# Practicals of Flow Models

CS180 Fall 2025

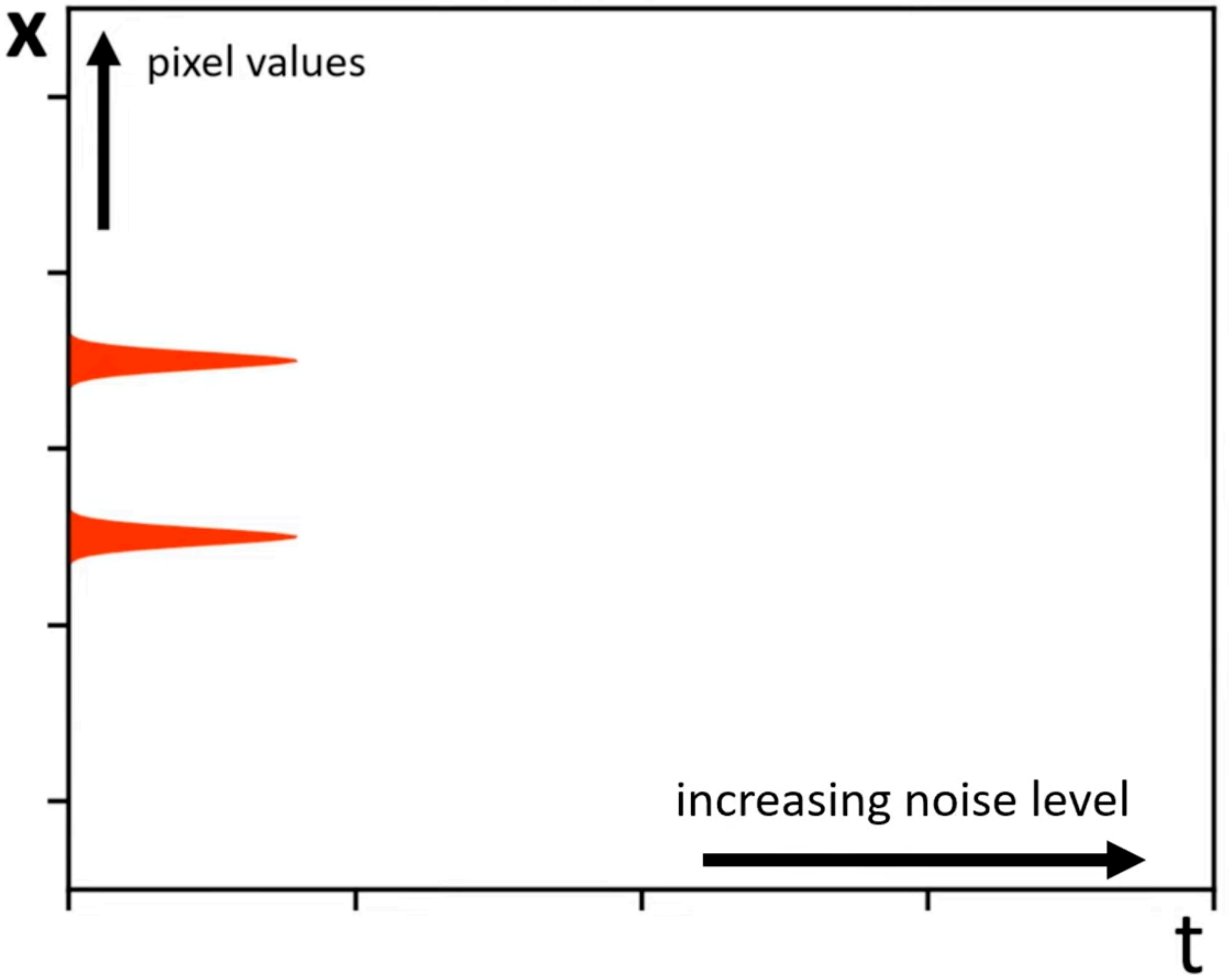
Angjoo Kanazawa & Alexei Efros

Slide heavily based on one made by Songwei Ge & David McAllister

# Outline

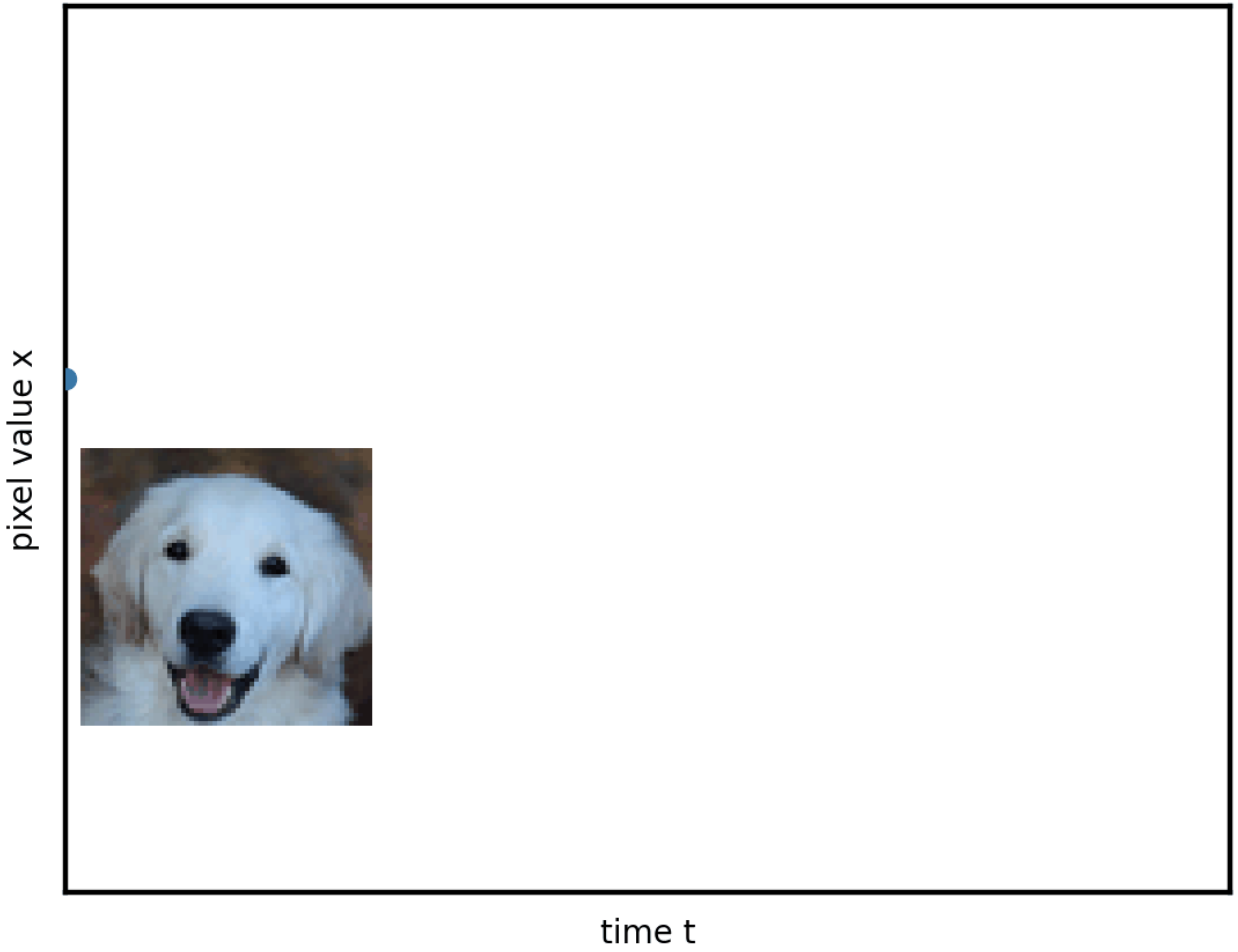
- Flow matching model samplers
- Inversion/Distillation
- Guidance
- Application

# Revisit diffusion models with a 1D example



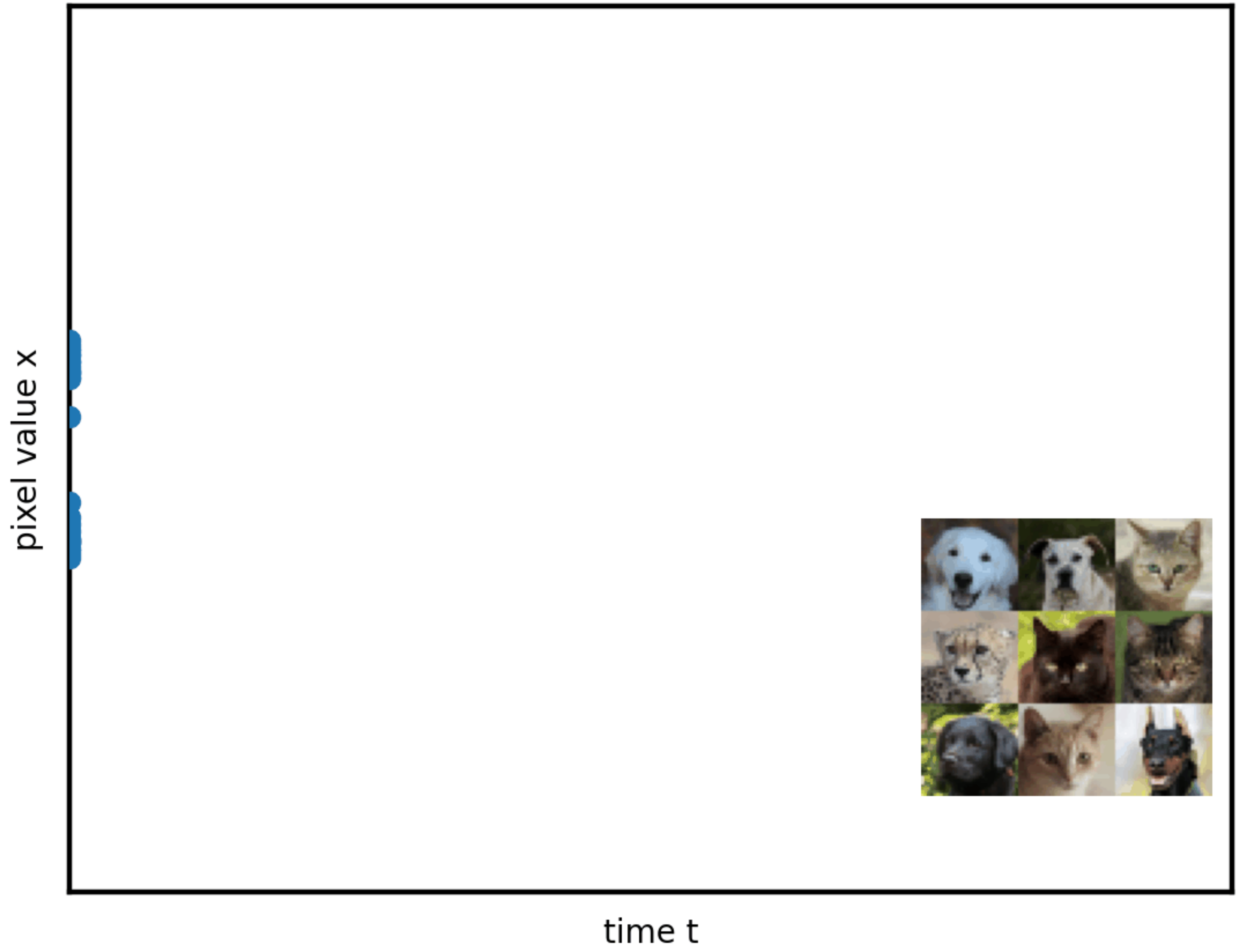

Credit: Miika Aittala

# Revisit diffusion models with a 1D example



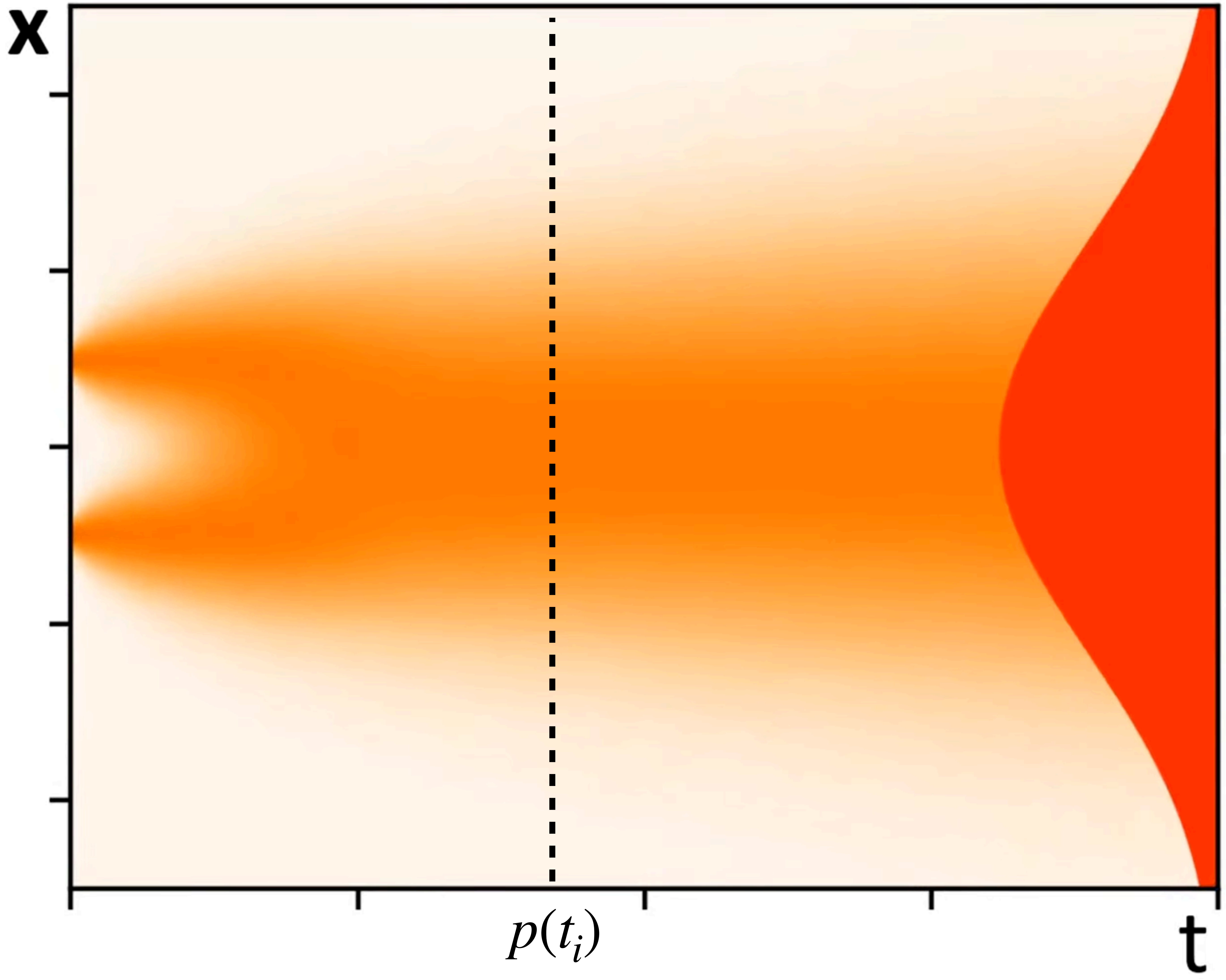

Credit: Miika Aittala

# Revisit diffusion models with a 1D example




Credit: Miika Aittala

# Revisit diffusion models with a 1D example

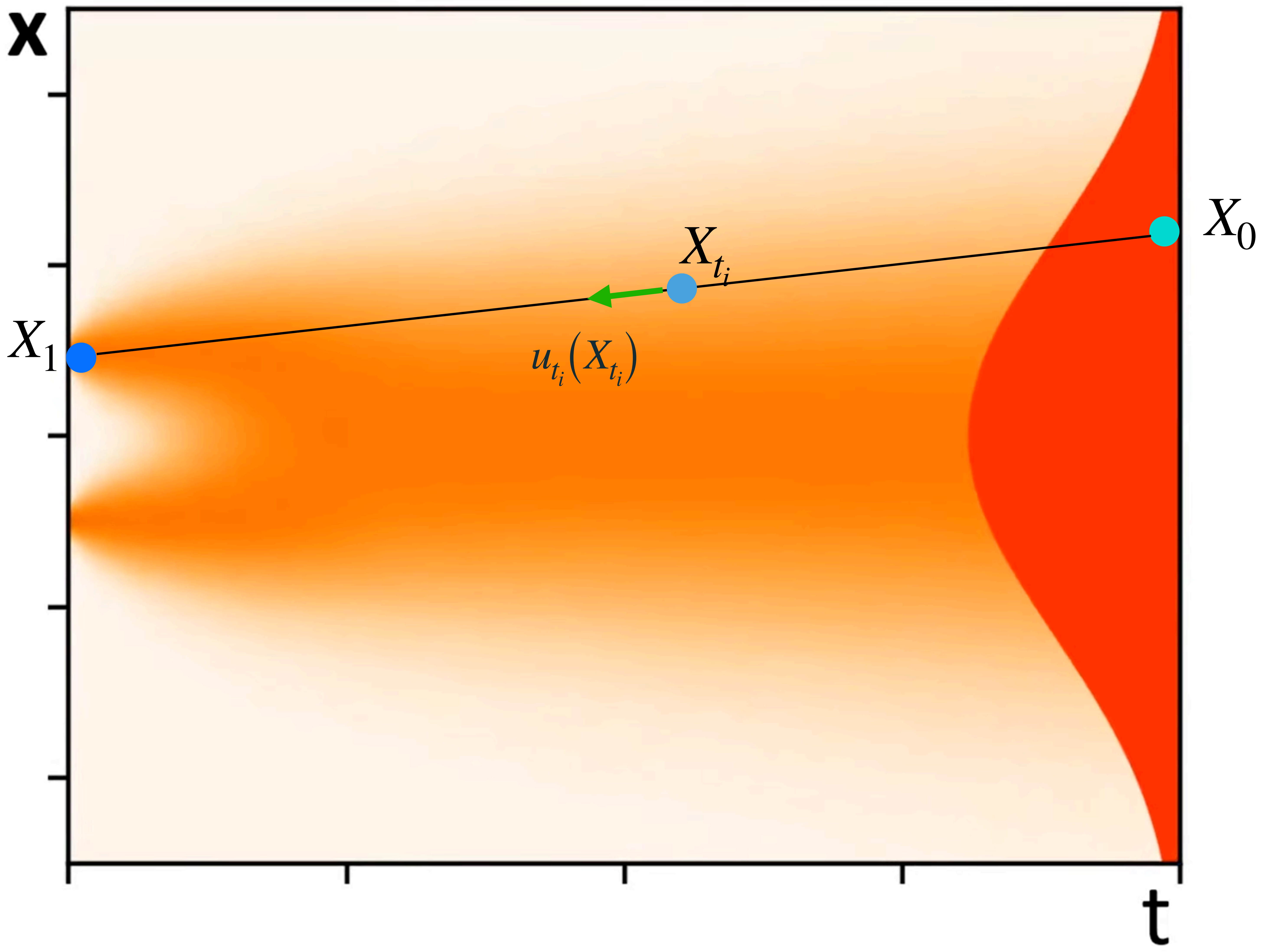



Credit: Miika Aittala

# Flow matching model training

Training Objective:

$$\mathbb{E}_{t,X_0,X_1}\left\| u_{t_i}^{\theta}(X_{t_i}) - u_{t_i}(X_{t_i}) \right\|^2$$

