Flow



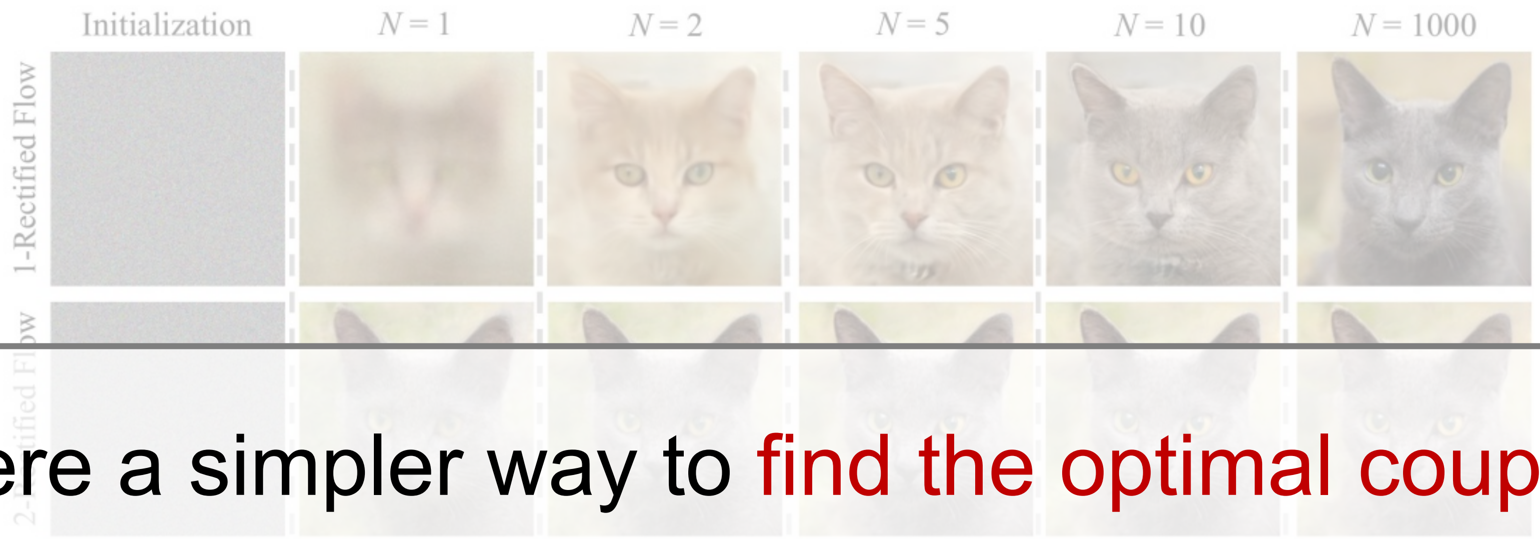
3-Rectified Flow

# Experiments

Result on CIFAR10

| Method | NFE(↓) | IS (↑) | FID (↓) | Recall (↑) |
|---|---|---|---|---|
| *ODE* | *One-Step Generation (Euler solver, N=1)* | | | |
| **1-Rectified Flow (+*Distill*)** | 1 | 1.13 (***9.08***) | 378 (*6.18*) | 0.0 (*0.45*) |
| **2-Rectified Flow (+*Distill*)** | 1 | 8.08 (*9.01*) | 12.21 (***4.85***) | 0.34 (*0.50*) |
| **3-Rectified Flow (+*Distill*)** | 1 | 8.47 (*8.79*) | 8.15 (*5.21*) | 0.41 (***0.51***) |
| VP ODE [73] (+*Distill*) | 1 | 1.20 (*8.73*) | 451 (*16.23*) | 0.0 (*0.29*) |
| sub-VP ODE [73] (+*Distill*) | 1 | 1.21 (*8.80*) | 451 (*14.32*) | 0.0 (*0.35*) |
| *ODE* | *Full Simulation (Runge–Kutta (RK45), Adaptive N)* | | | |
| **1-Rectified Flow** | 127 | **9.60** | **2.58** | **0.57** |
| **2-Rectified Flow** | 110 | 9.24 | 3.36 | 0.54 |
| **3-Rectified Flow** | 104 | 9.01 | 3.96 | 0.53 |
| VP ODE [73] | 140 | 9.37 | 3.93 | 0.51 |
| sub-VP ODE [73] | 146 | 9.46 | 3.16 | 0.55 |
| *SDE* | *Full Simulation (Euler solver, N=2000)* | | | |
| VP SDE [73] | 2000 | 9.58 | 2.55 | 0.58 |
| sub-VP SDE [73] | 2000 | 9.56 | 2.61 | 0.58 |