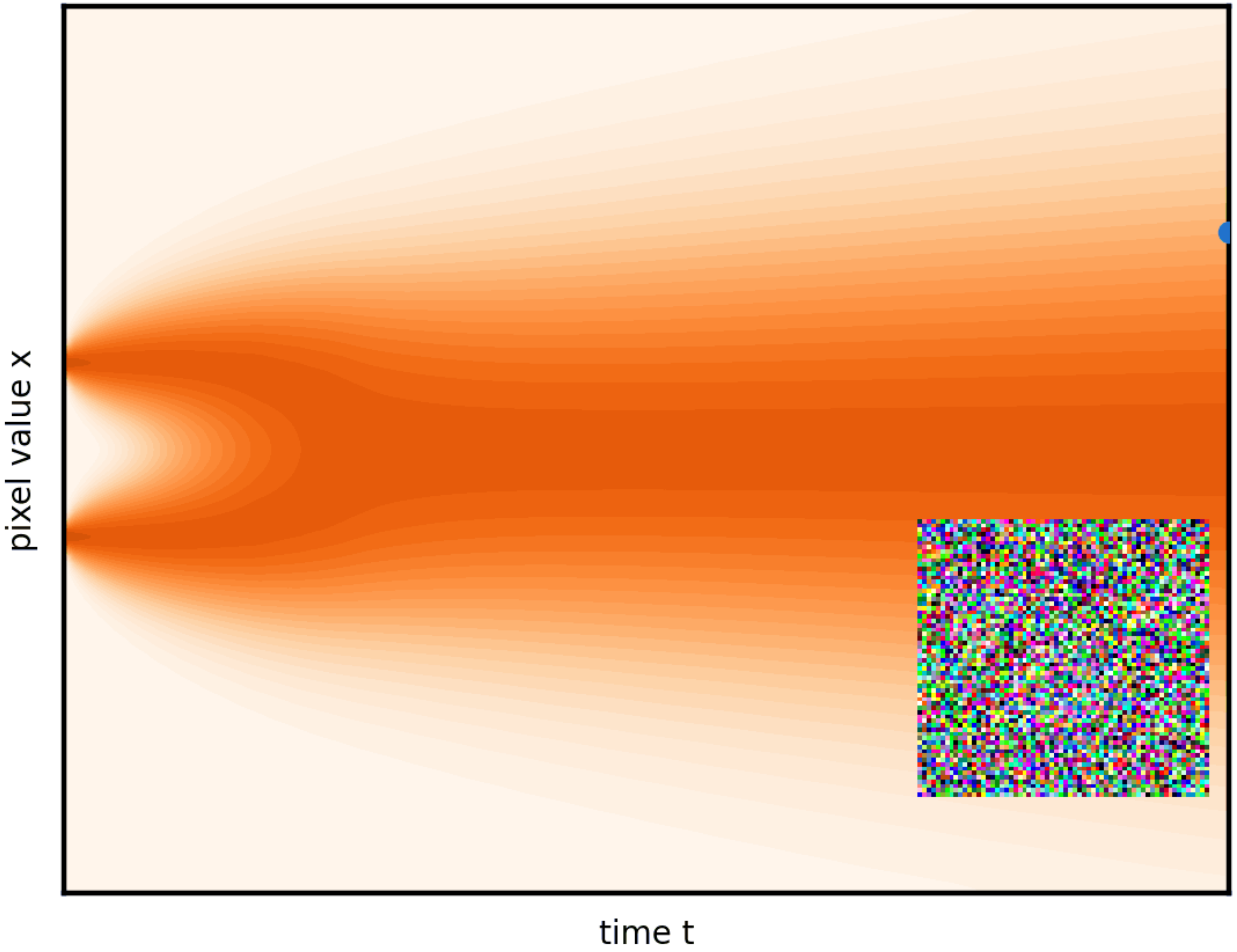



Credit: Miika Aittala

# Sampling by solving the flow ODE

Flow ODE

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t^{\theta}(x_t)$$

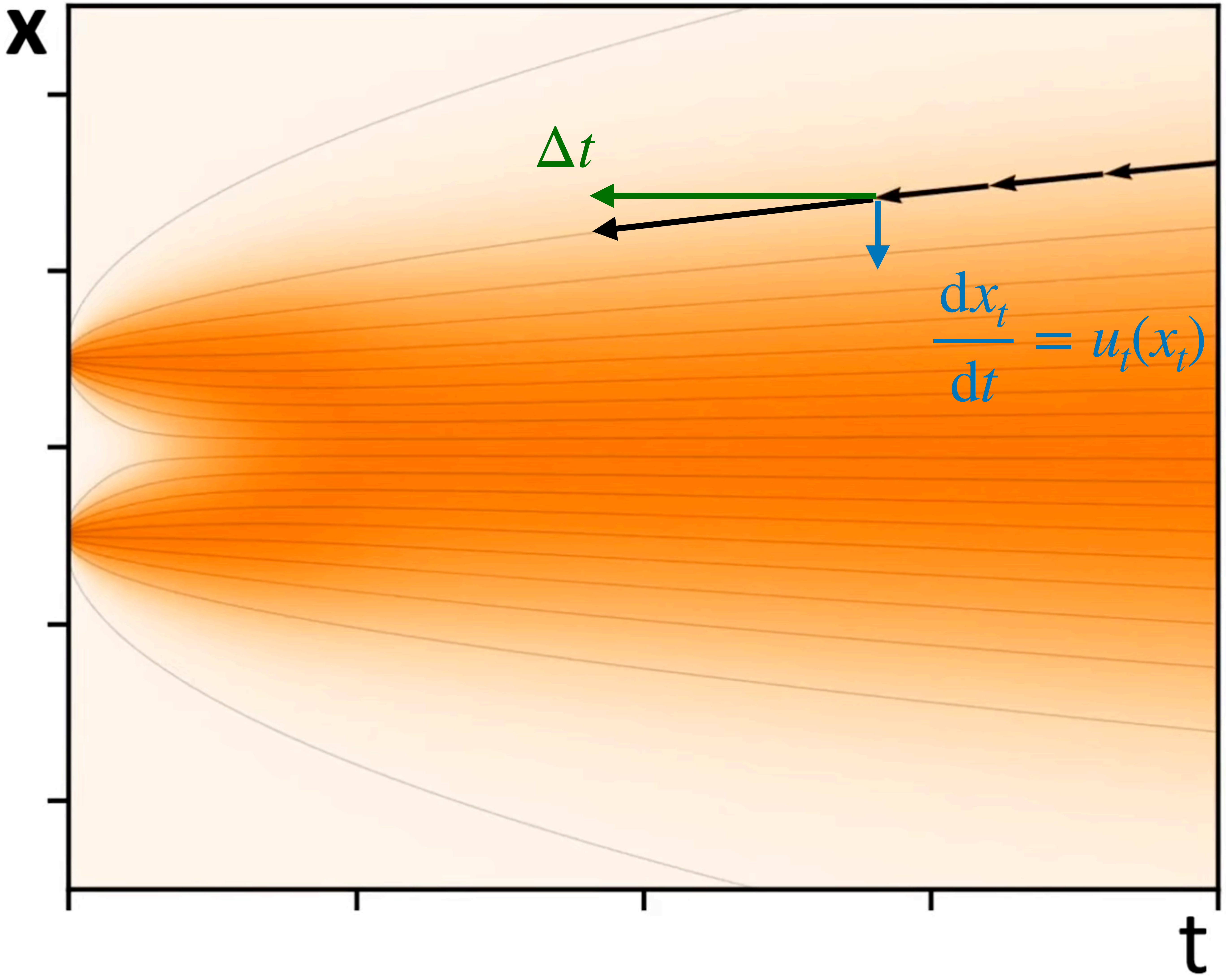



Credit: Miika Aittala

# Solving the flow ODE with discretization

Euler step:

$$x_{t+\Delta t} = x_t + \Delta t \cdot \frac{dx}{dt}\Big|_{x_t,t}$$

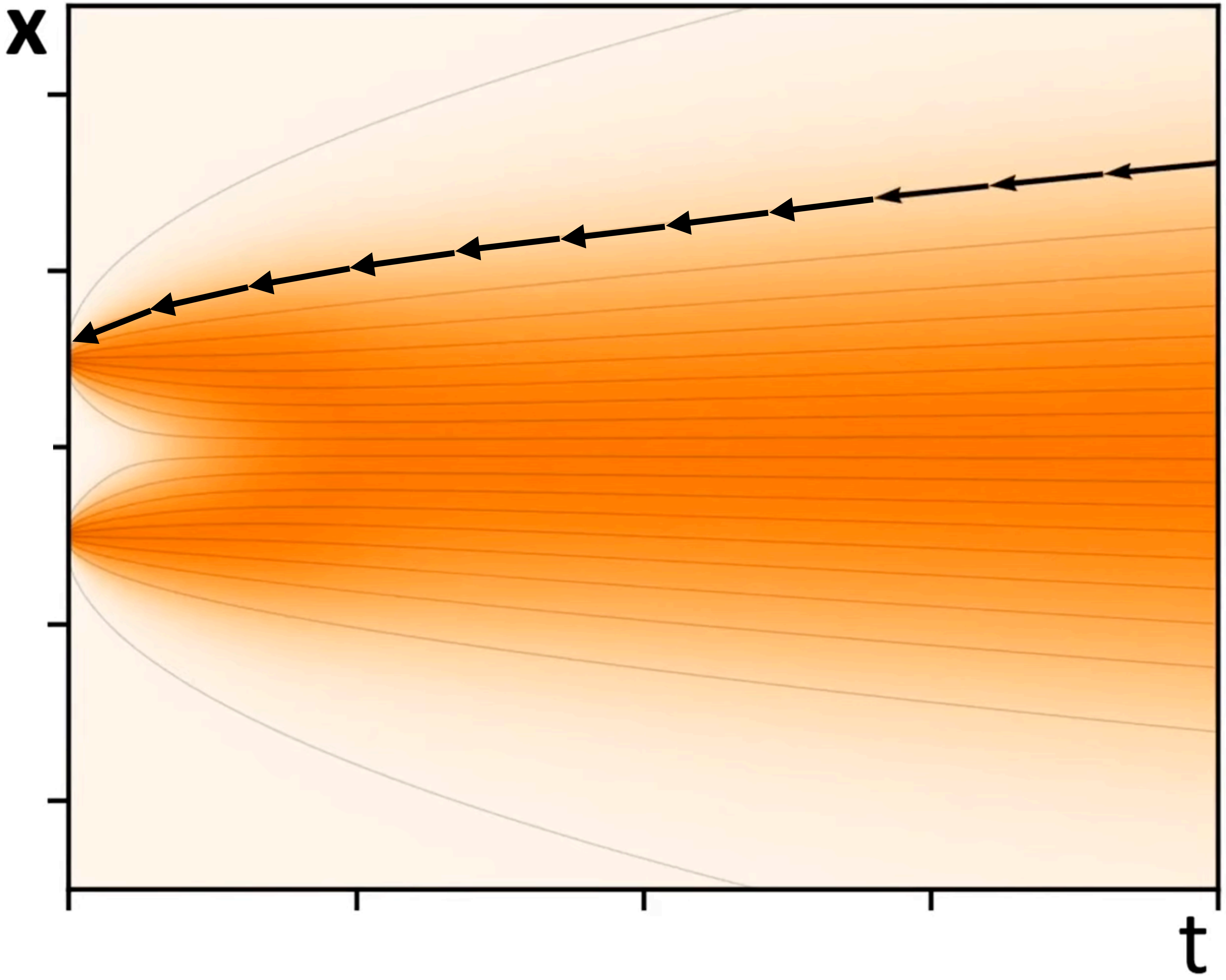



Credit: Miika Aittala

# Solving the flow ODE with discretization

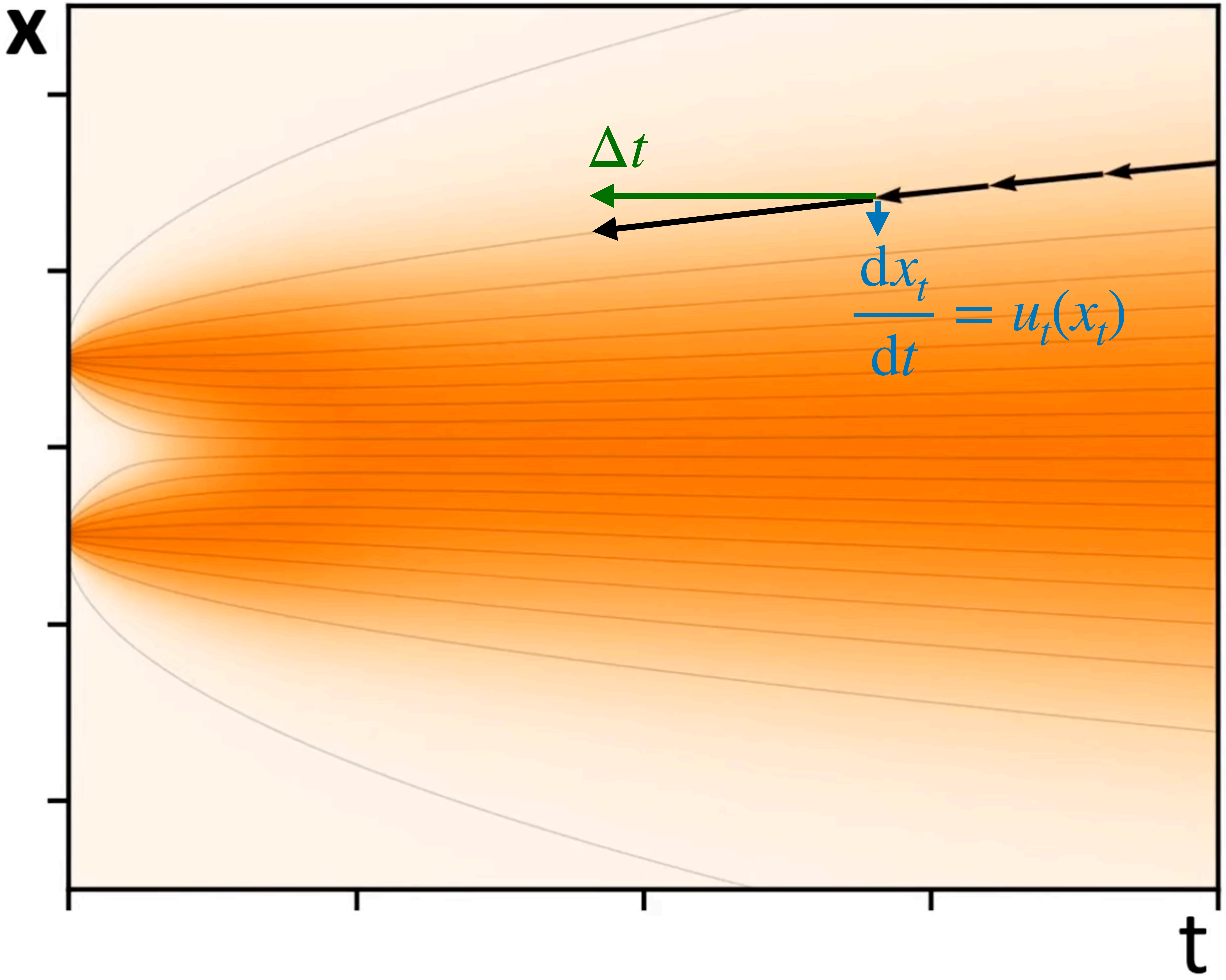
Euler step:

$$x_{t+\Delta t} = x_t + \Delta t \cdot \frac{dx}{dt}\Big|_{x_t,t}$$



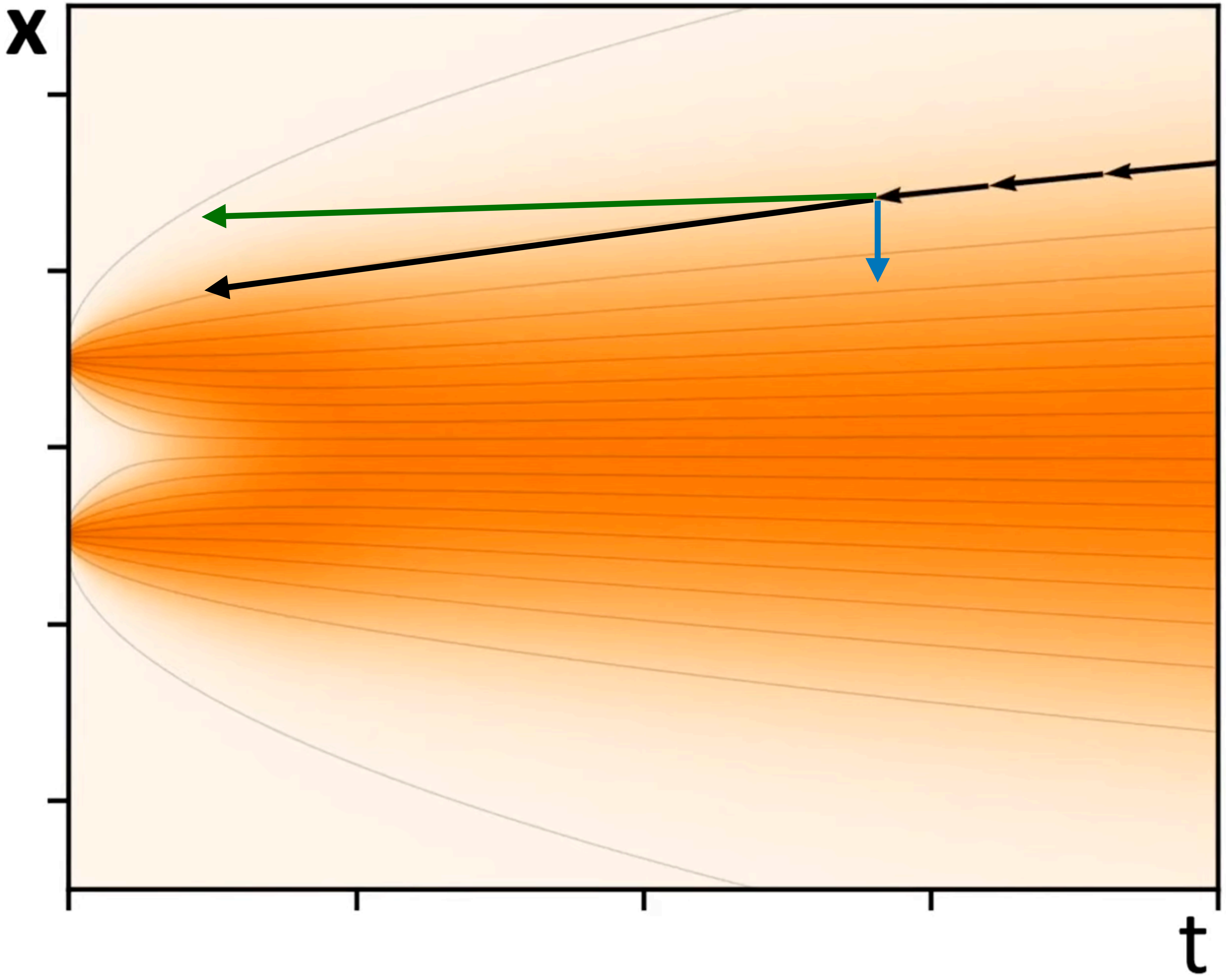

Credit: Miika Aittala

# What can go wrong with this sampling process



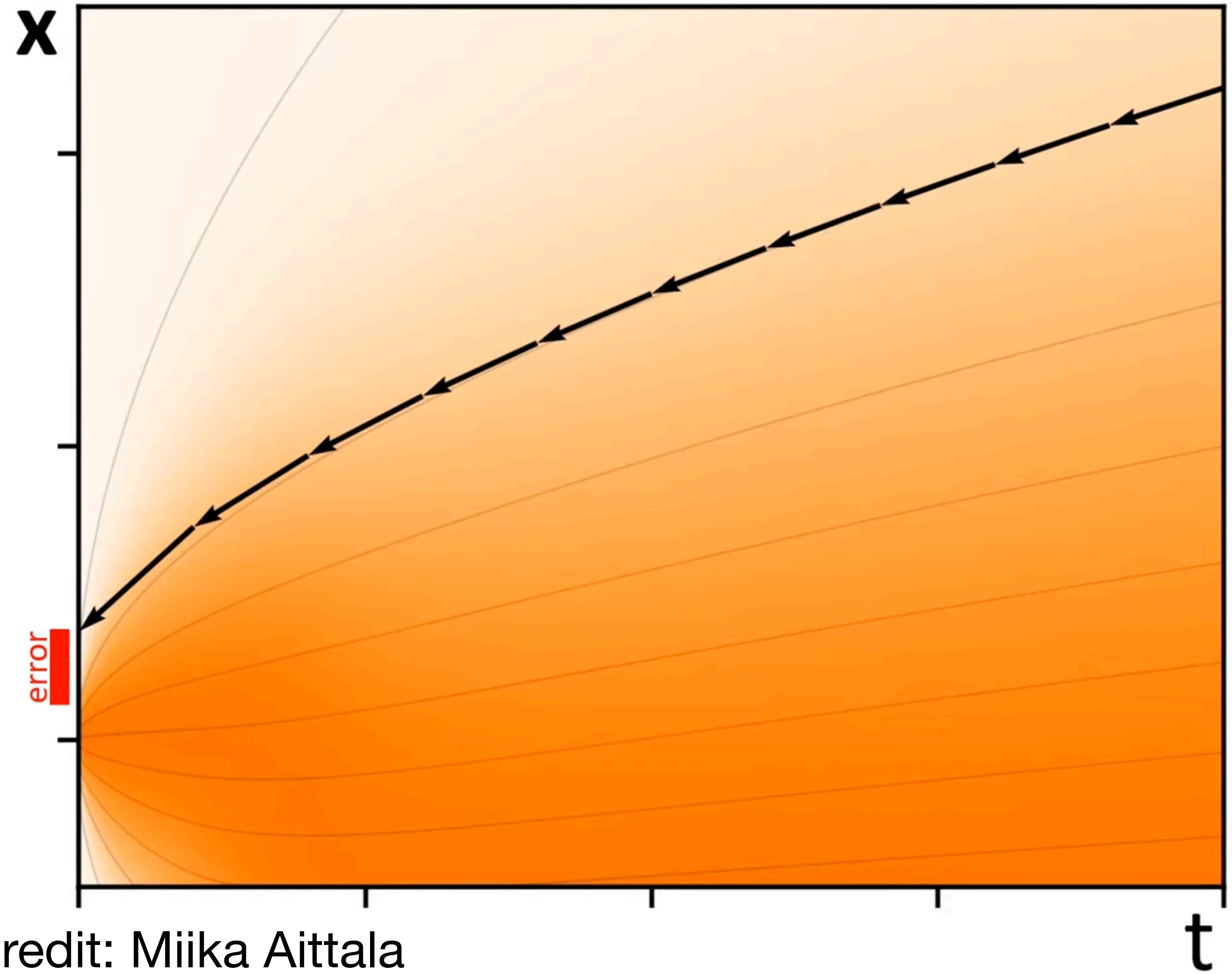

Credit: Miika Aittala

# What can go wrong with this sampling process




Credit: Miika Aittala

# Error Sources when Solving the flow ODE

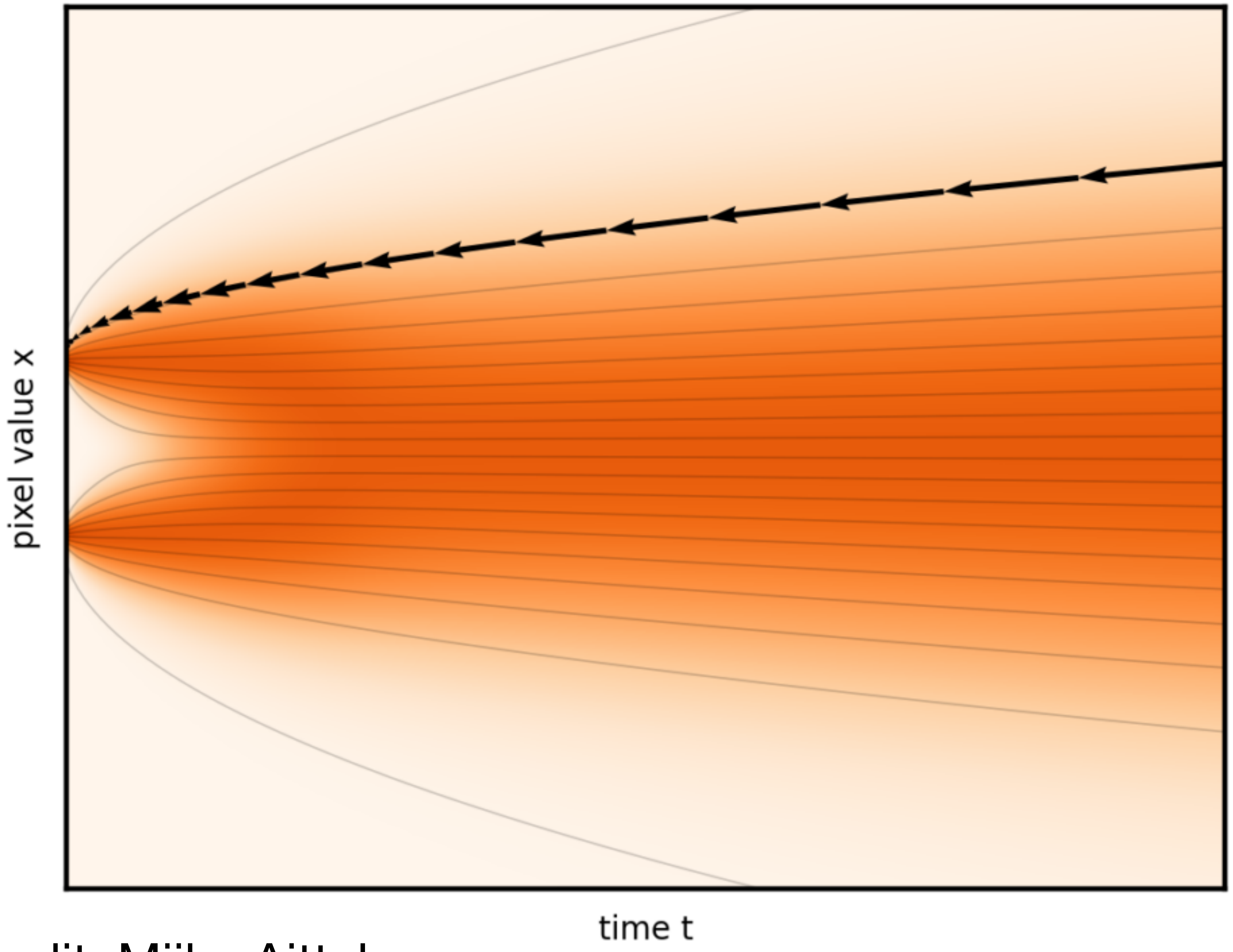



Credit: Miika Aittala

1. Truncation error: fails to approximate ideal trajectory by finite steps.

    1. Naive solution: sampling with more steps.

# Smart Time Step Schedule



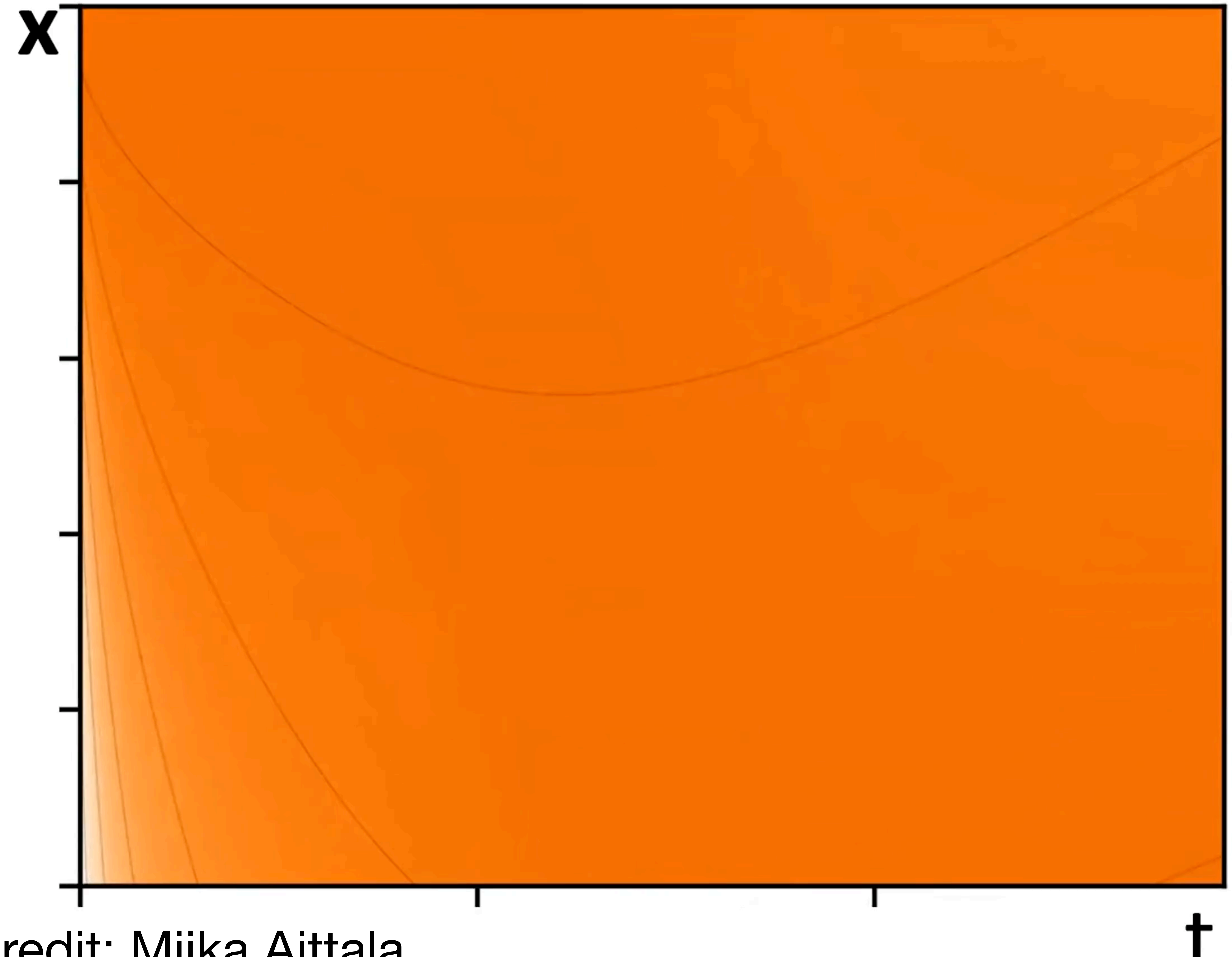

Credit: Miika Aittala

1. Truncation error: fails to approximate ideal trajectory by finite steps.

    1. Naive solution: sampling with more steps.

    2. Time steps are long at high noise levels and short at low noise levels

# Advanced ODE Solvers




Credit: Miika Aittala

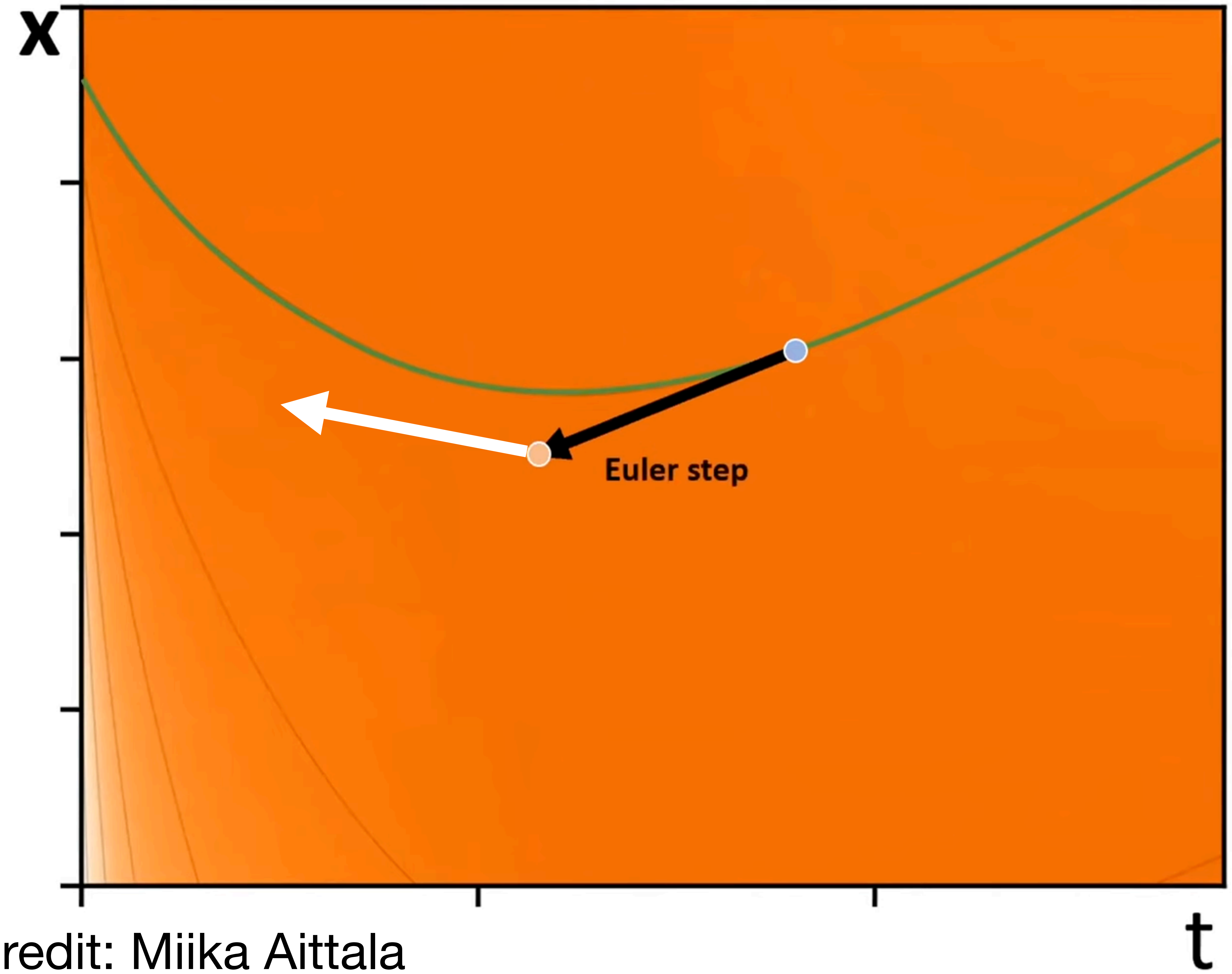
1. Truncation error: fails to approximate ideal trajectory by finite steps.

    1. Naive solution: sampling with more steps.

    2. Time steps are long at high noise levels and short at low noise levels

    3. Higher-order ODE solver

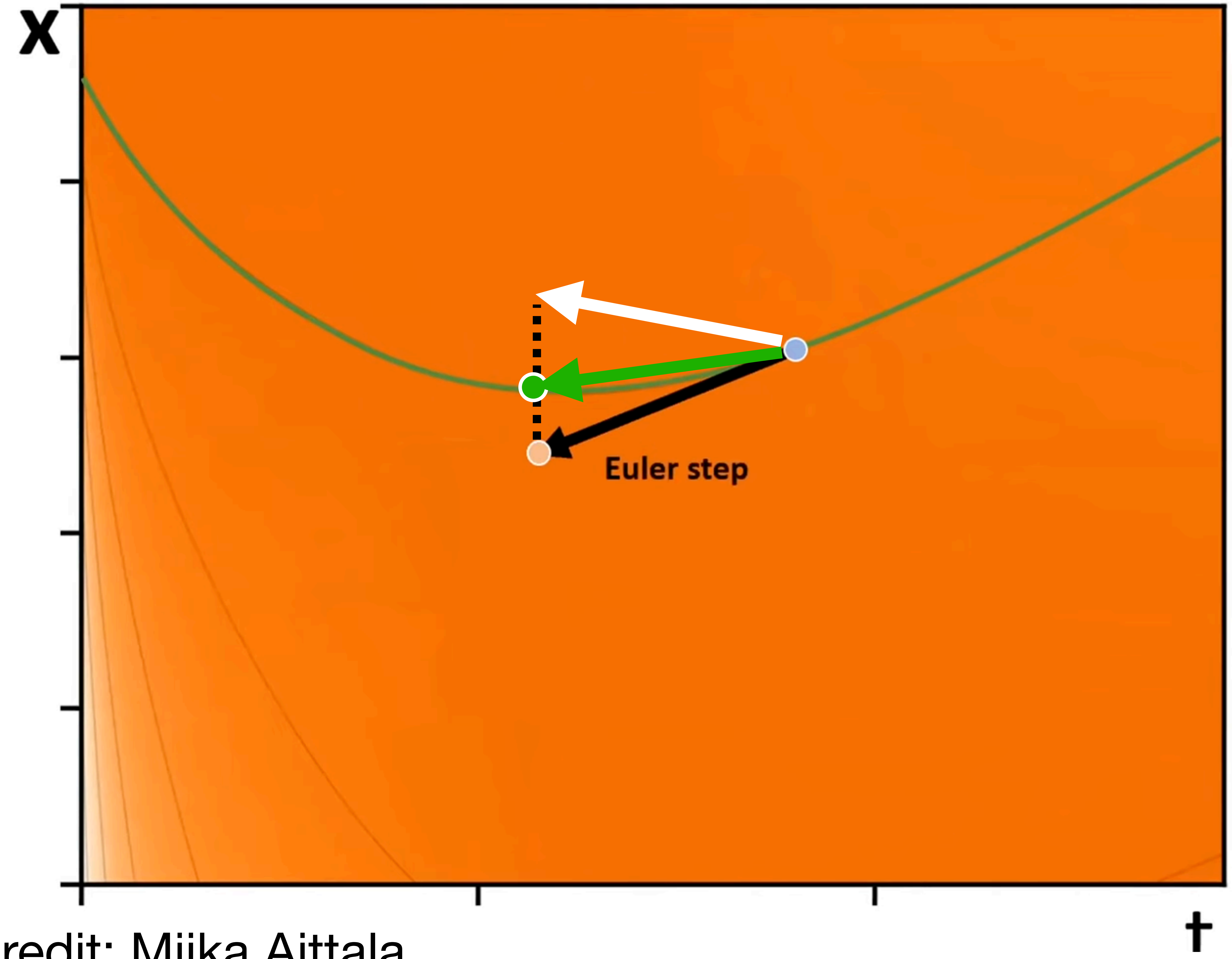
# Higher-order solvers




Credit: Miika Aittala

- Clever sub-steps -> higher accuracy though more cost
- 2nd order Heun method as an example

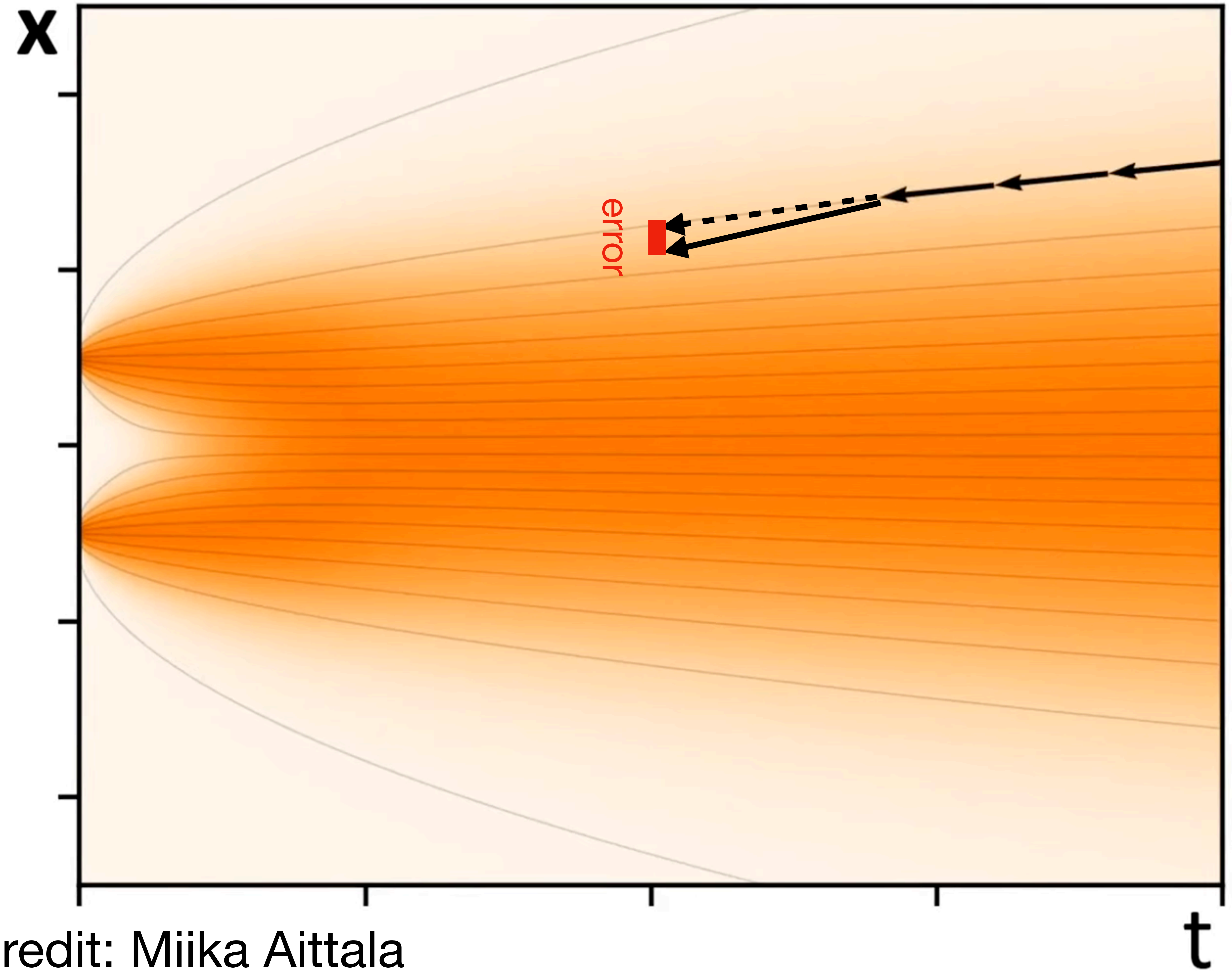
# Higher-order solvers




- Clever sub-steps -> higher accuracy though more cost
- 2nd order Heun method as an example