# *k*-Rectified Flow




Noise

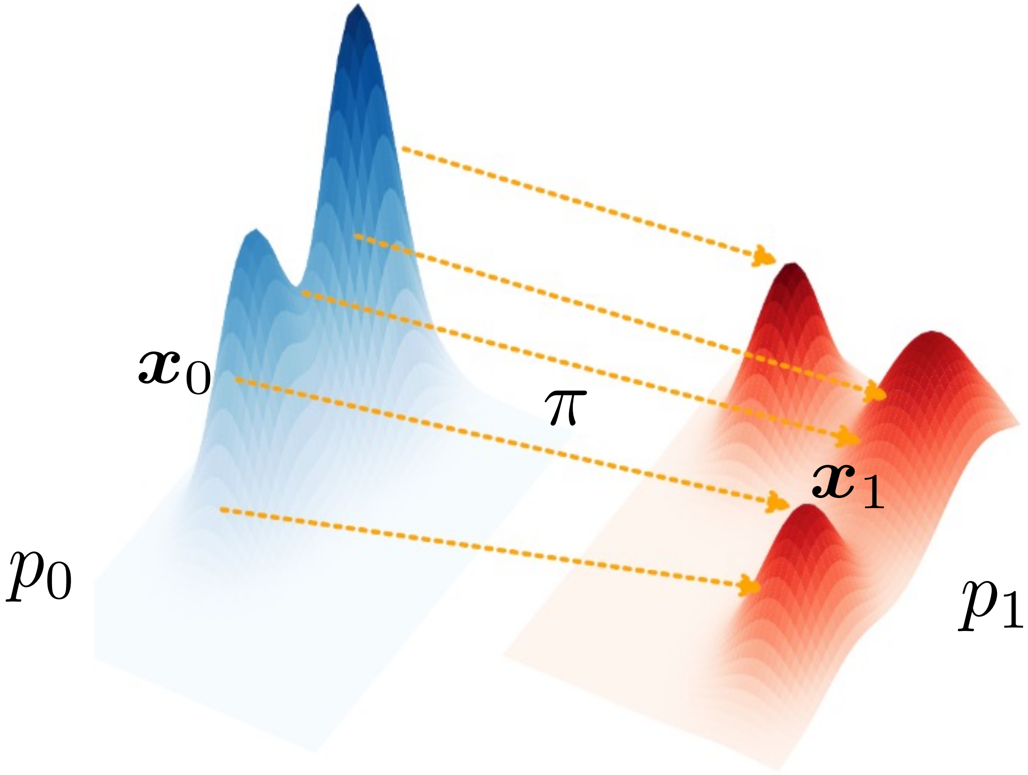
Cat

Is there a simpler way to find the optimal coupling?