Credit: Miika Aittala

# Error Sources when Solving the flow ODE

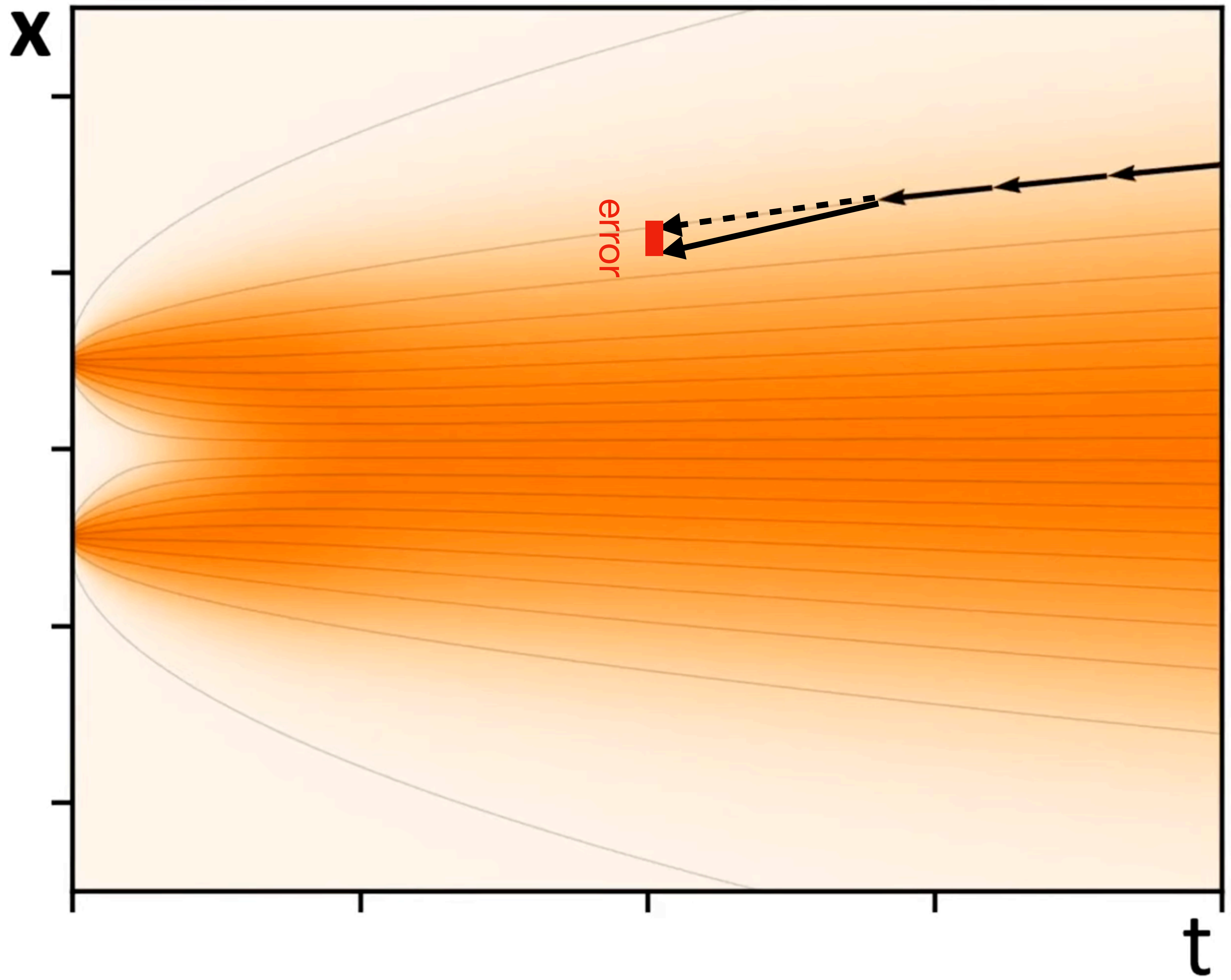



Credit: Miika Aittala

1. Truncation error: fails to approximate ideal trajectory by finite steps.

2. Model fails to approximate the marginal flow.

# Stochastic Sampler

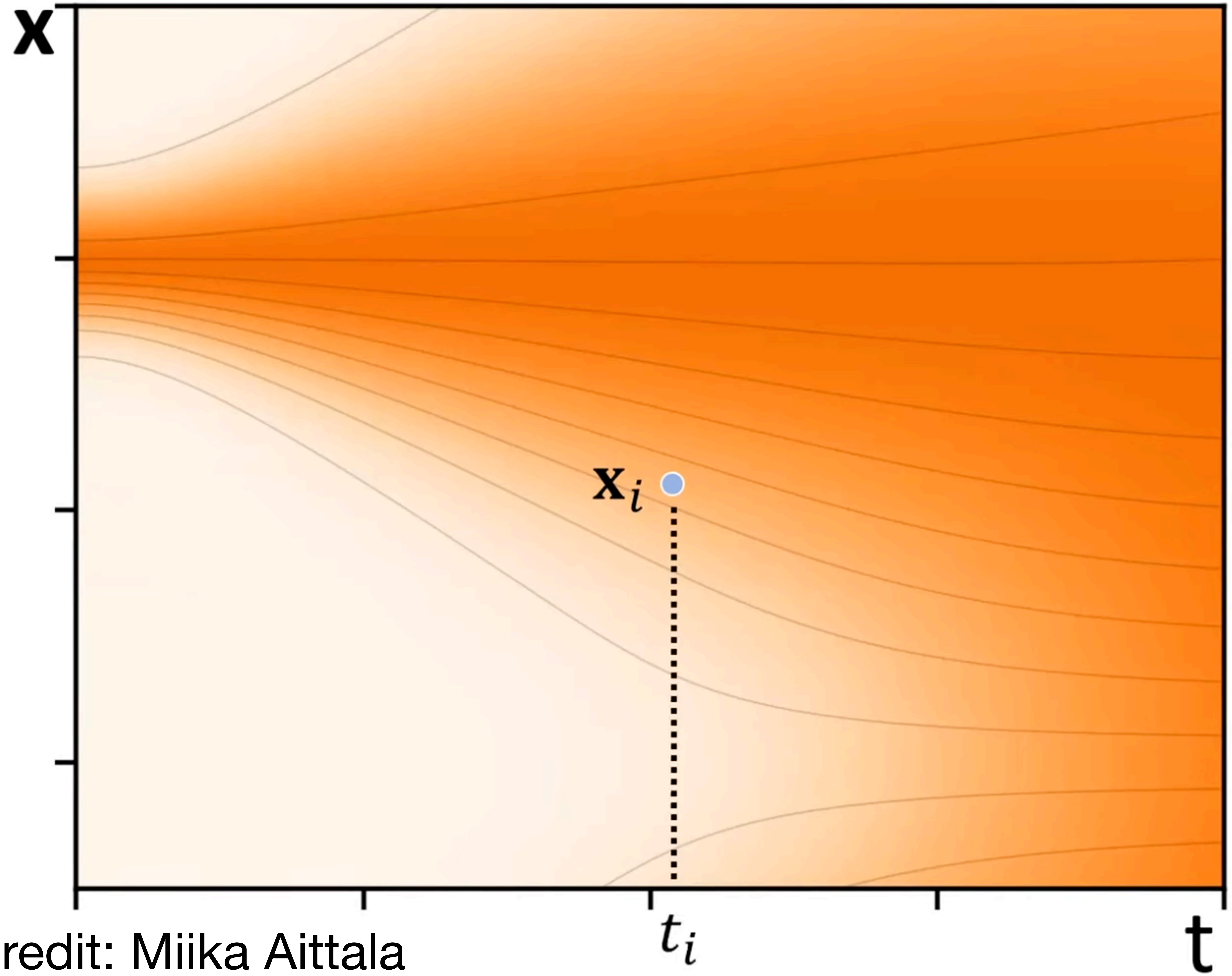



1. Truncation error: fails to approximate ideal trajectory by finite steps.

2. Model fails to approximate the marginal flow.

    1. Stochastic sampler (SDE) injects fresh noise throughout the evolution in addition to reducing the noise.

# Stochastic Sampler: why does it improve quality



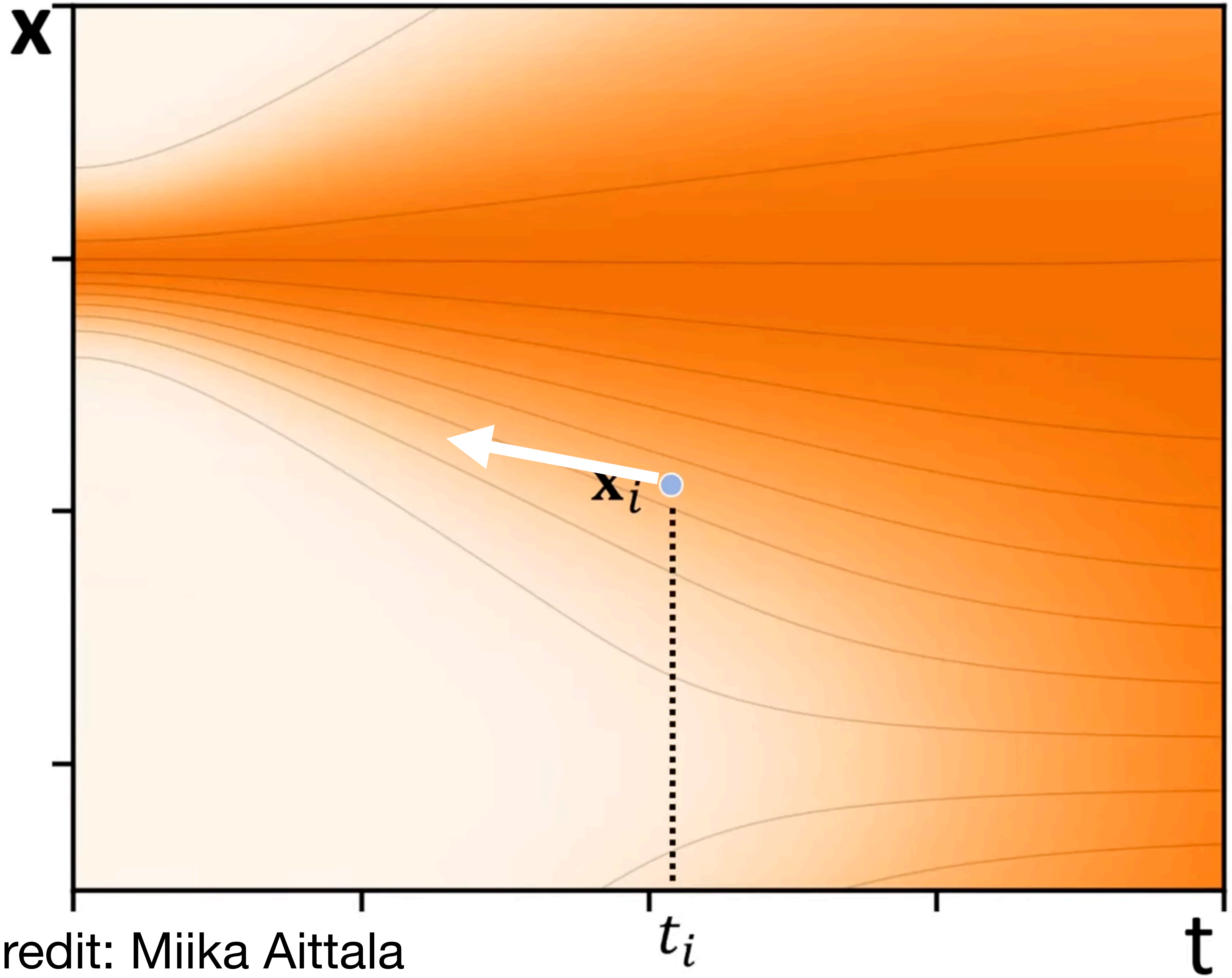

Credit: Miika Aittala

1. Truncation error: fails to approximate ideal trajectory by finite steps.

2. Model fails to approximate the marginal flow.

    1. Stochastic sampler (SDE) injects fresh noise throughout the evolution

# Stochastic Sampler: why does it improve quality



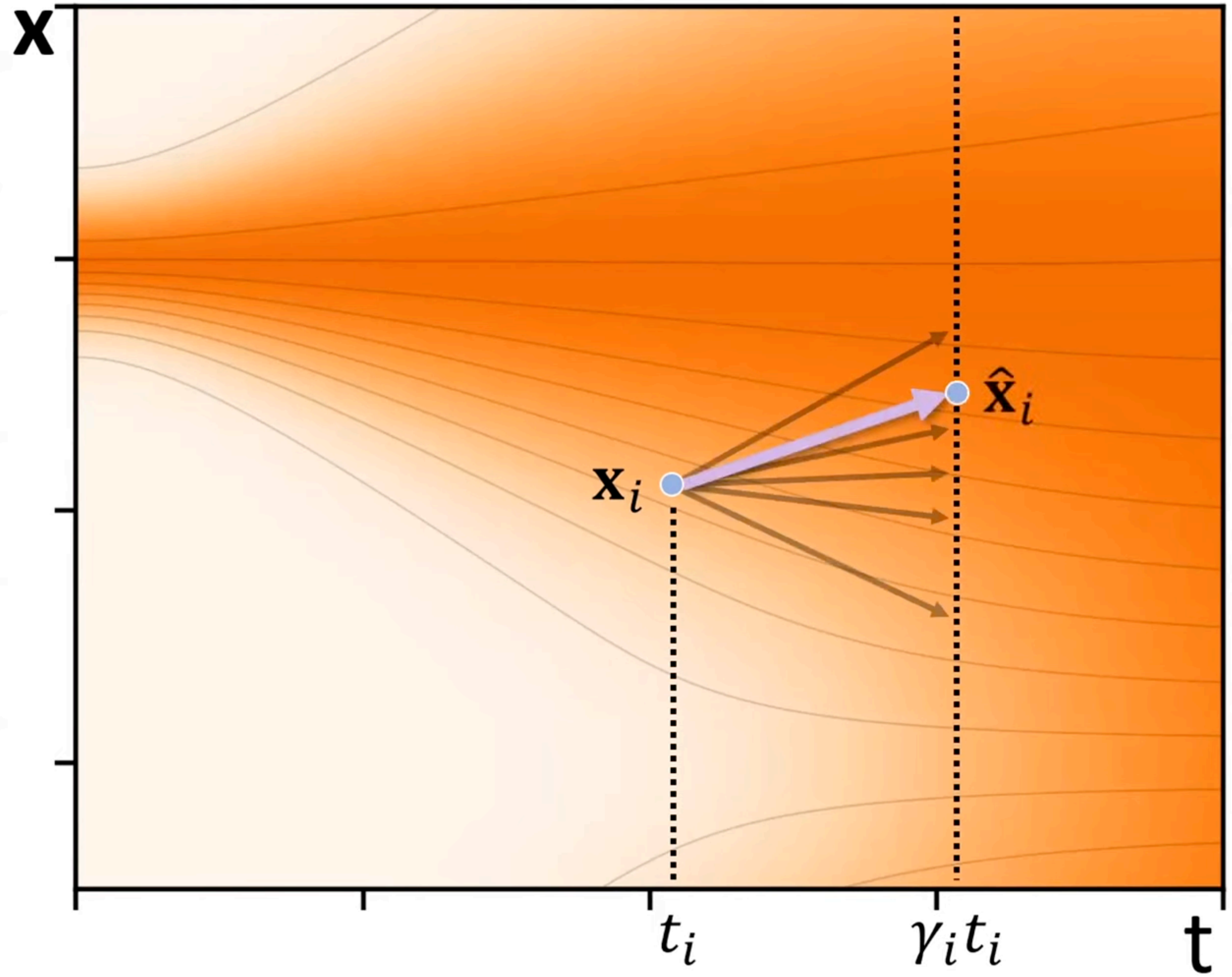

Credit: Miika Aittala

1. Truncation error: fails to approximate ideal trajectory by finite steps.

2. Model fails to approximate the marginal flow.

    1. Stochastic sampler (SDE) injects fresh noise throughout the evolution

# Stochastic Sampler: why does it improve quality

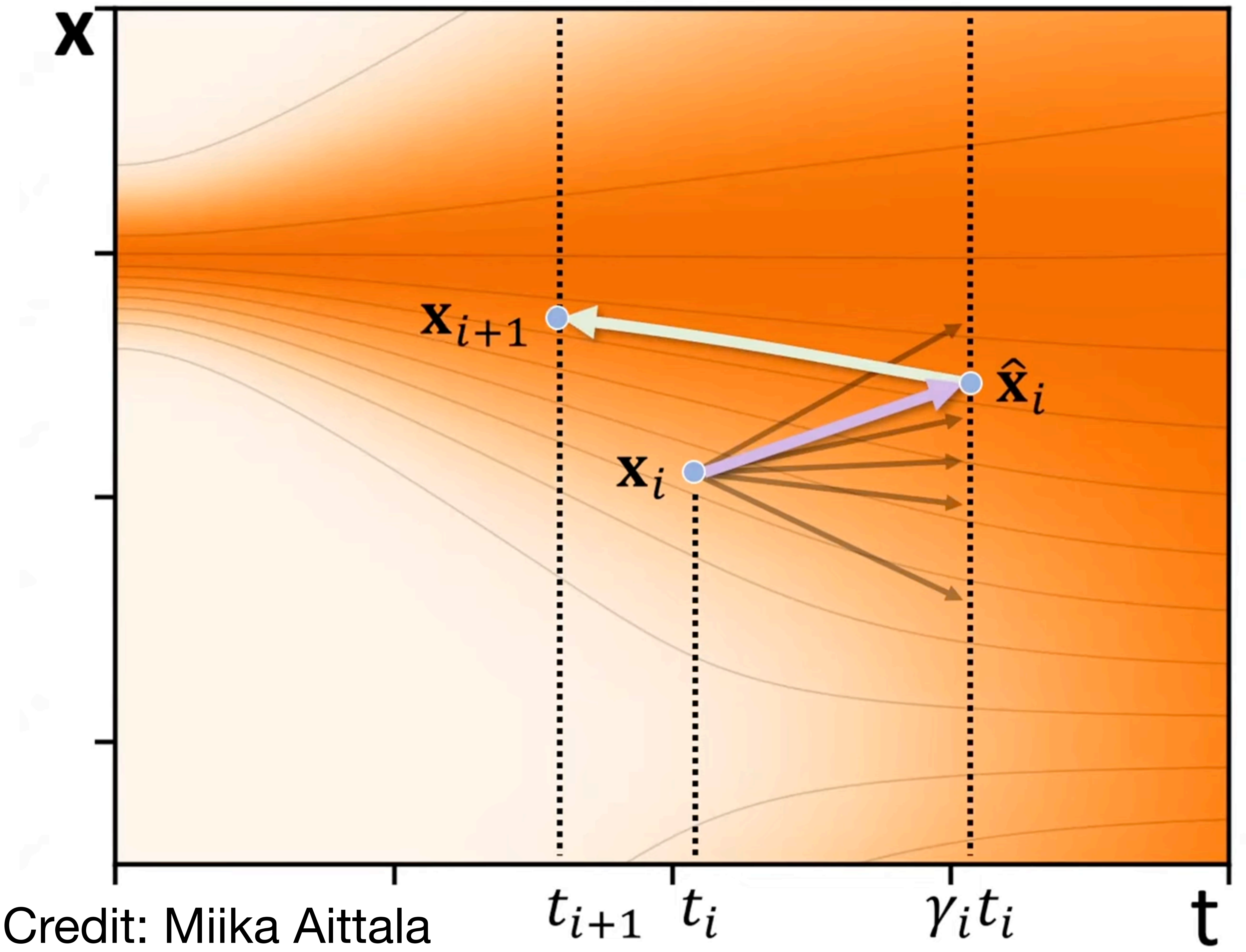



1. Truncation error: fails to approximate ideal trajectory by finite steps.

2. Model fails to approximate the marginal flow.

    1. Stochastic sampler (SDE) injects fresh noise throughout the evolution

# Stochastic Sampler: why does it improve quality



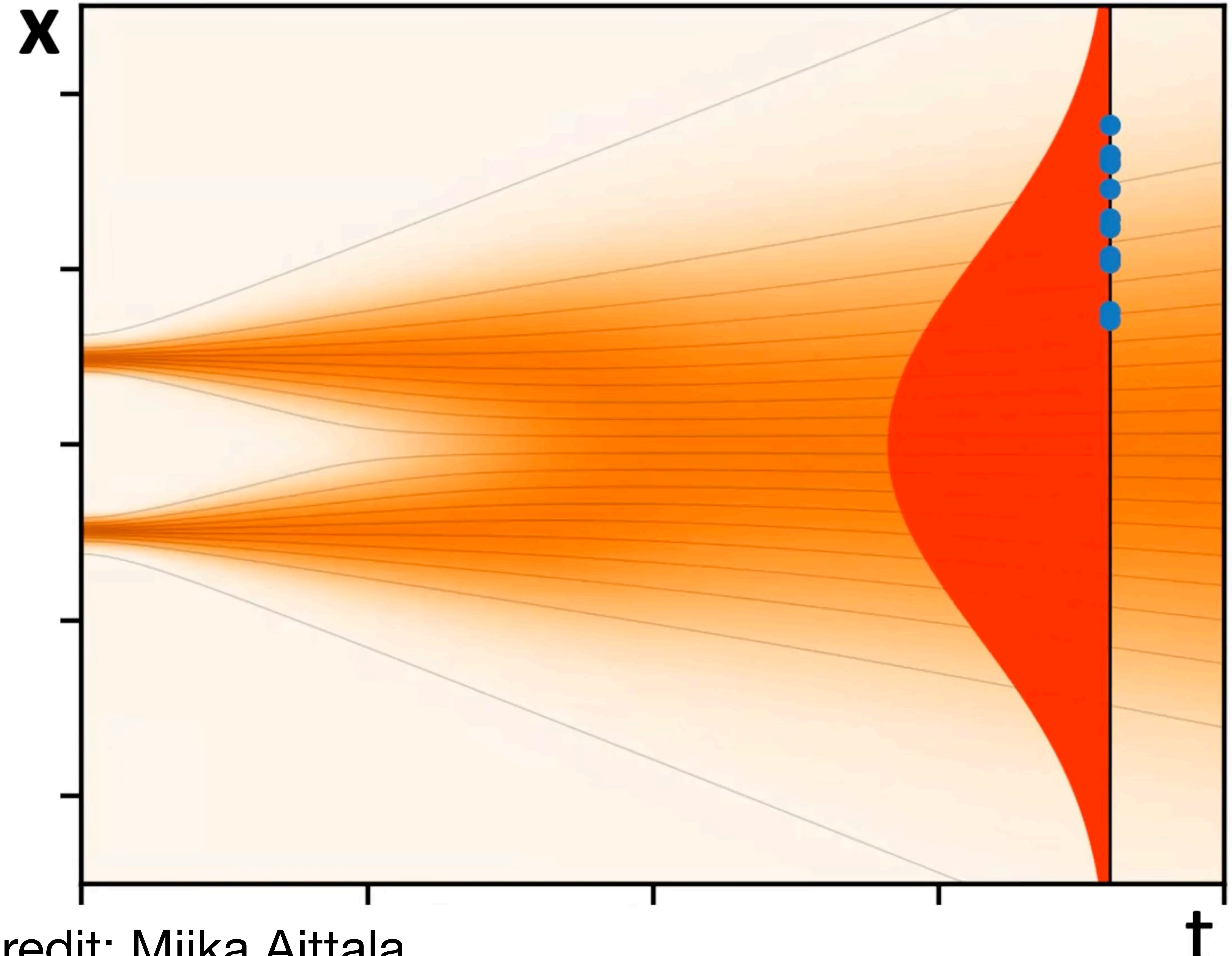

Credit: Miika Aittala

1. Truncation error: fails to approximate ideal trajectory by finite steps.

2. Model fails to approximate the marginal flow.

    1. Stochastic sampler (SDE) injects fresh noise throughout the evolution in addition to reducing the noise.

# Stochastic Sampler Helps Explore the Distribution



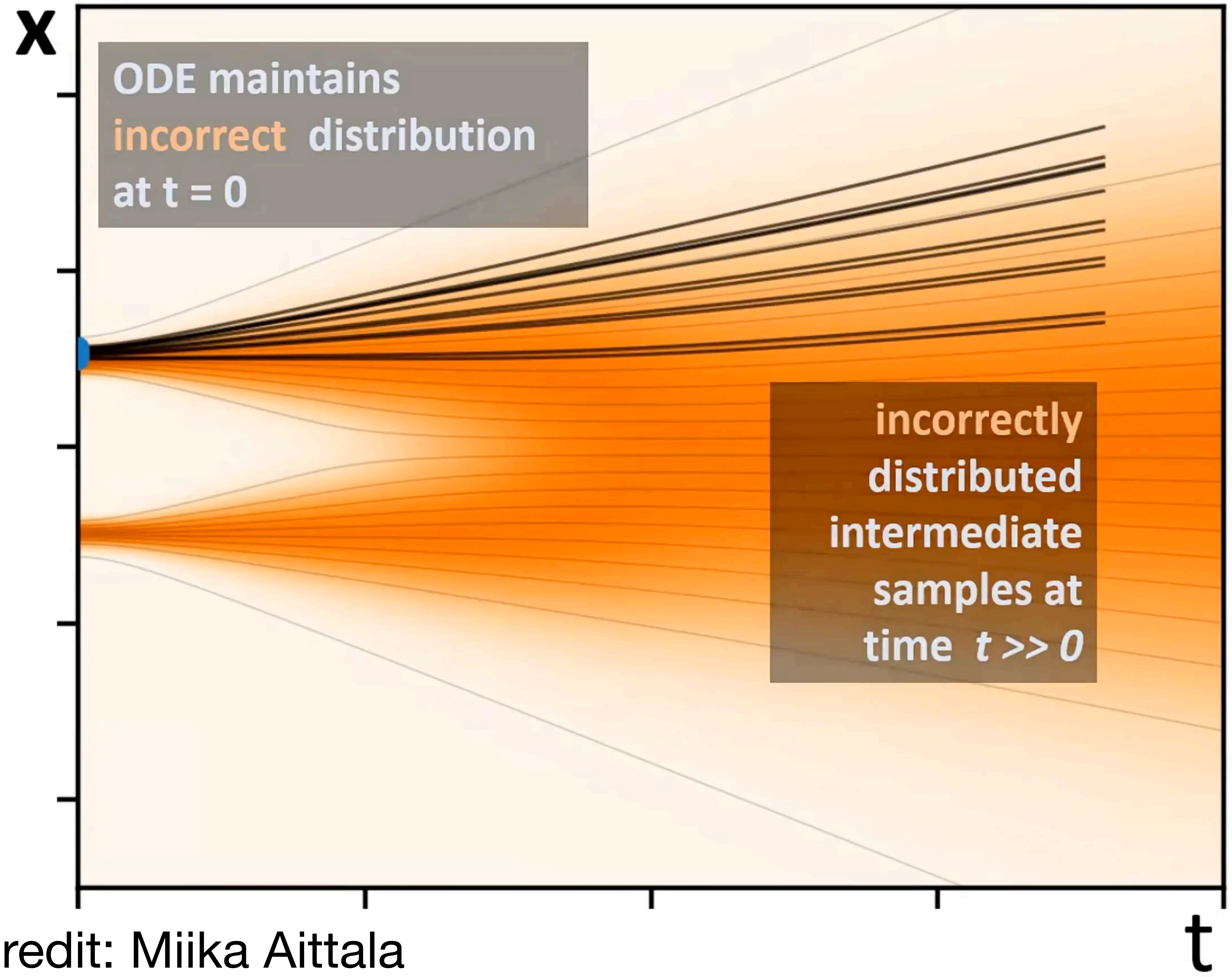

Credit: Miika Aittala

1. Truncation error: fails to approximate ideal trajectory by finite steps.

2. Model fails to approximate the marginal flow.

    1. Stochastic sampler (SDE) injects fresh noise throughout the evolution in addition to reducing the noise.

# Stochastic Sampler Helps Explore the Distribution




Credit: Miika Aittala

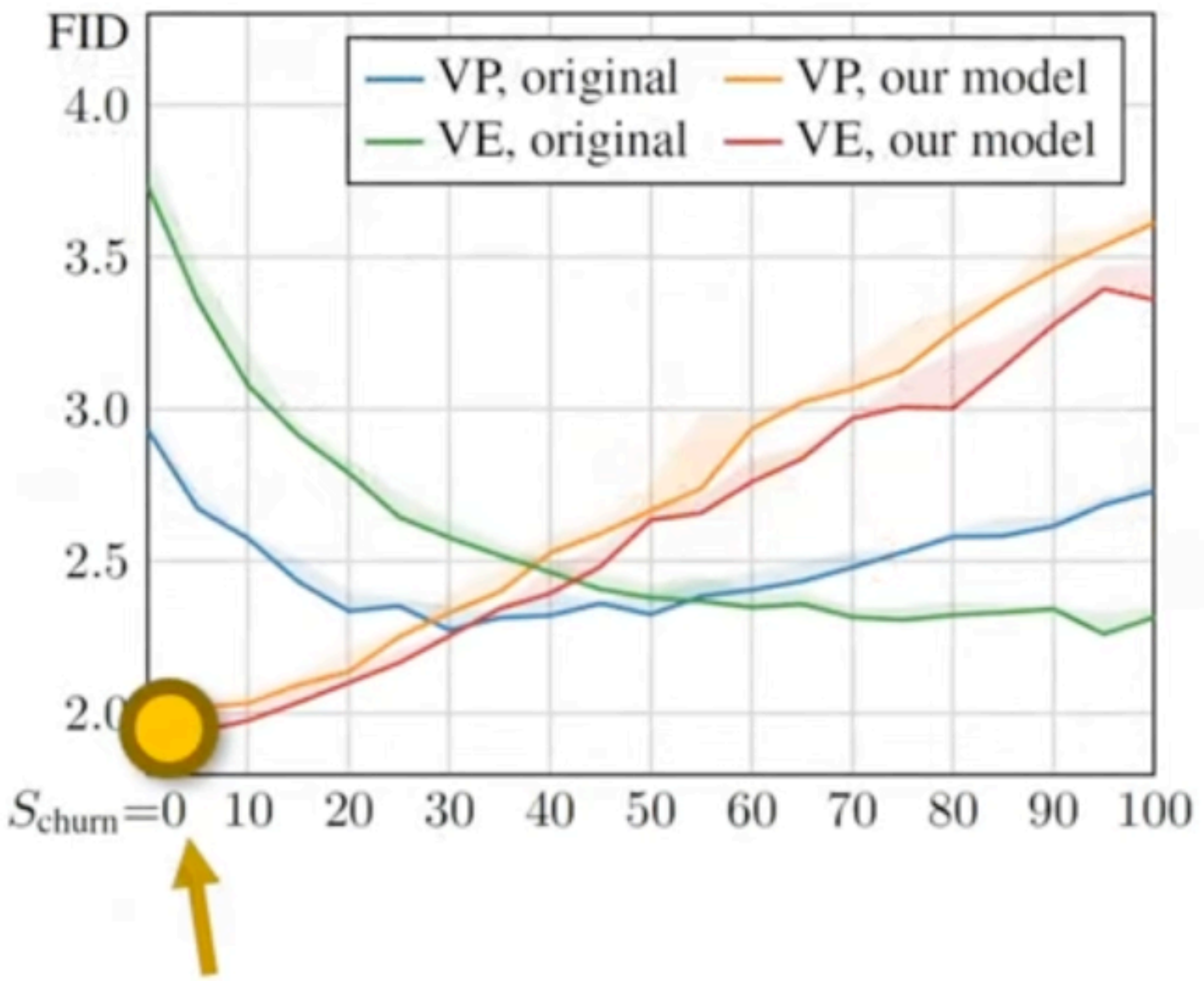
1. Truncation error: fails to approximate ideal trajectory by finite steps.

2. Model fails to approximate the marginal flow.

    1. Stochastic sampler (SDE) injects fresh noise throughout the evolution in addition to reducing the noise.

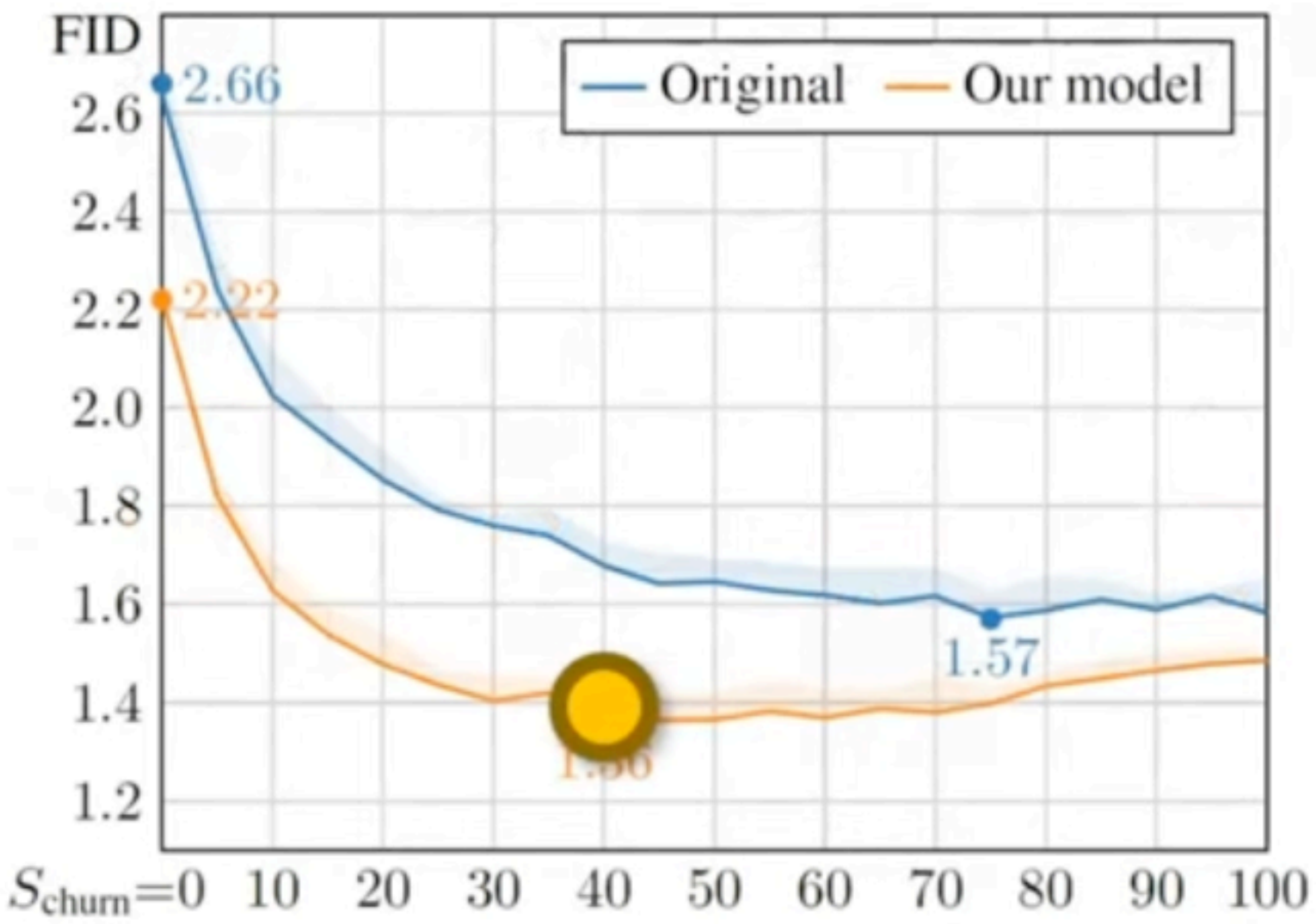
# Is stochasticity always helpful?

CIFAR-10: no



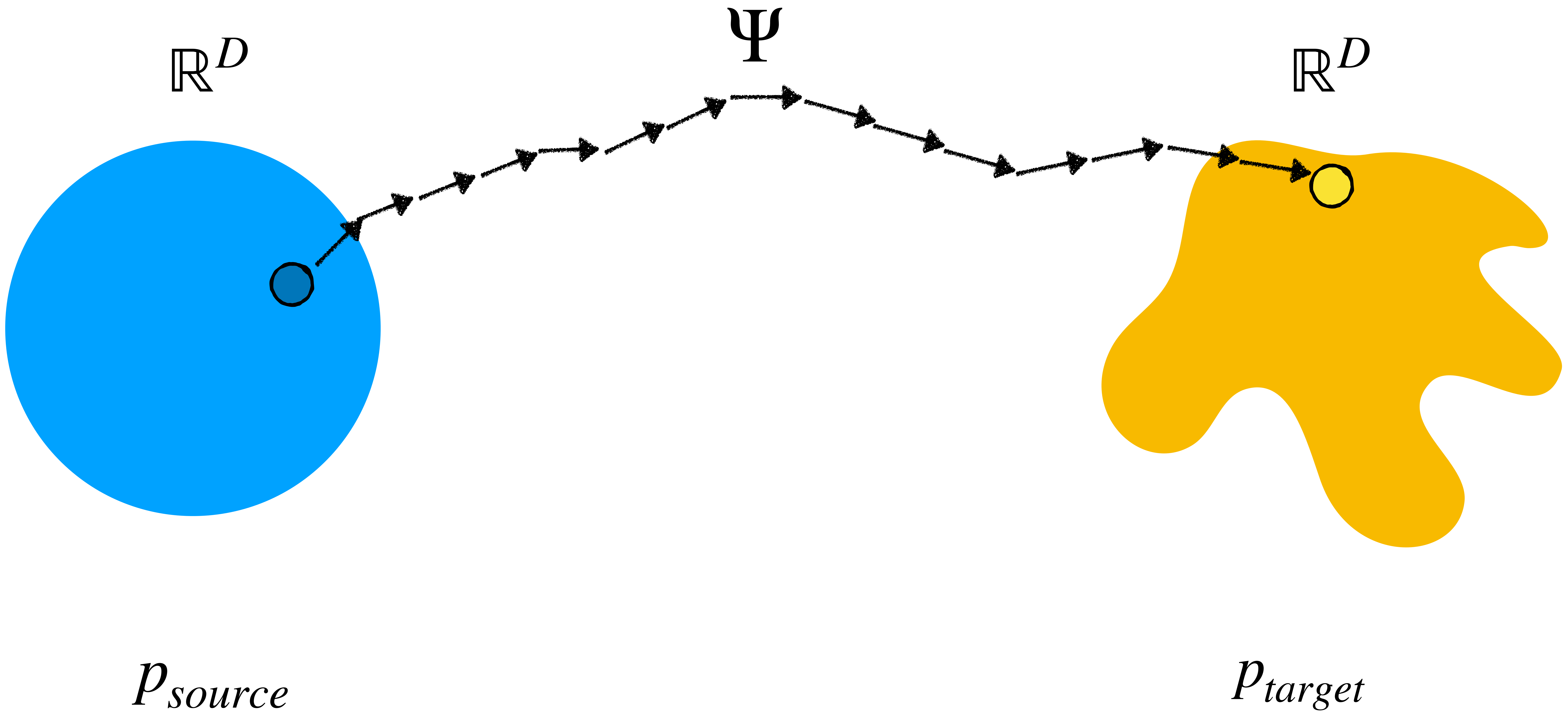

best FID at *zero* stochasticity

Imagenet: yes

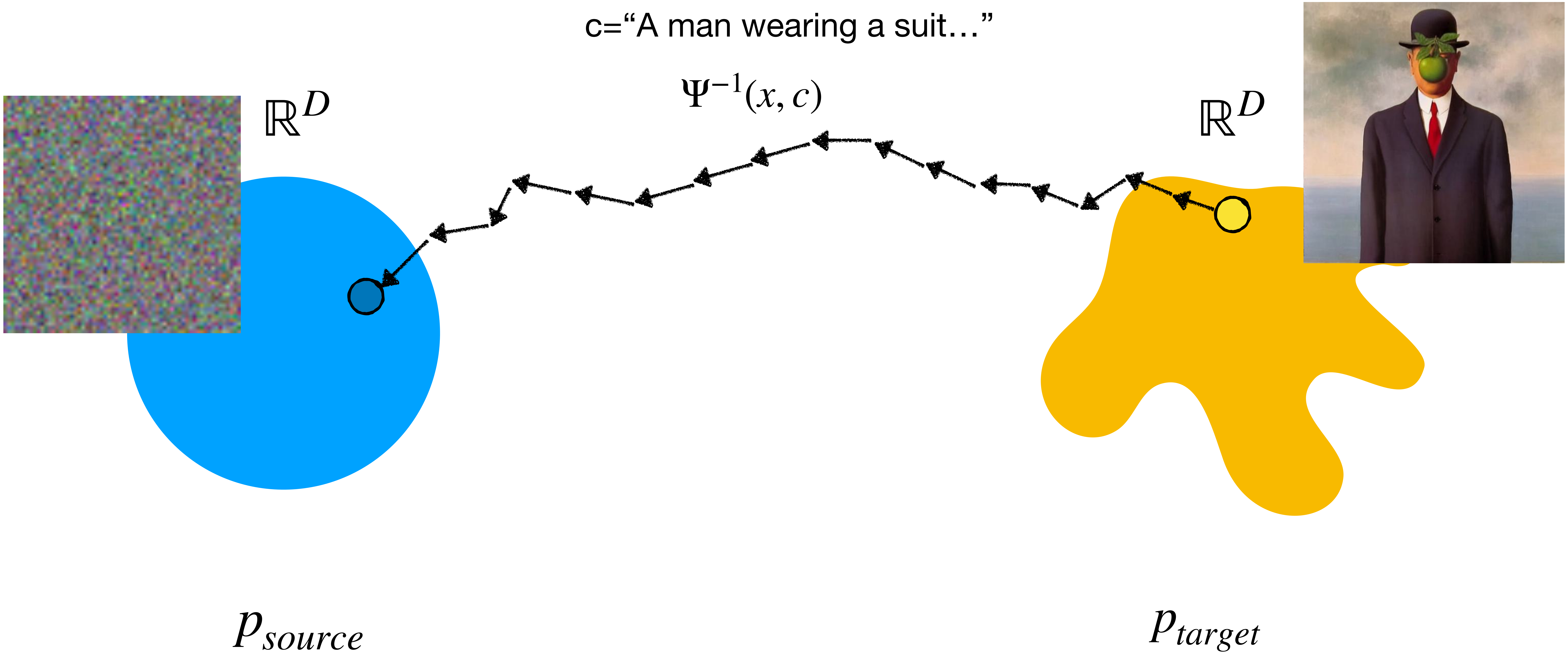



Credit: Miika Aittala

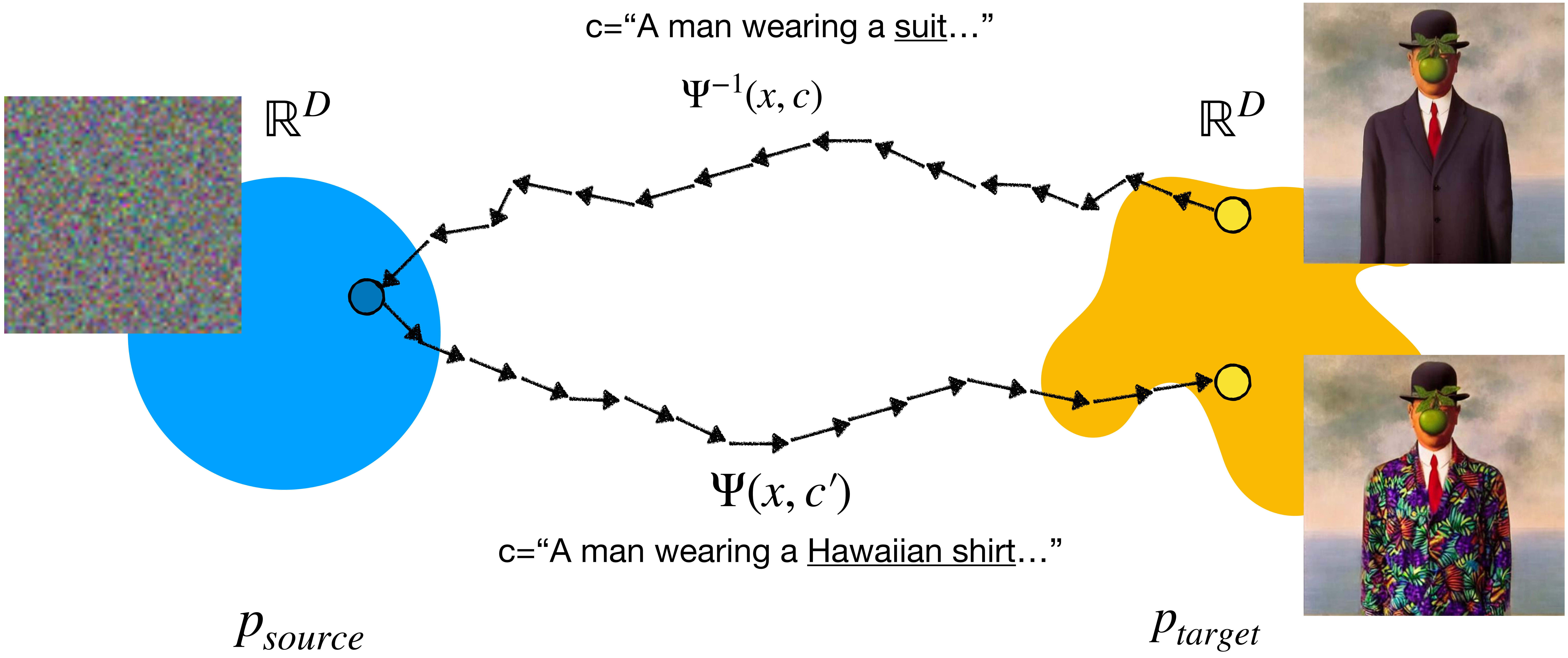
# Image Generation by Solving the Flow ODE



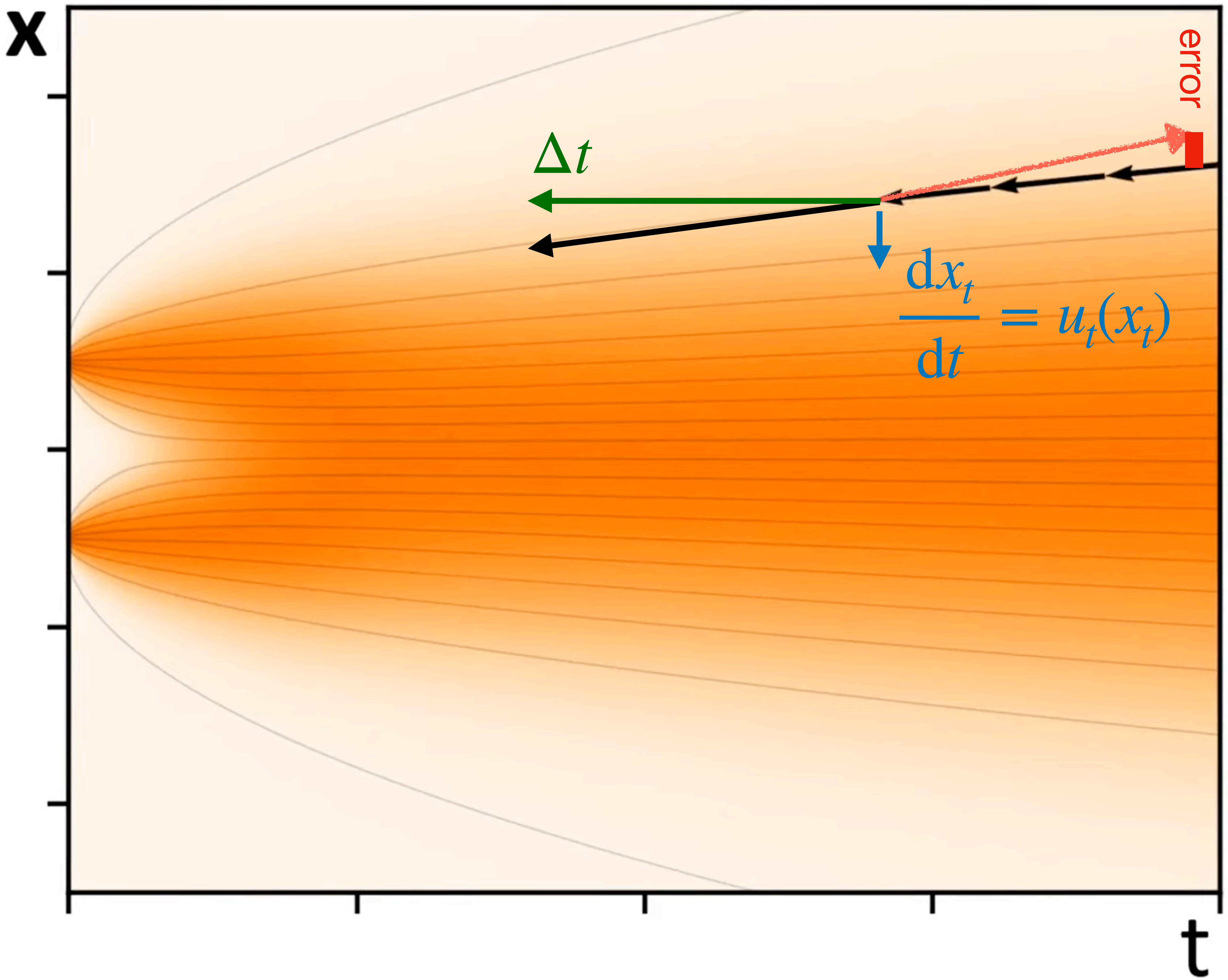
# Image Editing with Diffusion Inversion



# Image Editing with Diffusion Inversion



# Inverting a diffusion model is not easy




$$x_{t-\Delta t} = x_t - \Delta t \cdot \left.\frac{dx}{dt}\right|_{x_t,t}$$

Credit: Miika Aittala



"beret of raspberries"

# Conditioning & Guidance

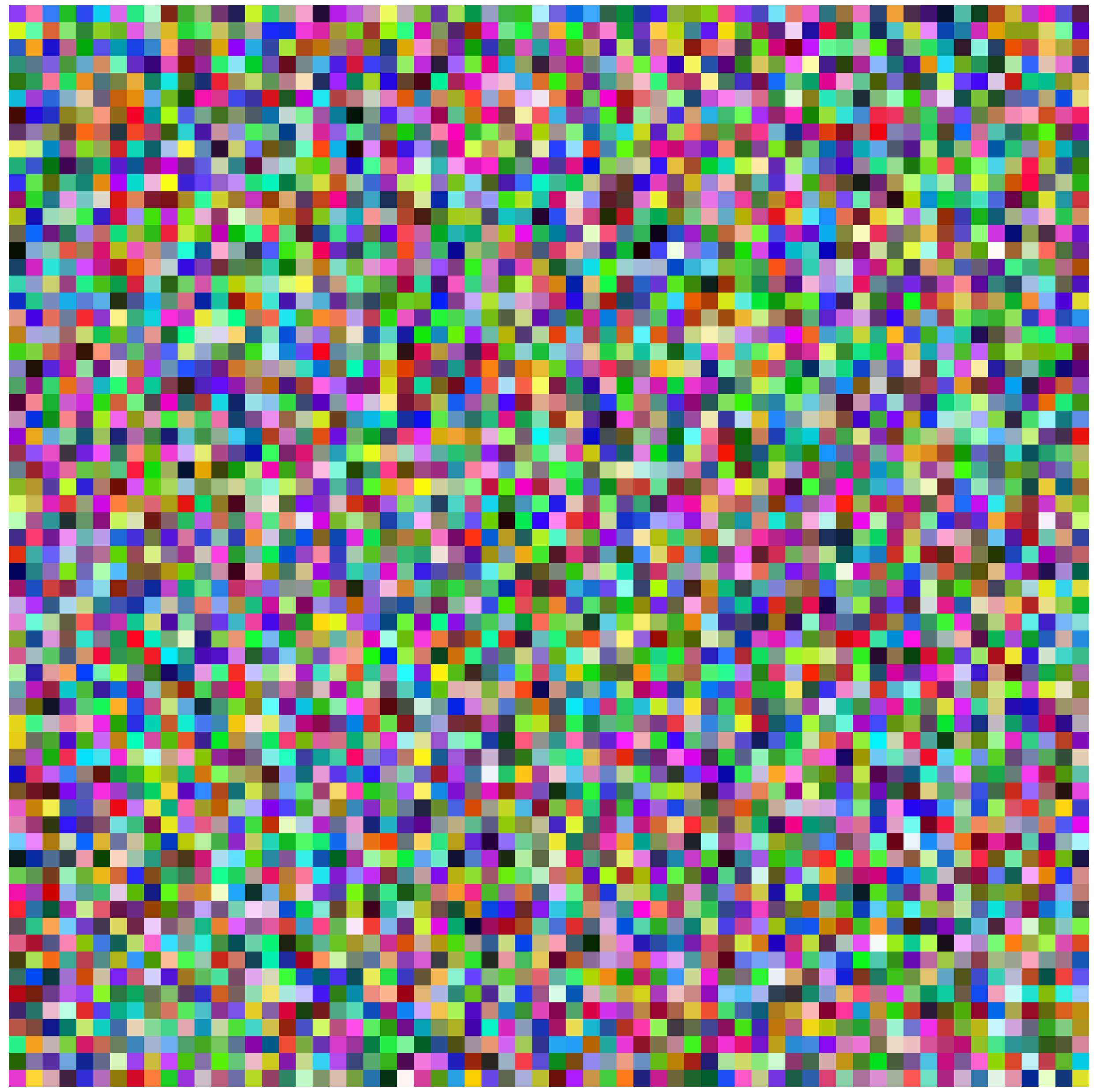
slide from Steve Seitz's [video]