Human

Cat

# Optimal Transport




$$\pi^* = \underset{\pi \in \Pi(p_0, p_1)}{\operatorname{argmin}} \int_{\mathbb{R}^n \times \mathbb{R}^n} \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|_2^2 \, \mathrm{d}\pi(\boldsymbol{x}_0, \boldsymbol{x}_1) =: \mathrm{OT}(p_0, p_1)$$

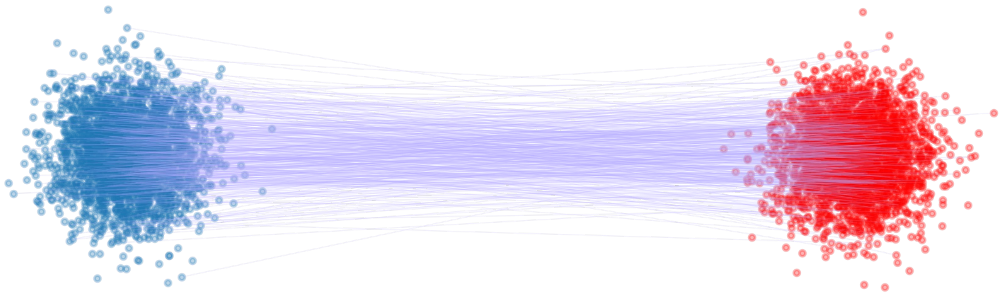
# Optimal Transport Flow Matching

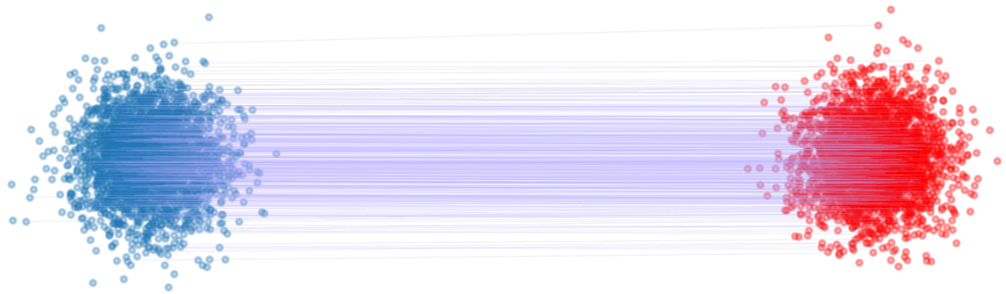
Random Coupling

$$\boldsymbol{x}_0 \sim p_0, \boldsymbol{x}_1 \sim p_1$$

Optimal Coupling