raspberry beret

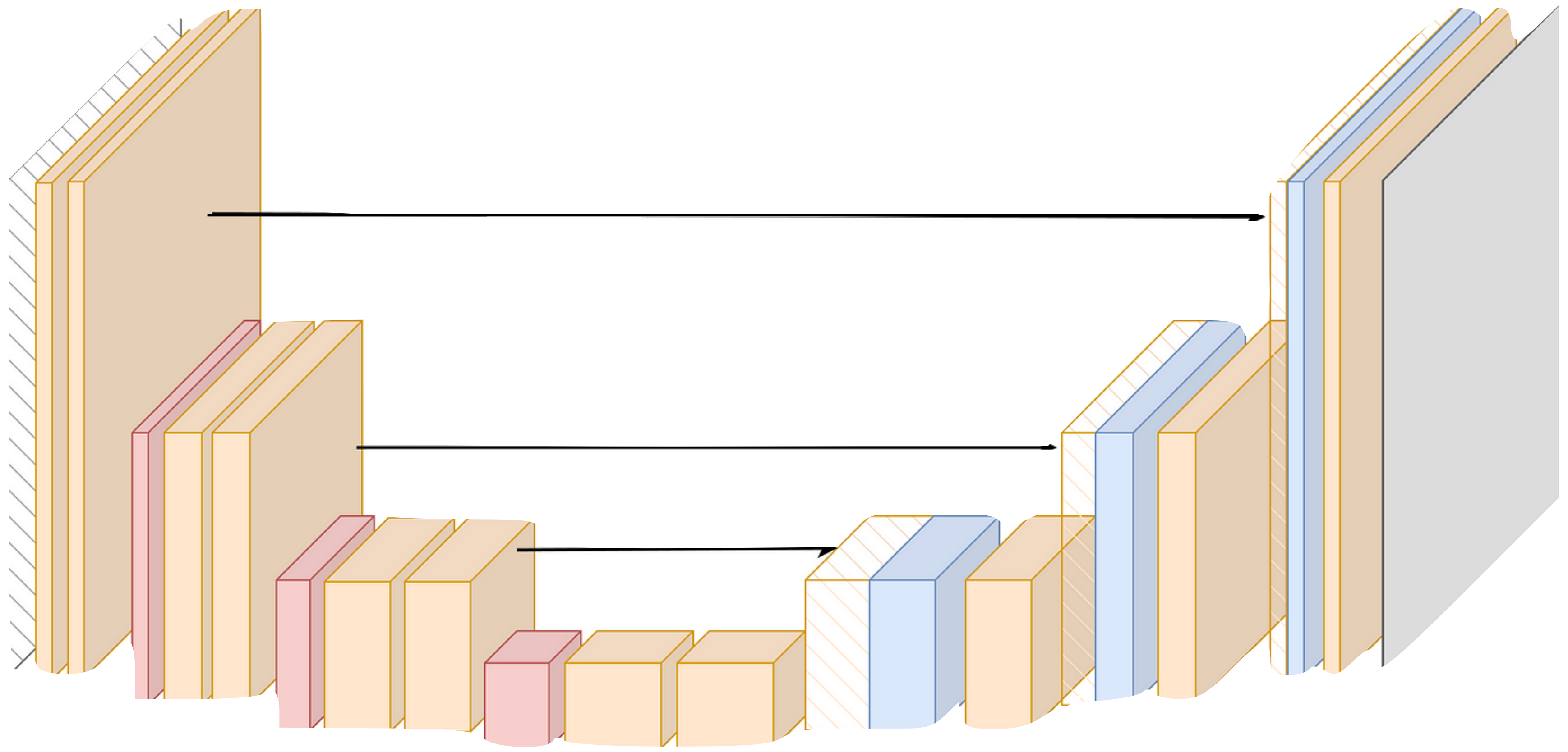
slide from Steve  Seitz's video

# Diffusion Guidance

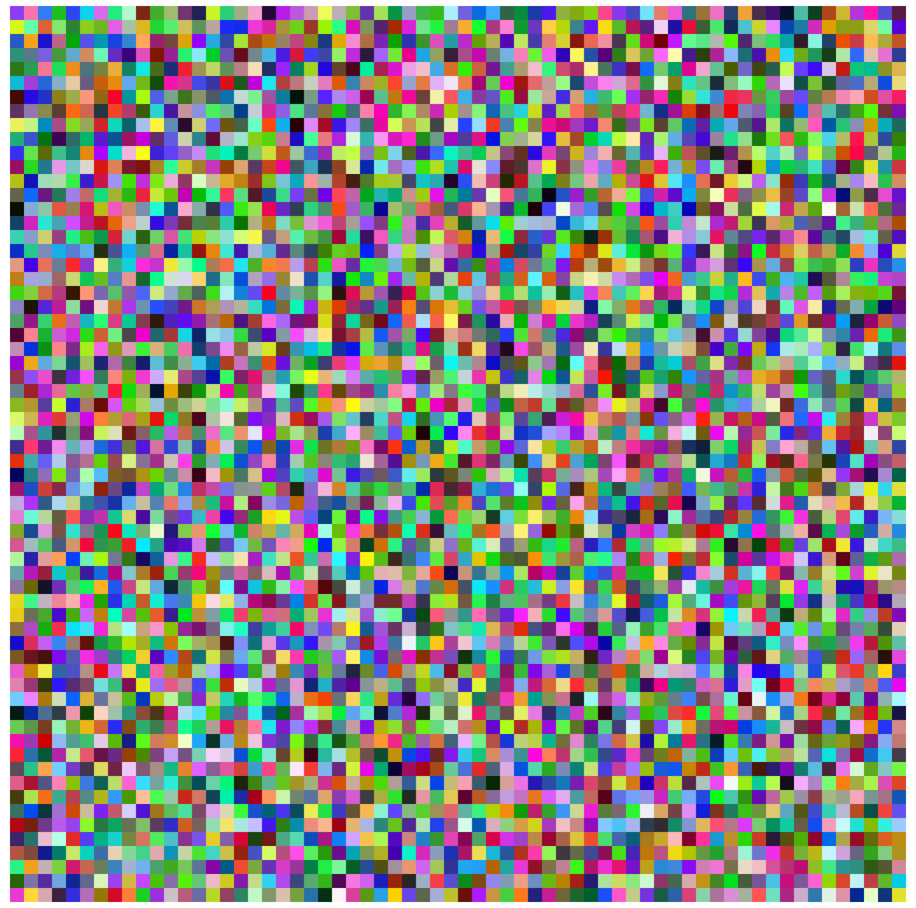
## Motivation



$\mathbf{x}_t$ Noise Image



Denoiser

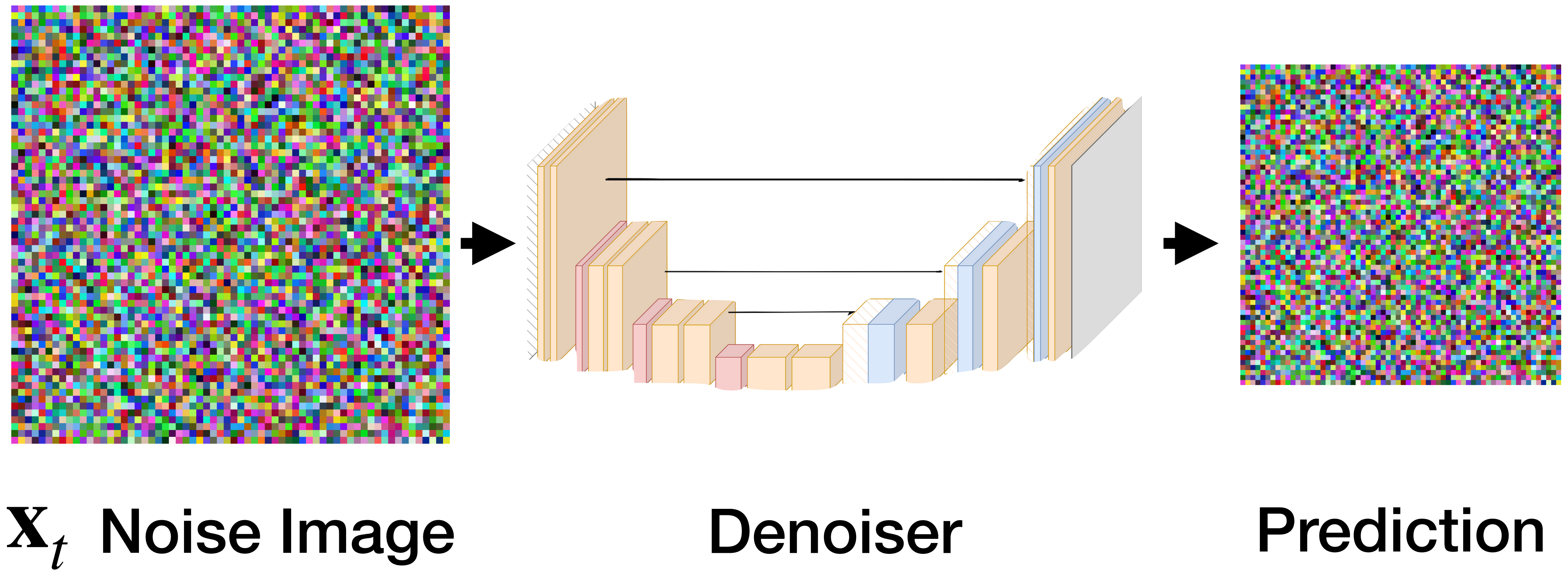


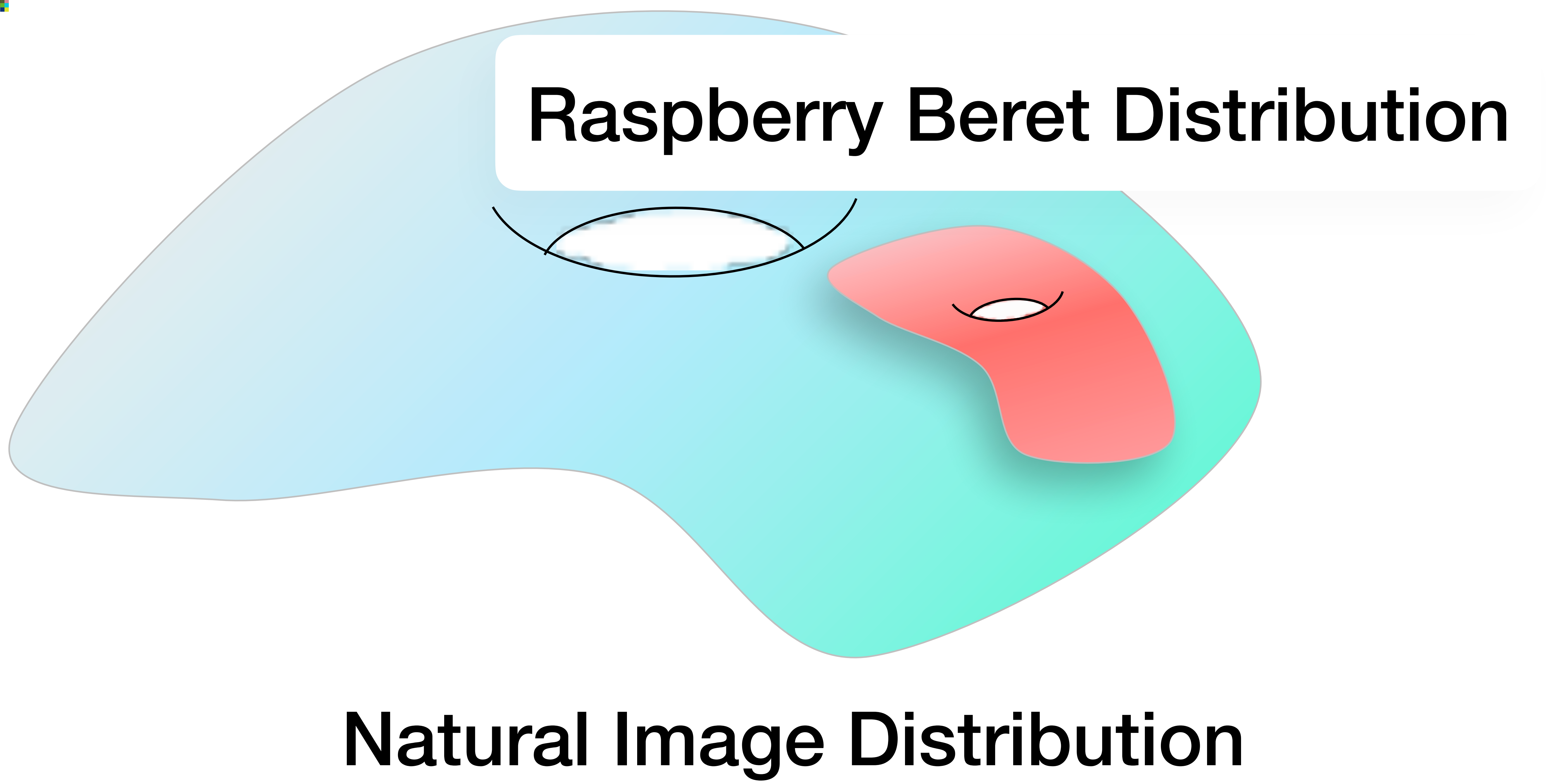
Prediction

(Flow, Epsilon etc.)

# Diffusion Guidance

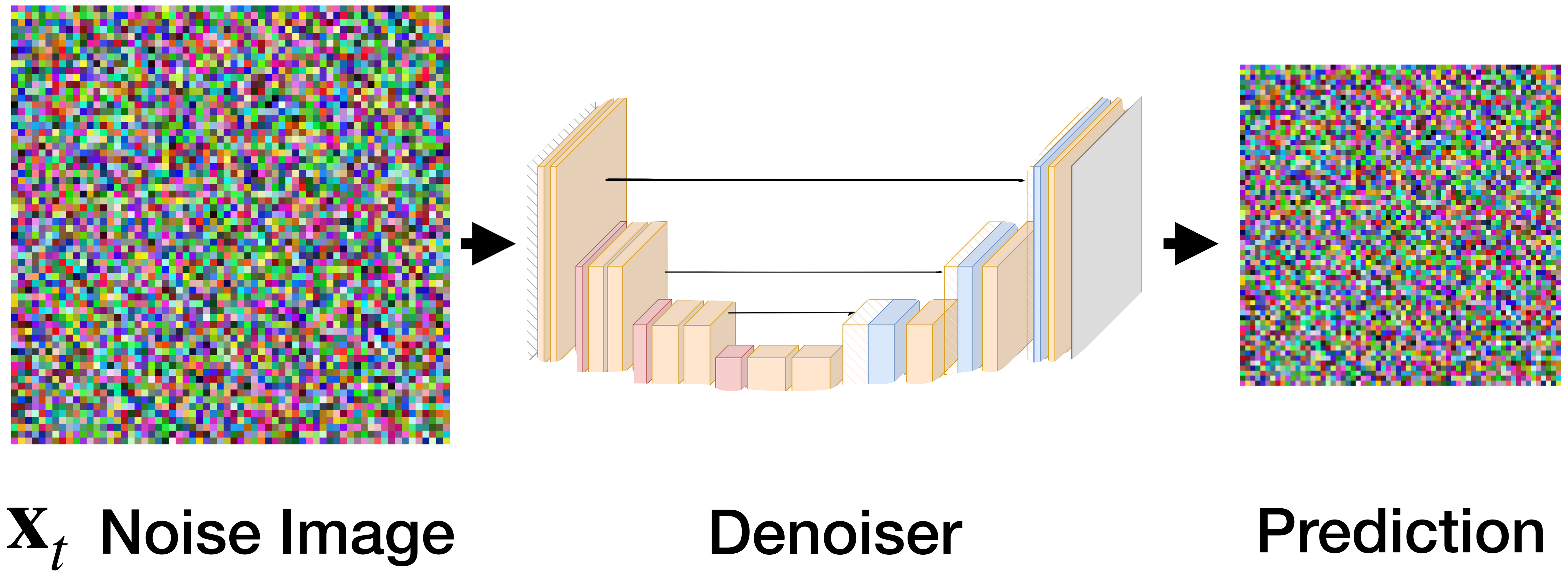
## Motivation

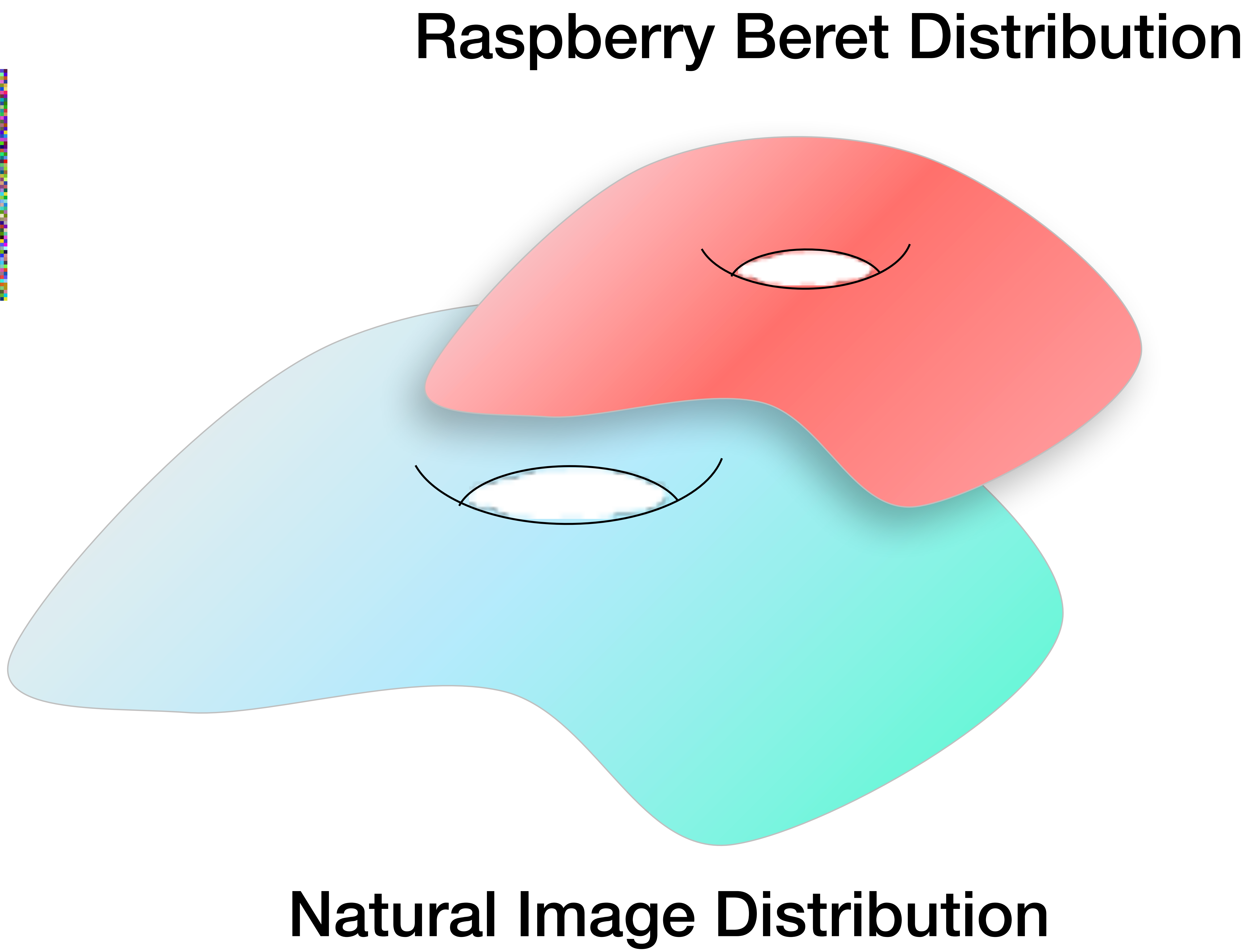


# Diffusion Guidance

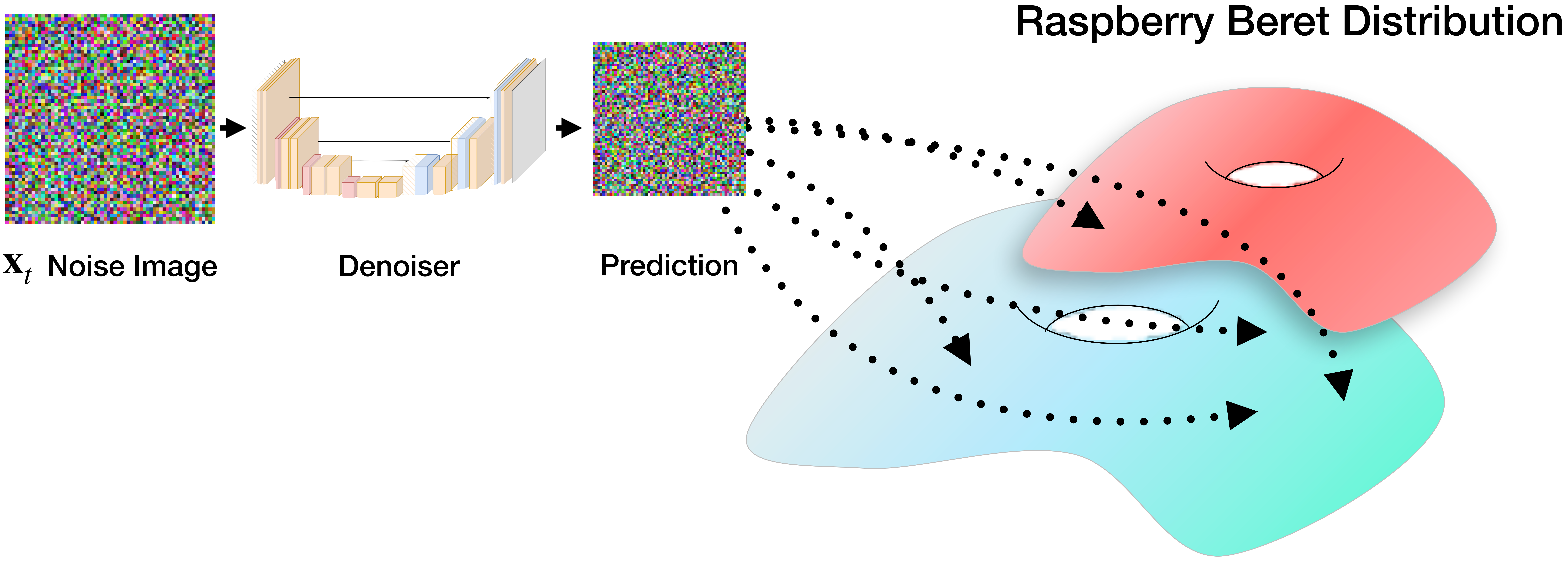
## Motivation



# Diffusion Guidance

## Motivation




Unguided Diffusion Isn't Too Useful!

# Flux Pro Unguided Samples

## Imitates Distribution of Internet Training Data

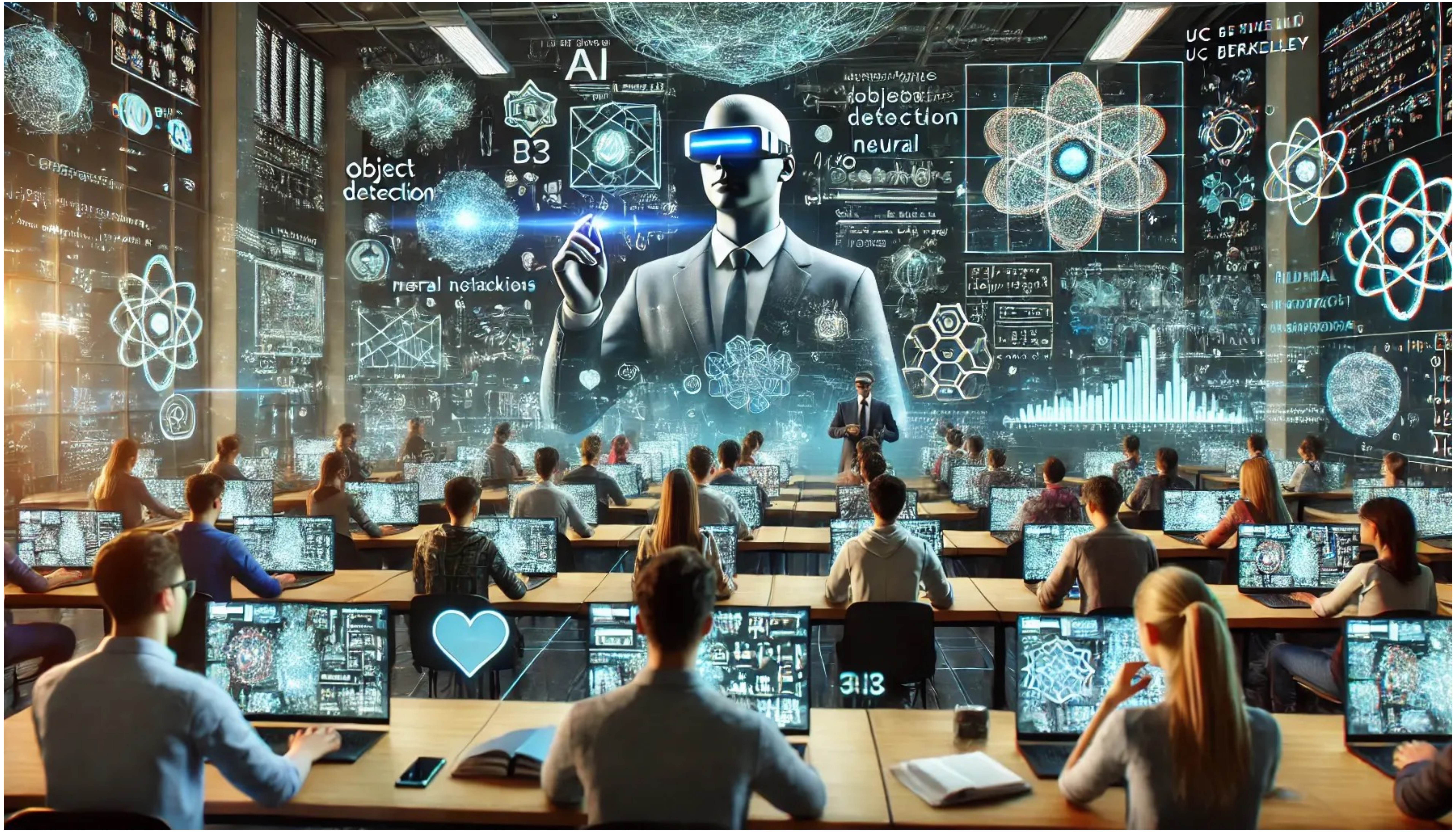





# Flux Pro Guided Samples

"beret of raspberries"


Generate a photo of a Berkeley computer vision class

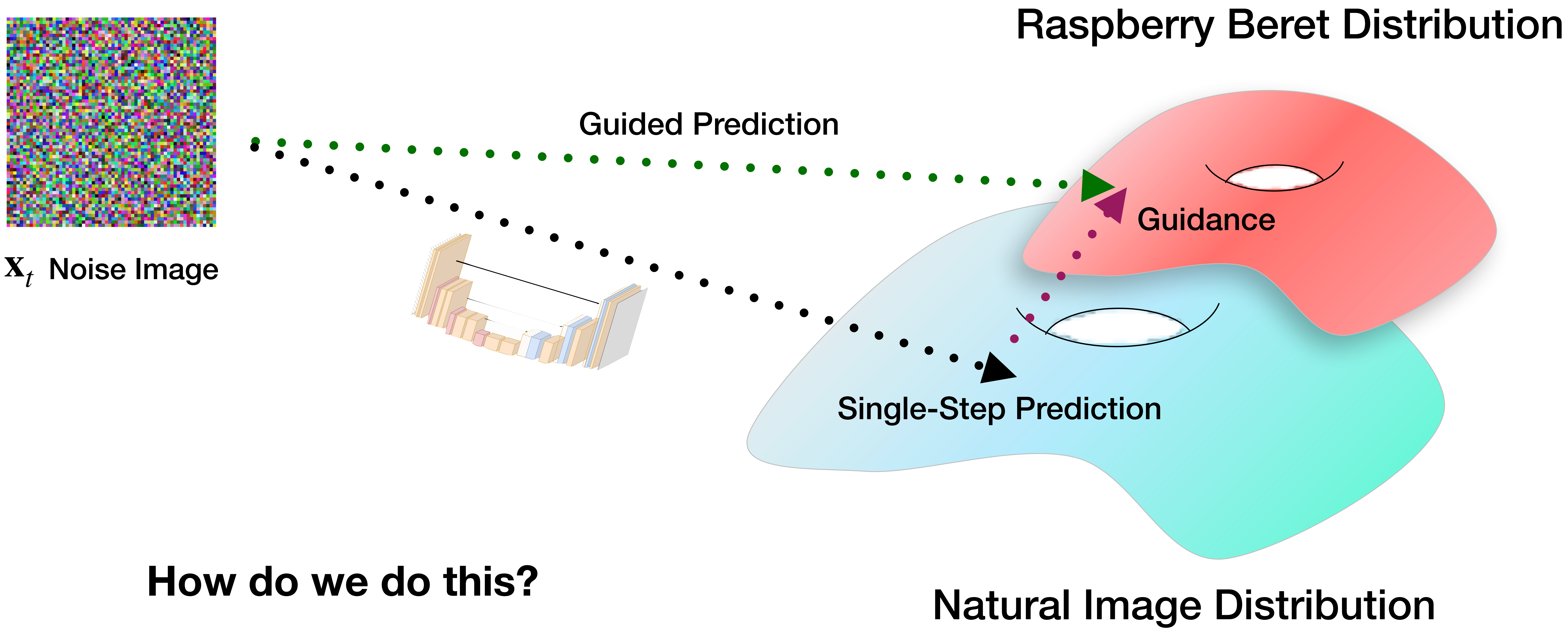



Make it more epic


Make it THE MOST EPIC computer vision class that you can ever think of

# Diffusion Guidance

## Push Toward a Conditional Mode




**How do we do this?**

# Two Approaches

**Original**

## Classifier Guidance

Guide with a pretrained classifier.

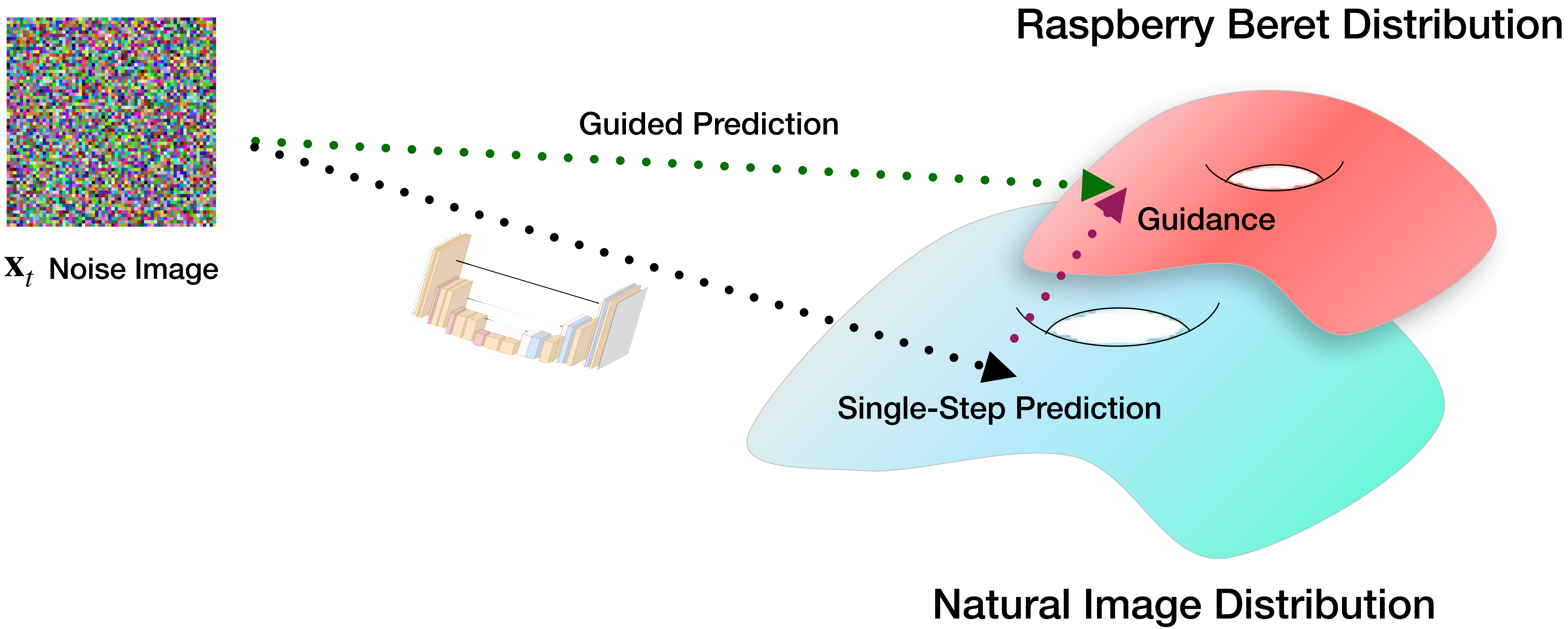
**Current**

## Classifier-Free Guidance

Guide a diffusion model with itself.