$$(\boldsymbol{x}_0, \boldsymbol{x}_1) \sim \pi^* = \mathrm{OT}(p_0, p_1)$$





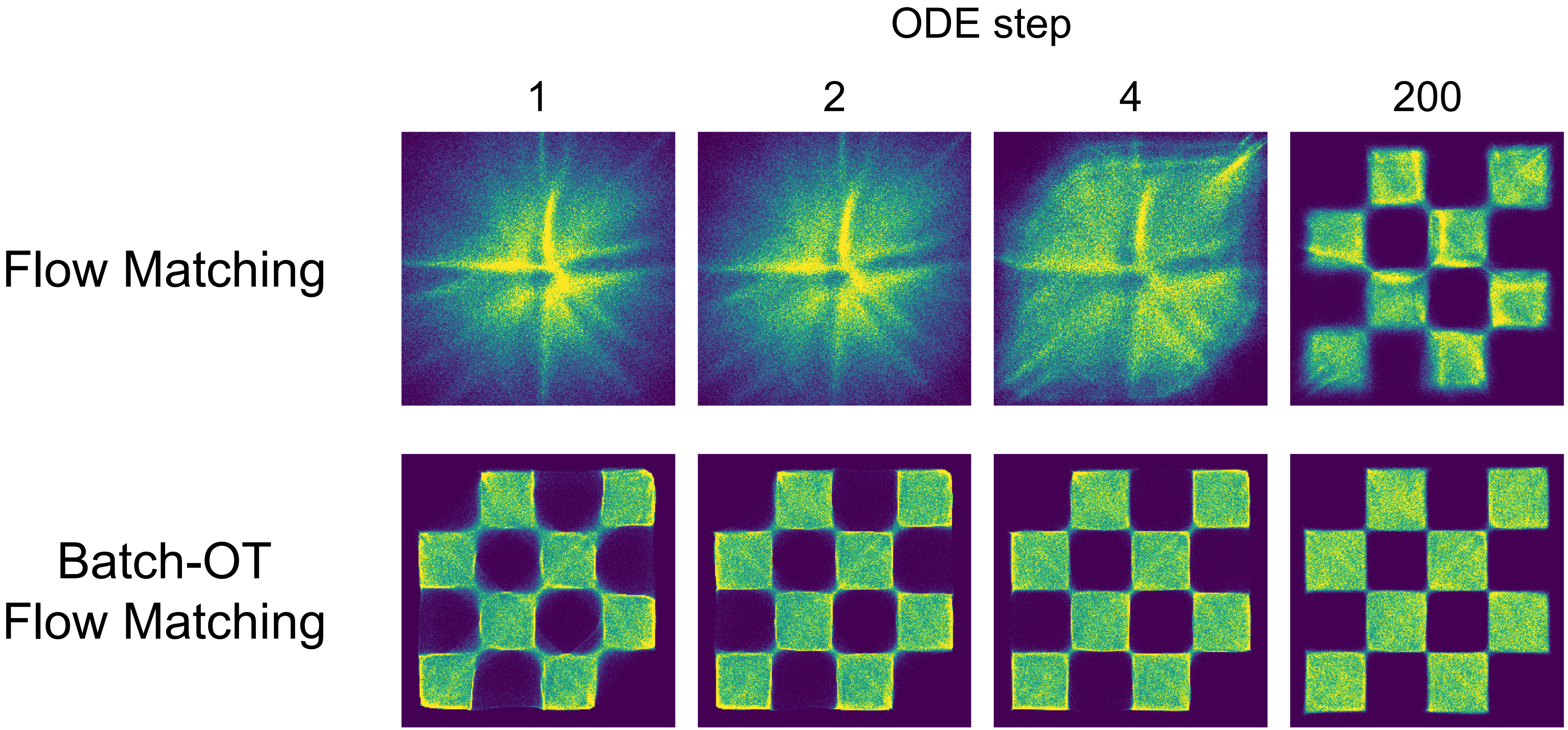
# Batch Optimal Transport Flow Matching

Computing optimal transport over the entire dataset is prohibitively expensive. We instead compute optimal transport on each sampled mini-batch during training.

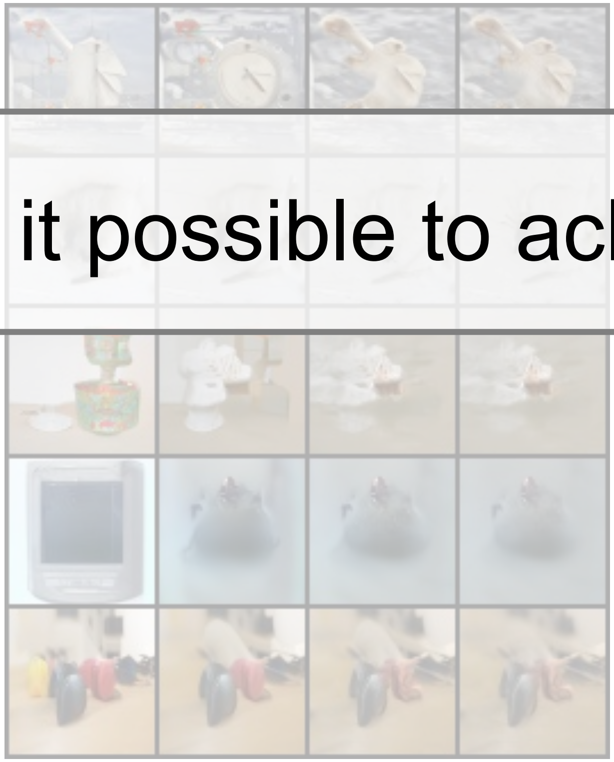
$$(\boldsymbol{x}_0, \boldsymbol{x}_1) \sim \pi^*_{\textsf{batch}} = \mathrm{OT}(\{\boldsymbol{x}_0^{(i)}\}_{i=1}^{b}, \{\boldsymbol{x}_1^{(i)}\}_{i=1}^{b})$$

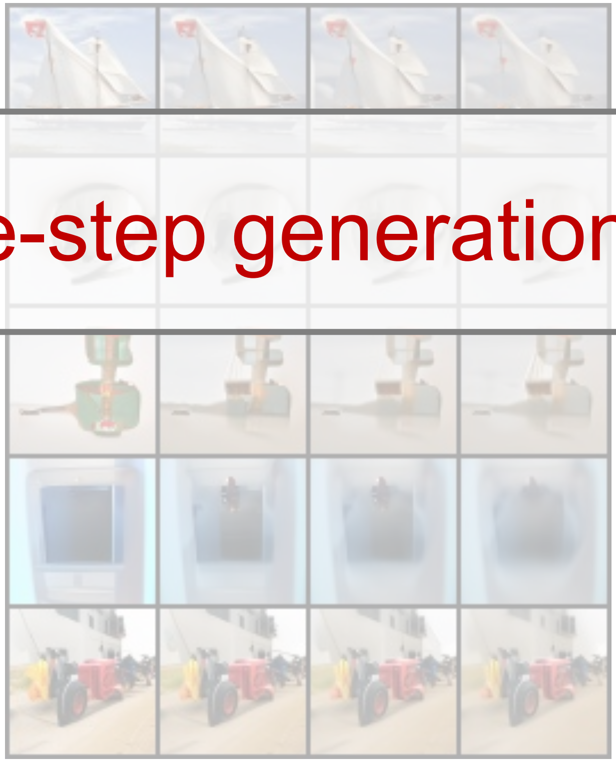
$$\mathcal{L}(\theta) = \mathbb{E}_{t,(\boldsymbol{x}_0,\boldsymbol{x}_1)\sim\pi^*_{\textsf{batch}}} \|\boldsymbol{v}_\theta(\boldsymbol{x}_t, t) - \boldsymbol{v}_{\mathrm{tgt}}\|_2^2$$

# Diffusion vs Batch-OT Flow Matching



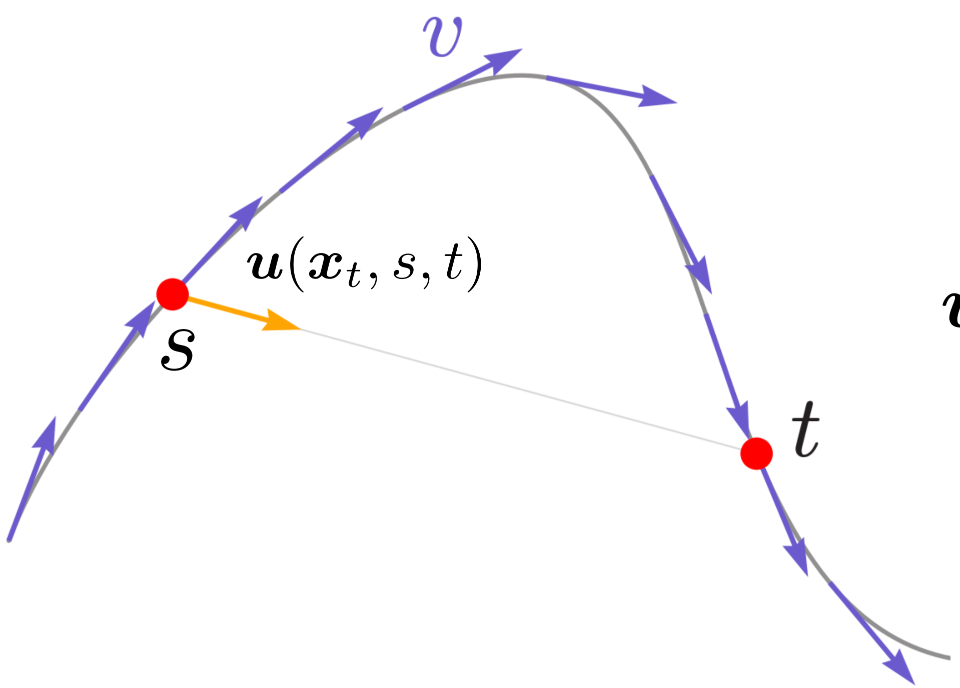
# Flow Matching vs Batch OT Flow Matching




Is it possible to achieve one-step generation?

# MeanFlow: Average Velocity




$$\boldsymbol{u}(\boldsymbol{x}_t, s, t) := \frac{1}{t-s}\int_s^t \boldsymbol{v}(\boldsymbol{x}_\tau, \tau)\mathrm{d}\tau, \quad t > s$$

# MeanFlow Identity

Differential

$$\boldsymbol{u}(\boldsymbol{x}_t, s, t) = \frac{1}{t-s}\int_s^t \boldsymbol{v}(\boldsymbol{x}_\tau, \tau)\mathrm{d}\tau$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(t-s)\boldsymbol{u}(\boldsymbol{x}_t, s, t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_s^t \boldsymbol{v}(\boldsymbol{x}_\tau, \tau)\mathrm{d}\tau$$

$$\boldsymbol{u}(\boldsymbol{x}_t, s, t) + (t-s)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(\boldsymbol{x}_t, s, t) = \boldsymbol{v}(\boldsymbol{x}_t, t)$$

$$\boldsymbol{u}(\boldsymbol{x}_t, s, t) = \boldsymbol{v}(\boldsymbol{x}_t, t) - (t-s)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(\boldsymbol{x}_t, s, t)$$

Integral

# MeanFlow: Time Derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(\boldsymbol{x}_t, s, t) = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}_t} \cdot \frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} + \frac{\partial \boldsymbol{u}}{\partial s} \cdot \frac{\mathrm{d}s}{\mathrm{d}t} + \frac{\partial \boldsymbol{u}}{\partial t} \cdot \frac{\mathrm{d}t}{\mathrm{d}t}$$

$$= \boldsymbol{v}(\boldsymbol{x}_t, t)\partial_{\boldsymbol{x}_t}\boldsymbol{u} + \partial_t \boldsymbol{u}$$

$$= \left[\frac{\partial \boldsymbol{u}(\boldsymbol{x}_t, s, t)}{\partial(\boldsymbol{x}_t, s, t)}\right] \begin{bmatrix}\boldsymbol{v}(\boldsymbol{x}_t, t) & 0 & 1\end{bmatrix}^\top$$ (Jacobian-Vector Product)

$$\boldsymbol{u}(\boldsymbol{x}_t, s, t) = \boldsymbol{v}(\boldsymbol{x}_t, t) - (t - s)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(\boldsymbol{x}_t, s, t)$$

$$= (\boldsymbol{x}_1 - \boldsymbol{x}_0) - (t - s)(\boldsymbol{v}(\boldsymbol{x}_t, t)\partial_{\boldsymbol{x}_t}\boldsymbol{u} + \partial_t \boldsymbol{u})$$

# MeanFlow: Training Objective

$$\boldsymbol{u}(\boldsymbol{x}_t, s, t) = (\boldsymbol{x}_1 - \boldsymbol{x}_0) - (t - s)(\boldsymbol{v}(\boldsymbol{x}_t, t)\partial_{\boldsymbol{x}_t}\boldsymbol{u} + \partial_t \boldsymbol{u})$$

$$\mathcal{L}(\theta) = \|\boldsymbol{u}_\theta(\boldsymbol{x}_t, s, t) - \mathrm{sg}[\boldsymbol{u}_{\mathrm{tgt}}]\|_2^2$$

$$\boldsymbol{u}_{\mathrm{tgt}}(\boldsymbol{x}_t, s, t) = (\boldsymbol{x}_1 - \boldsymbol{x}_0) - (t - s)(\boldsymbol{v}(\boldsymbol{x}_t, t)\partial_{\boldsymbol{x}_t}\boldsymbol{u}_\theta + \partial_t \boldsymbol{u}_\theta)$$

# MeanFlow: Sampling

$$\boldsymbol{u}_\theta(\boldsymbol{x}_t, s, t) = \frac{1}{t-s}\int_s^t \boldsymbol{v}(\boldsymbol{x}_\tau, \tau)\mathrm{d}\tau$$

$$(t-s)\boldsymbol{u}_\theta(\boldsymbol{x}_t, s, t) = \int_s^t \boldsymbol{v}(\boldsymbol{x}_\tau, \tau)\mathrm{d}\tau = \boldsymbol{x}_t - \boldsymbol{x}_s$$

$$\boldsymbol{x}_s = \boldsymbol{x}_t - (t-s)\boldsymbol{u}_\theta(\boldsymbol{x}_t, s, t)$$

Multi-step

$$\boldsymbol{x}_{t_i} = \boldsymbol{x}_{t_{i+1}} - (t_{i+1} - t_i)\boldsymbol{u}_\theta(\boldsymbol{x}_{t_{i+1}}, t_i, t_{i+1})$$

One-step

$$\boldsymbol{x}_0 = \boldsymbol{x}_1 - \boldsymbol{u}_\theta(\boldsymbol{x}_1, 0, 1)$$

# Experiments

Result on ImageNet-256 x 256

| method | params | NFE | FID |
|---|---|---|---|
| *1-NFE diffusion/flow from scratch* | | | |
| iCT-XL/2 [43]† | 675M | 1 | 34.24 |
| Shortcut-XL/2 [13] | 675M | 1 | 10.60 |
| MeanFlow-B/2 | 131M | 1 | 6.17 |
| MeanFlow-M/2 | 308M | 1 | 5.01 |
| MeanFlow-L/2 | 459M | 1 | 3.84 |
| MeanFlow-XL/2 | 676M | 1 | **3.43** |
| *2-NFE diffusion/flow from scratch* | | | |
| iCT-XL/2 [43]† | 675M | 2 | 20.30 |
| iMM-XL/2 [52] | 675M | 1×2 | 7.77 |
| MeanFlow-XL/2 | 676M | 2 | 2.93 |
| MeanFlow-XL/2+ | 676M | 2 | **2.20** |

| method | params | NFE | FID |
|---|---|---|---|
| *GANs* | | | |
| BigGAN [5] | 112M | 1 | 6.95 |
| GigaGAN [21] | 569M | 1 | 3.45 |
| StyleGAN-XL [40] | 166M | 1 | 2.30 |
| **autoregressive/masking** | | | |
| AR w/ VQGAN [10] | 227M | 1024 | 26.52 |
| MaskGIT [6] | 227M | 8 | 6.18 |
| VAR-*d*30 [47] | 2B | 10×2 | 1.92 |
| MAR-H [27] | 943M | 256×2 | 1.55 |
| *diffusion/flow* | | | |
| ADM [8] | 554M | 250×2 | 10.94 |
| LDM-4-G [37] | 400M | 250×2 | 3.60 |
| SimDiff [20] | 2B | 512×2 | 2.77 |
| DiT-XL/2 [34] | 675M | 250×2 | 2.27 |
| SiT-XL/2 [33] | 675M | 250×2 | 2.06 |
| SiT-XL/2+REPA [51] | 675M | 250×2 | **1.42** |

# Summary

- Flow Matching rethinks generation as learning a vector field that directly transports noise to data, enabling more efficient and controllable sampling.
- Rectified and OT-based flows improve the alignment between source and target distributions, preventing path crossings and preserving structure during transport.
- MeanFlow introduces integral-based velocity averaging to further improve sample quality and stability, moving us closer to fast and high-fidelity generative models.

# Thank you