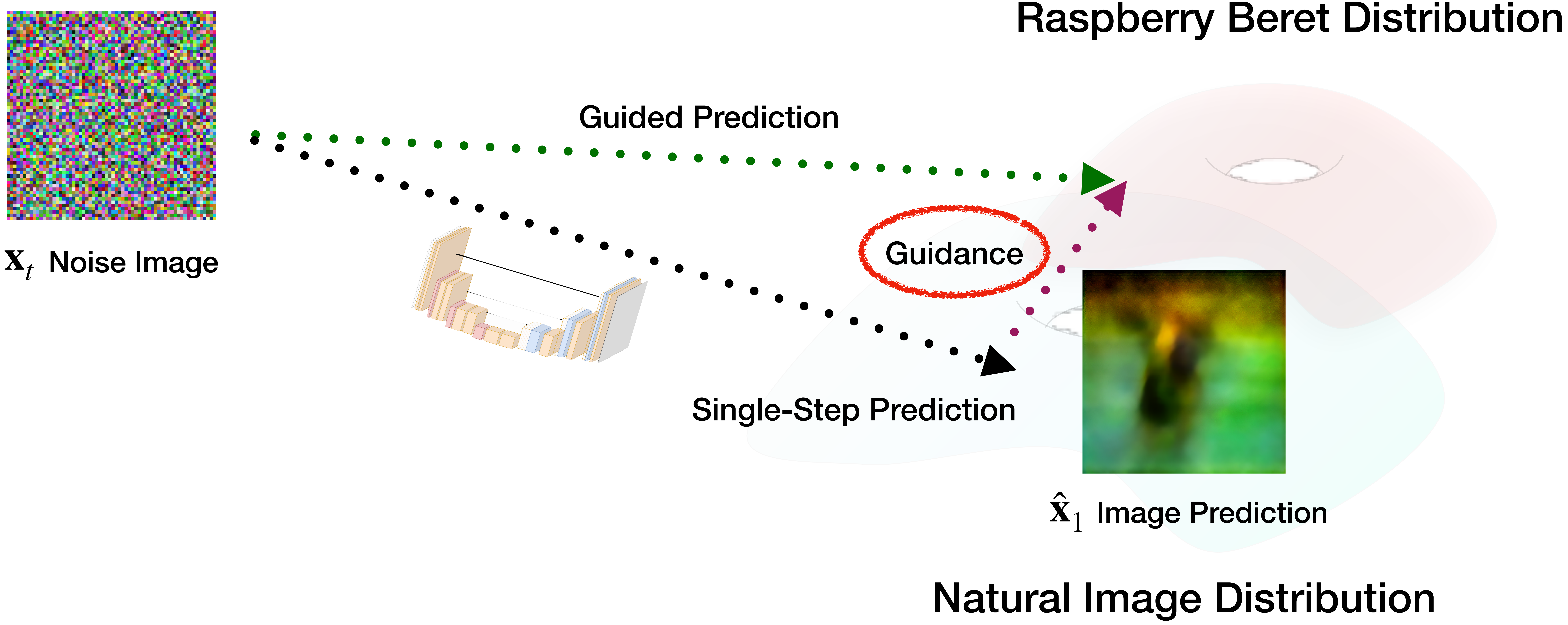
# Diffusion Guidance

## Push Toward a Conditional Mode



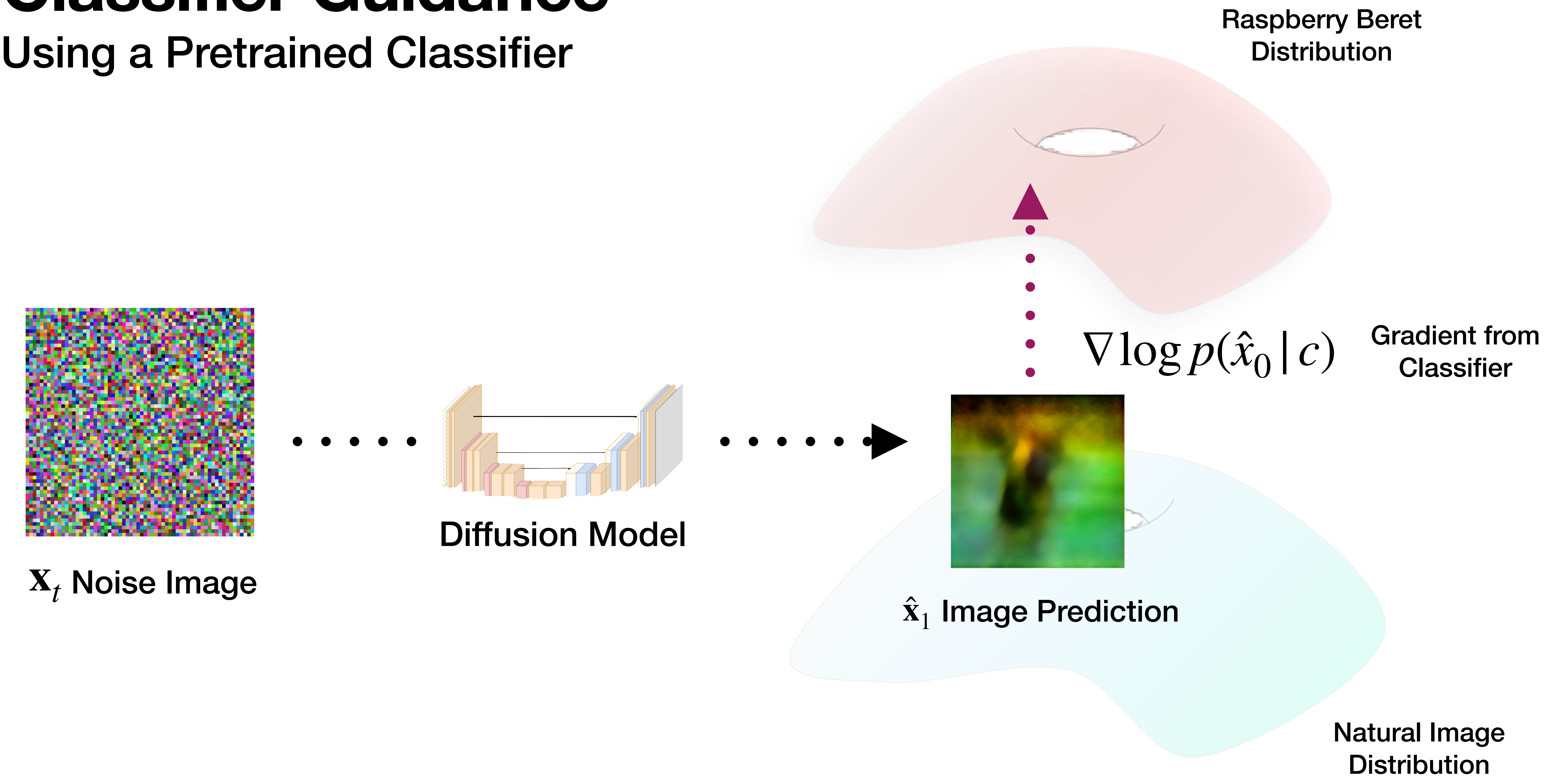
# Classifier Guidance

## Using a Pretrained Classifier

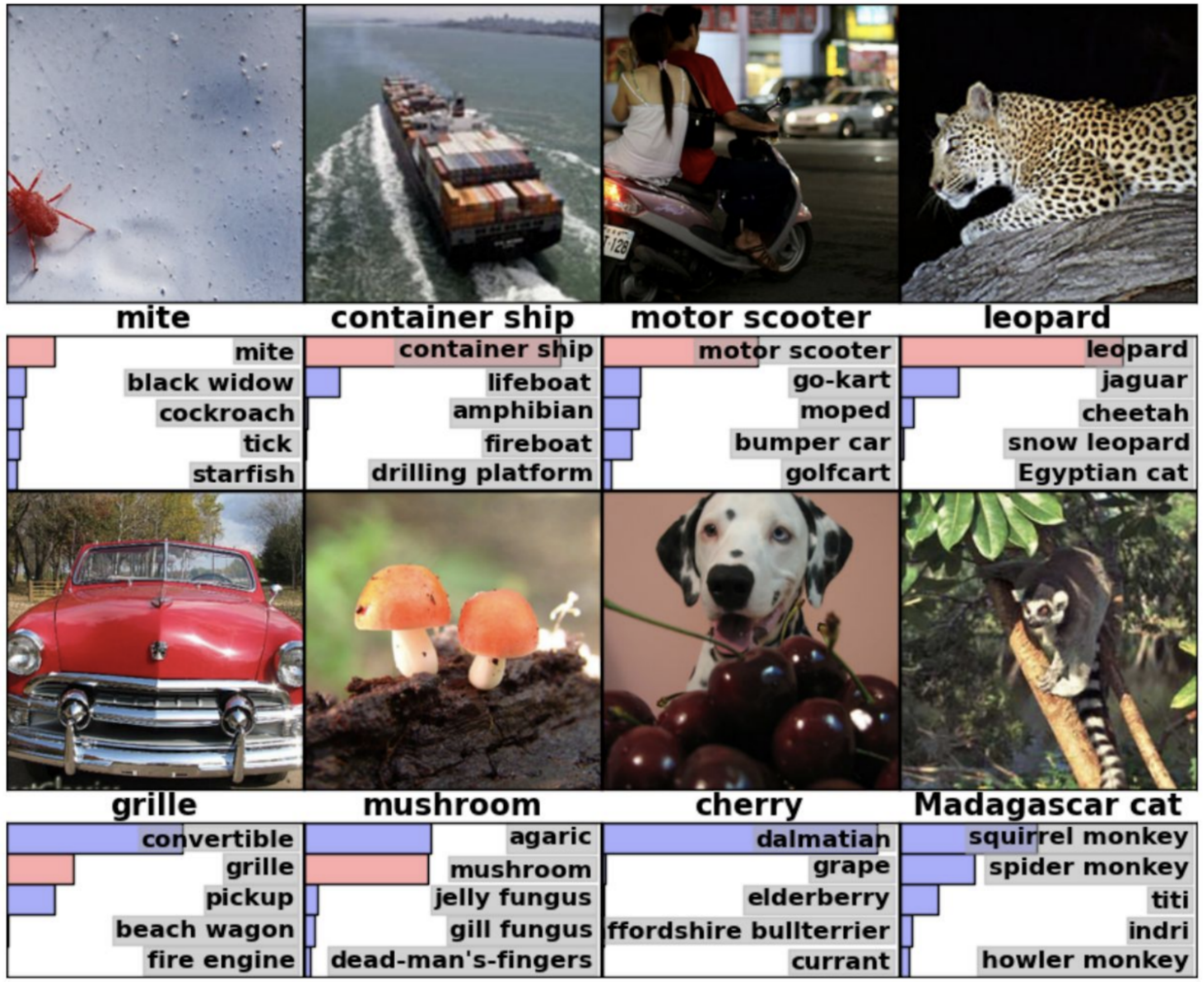


# Classifier Guidance

## Using a Pretrained Classifier



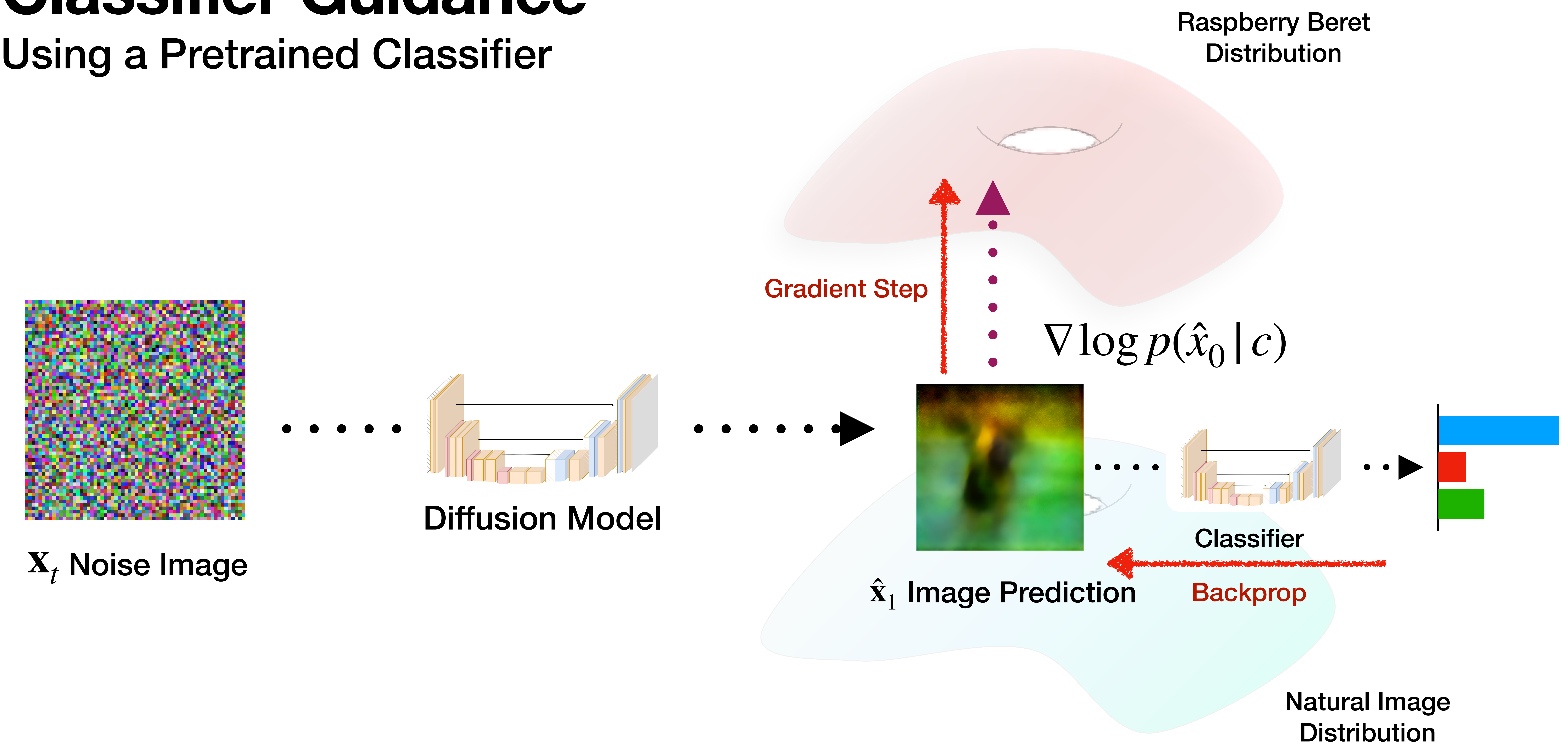
# For Example…



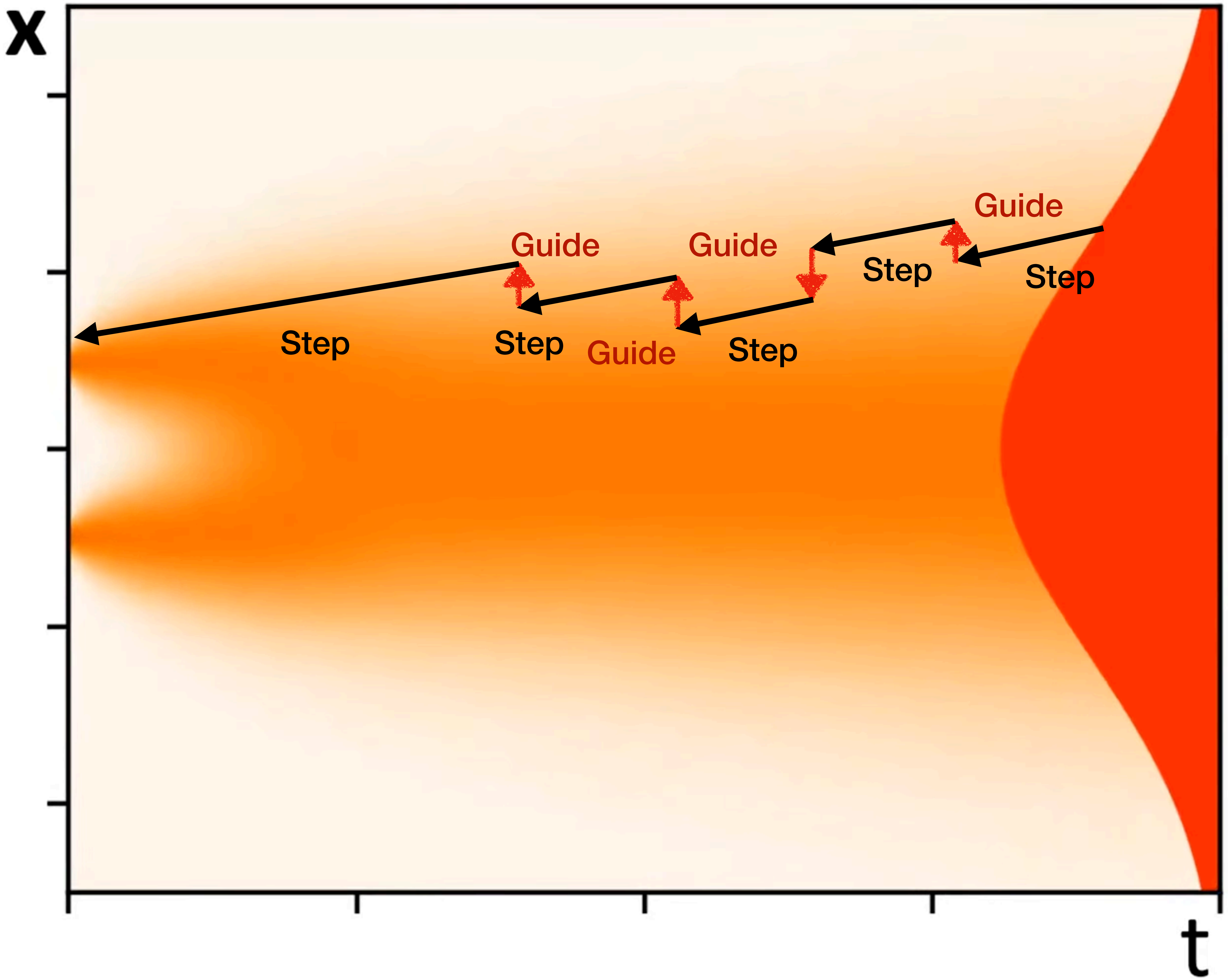

IMAGENET

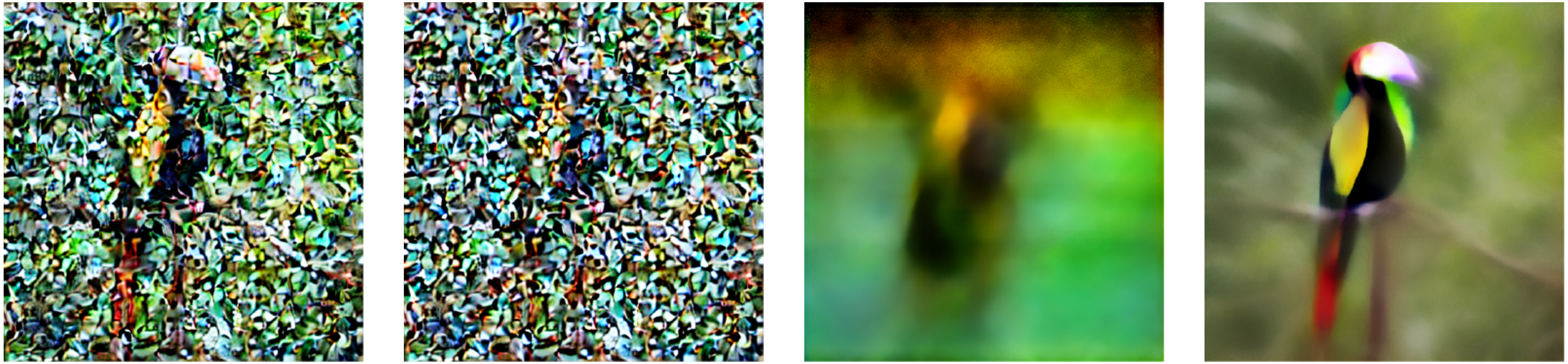
# Classifier Guidance

## Using a Pretrained Classifier



# Sampling ODE View





**Classify These Images!**

# Classifier Guidance Pathologies

## Intermediate Diffusion Steps are O.O.D. for Classifier


ImageNet Training Data

$\mathbf{x}_t$ Noise Image $\hat{\mathbf{x}}_1$ Image Prediction

# Two Approaches

**Original**

## Classifier Guidance

Guide with a pretrained classifier.

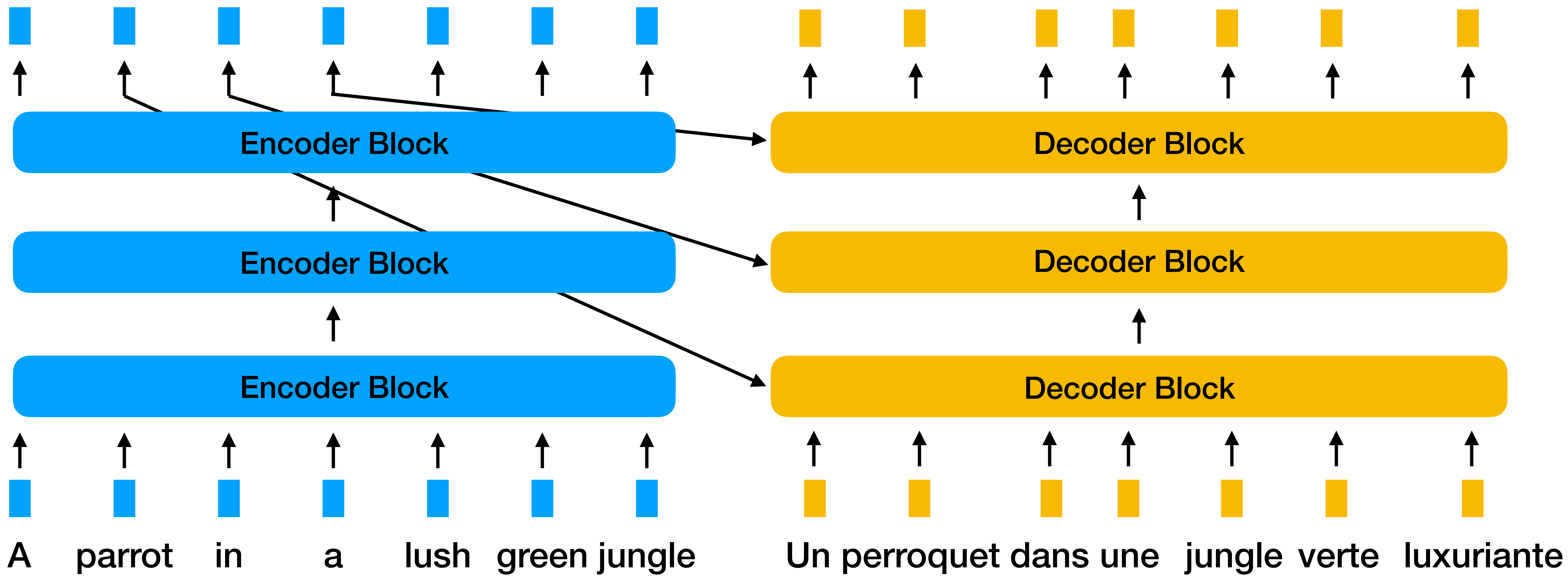
**Current**

## Classifier-Free Guidance

Guide a diffusion model with itself.

# Encoder-Decoder Architecture

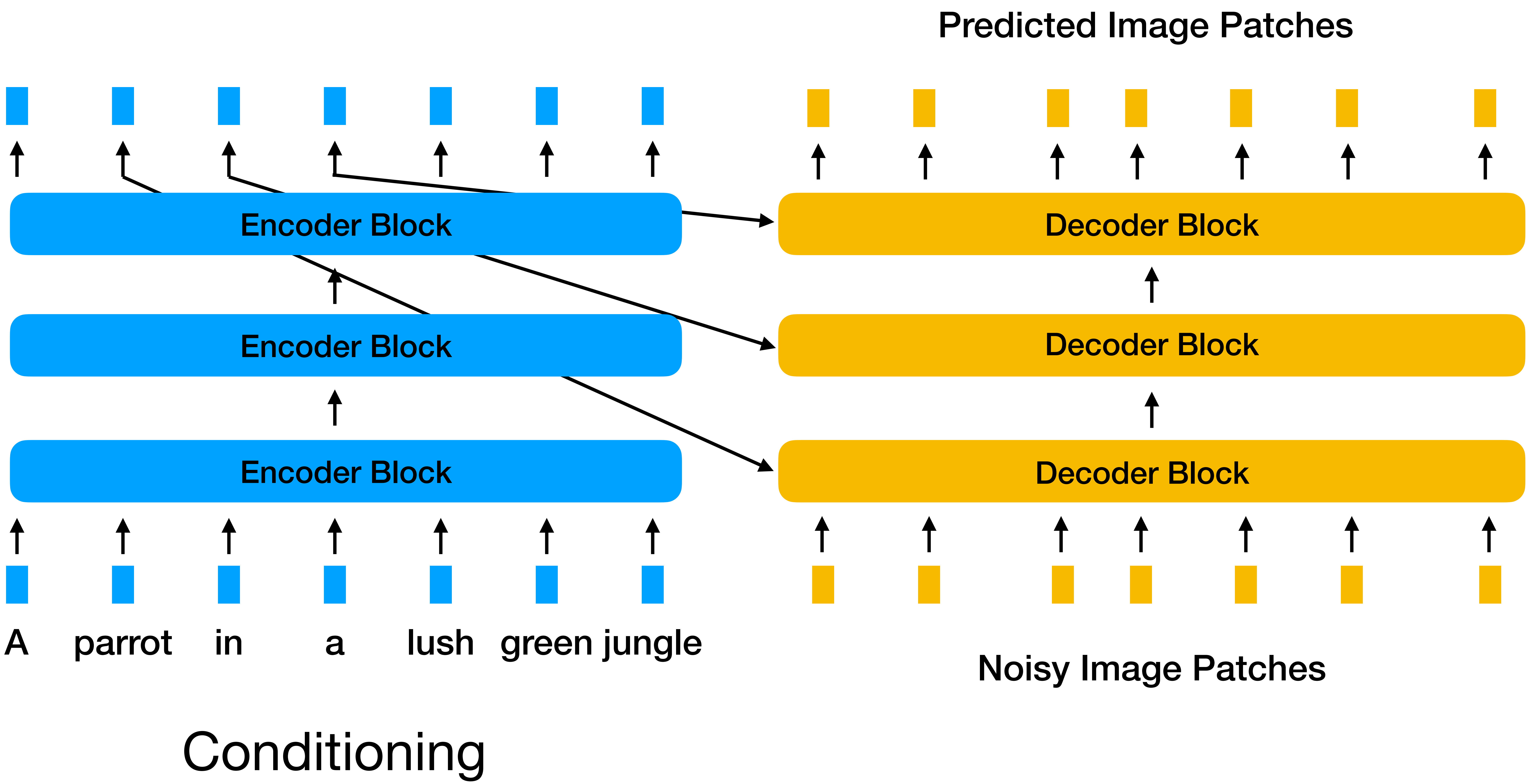
## Attention is All You Need

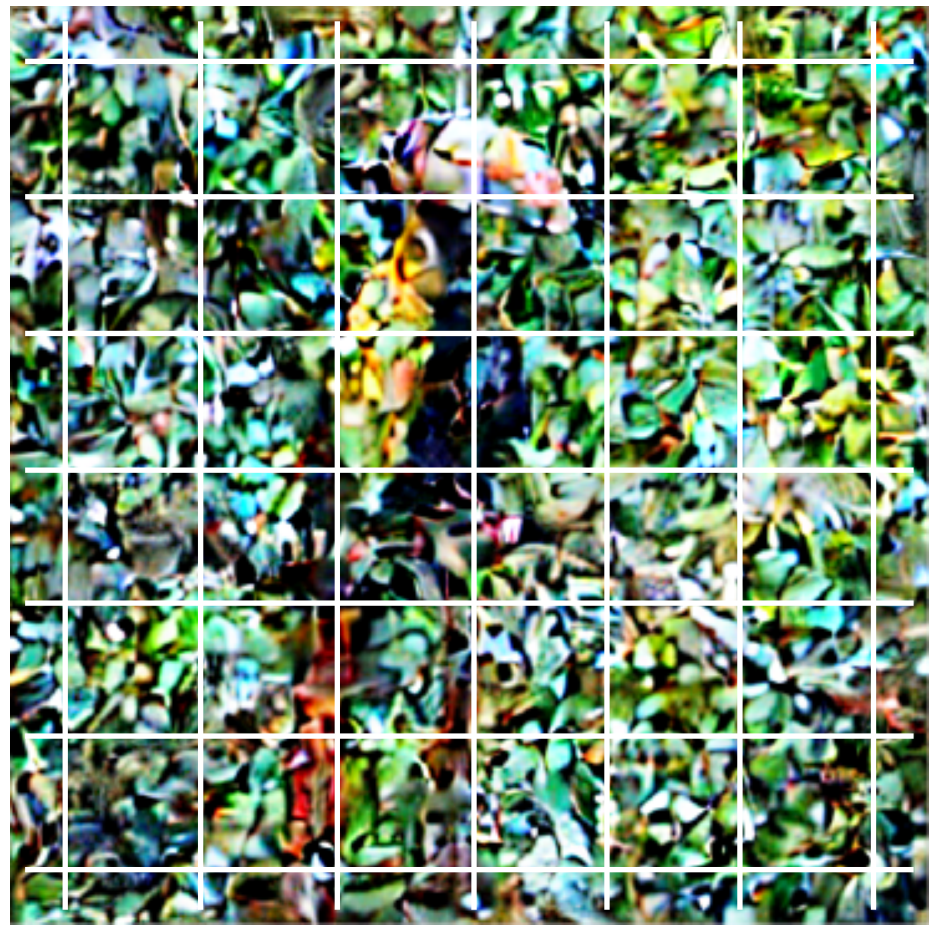



Adapted from HuggingFace Transformers-based Encoder-Decoder Models

# Diffusion Transformer Architecture

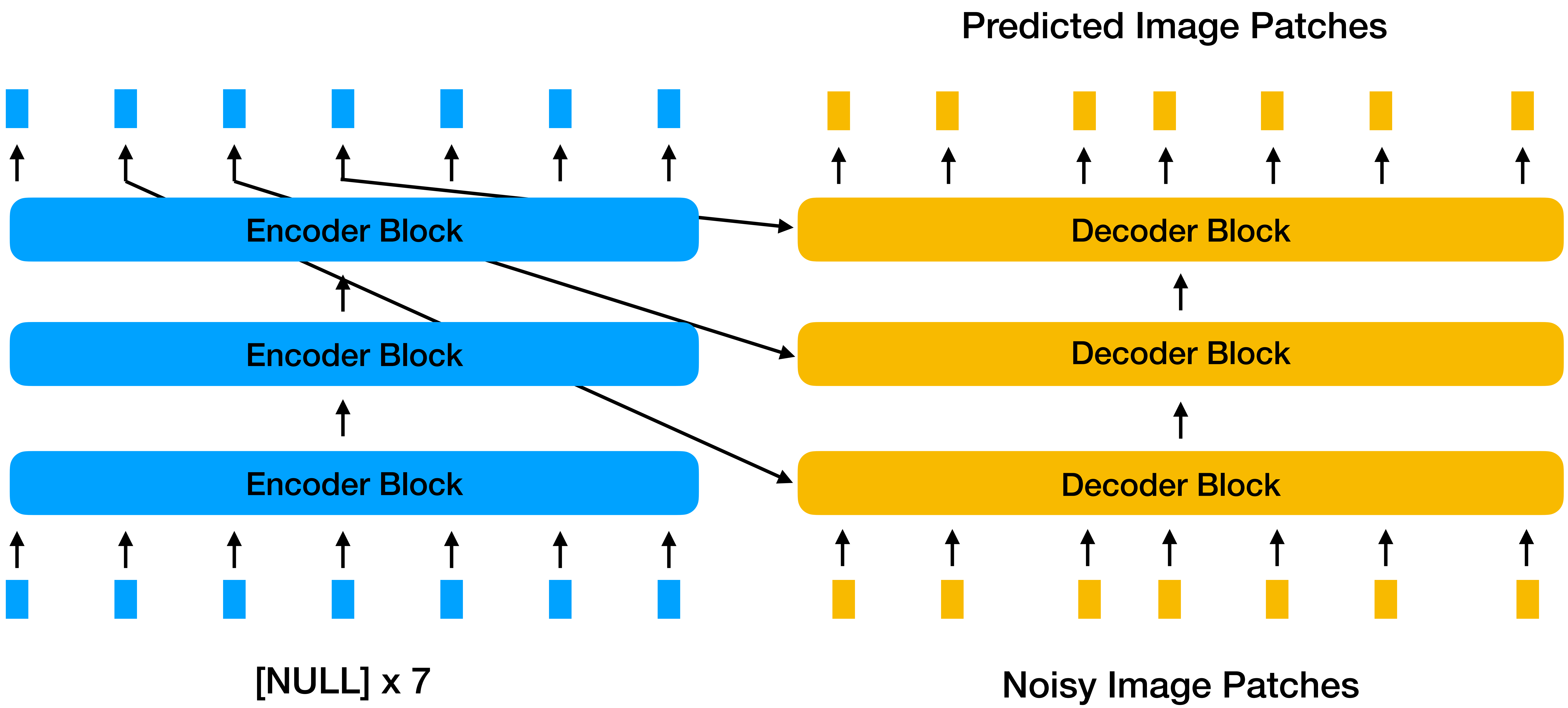
## DiT, PixArt Alpha, MMDiT, etc.

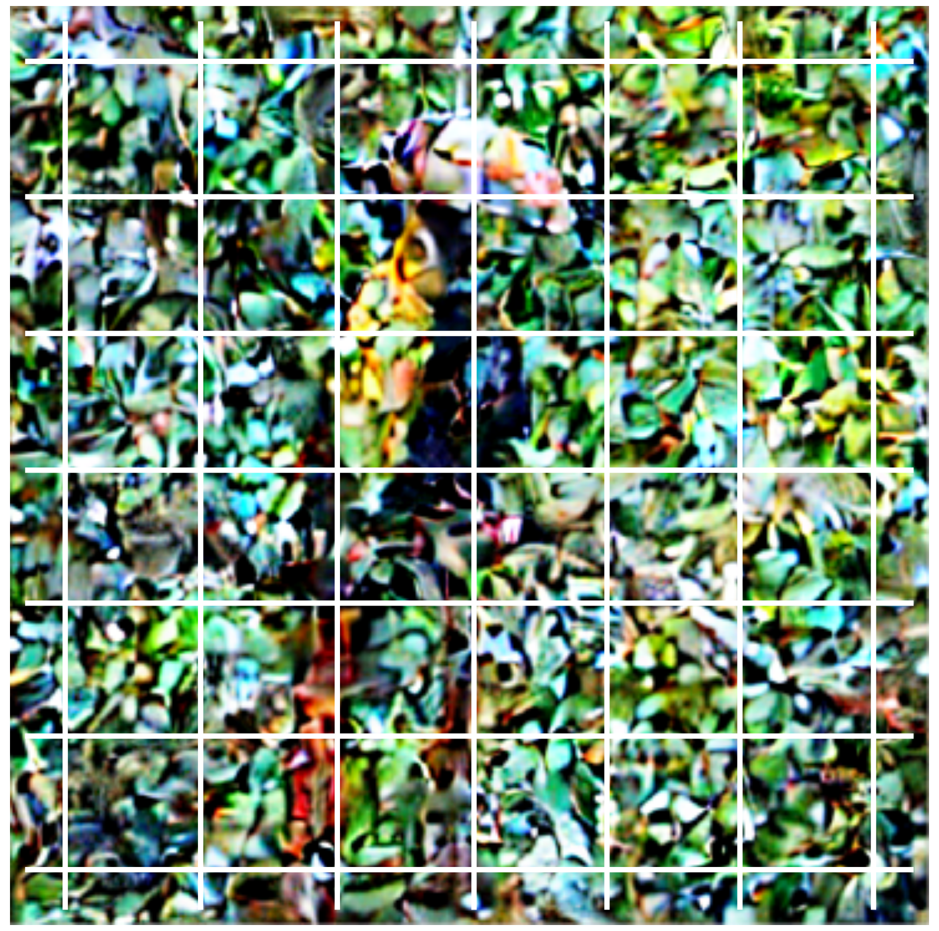


# Dropout Conditioning

## Maybe 30% of Training Samples



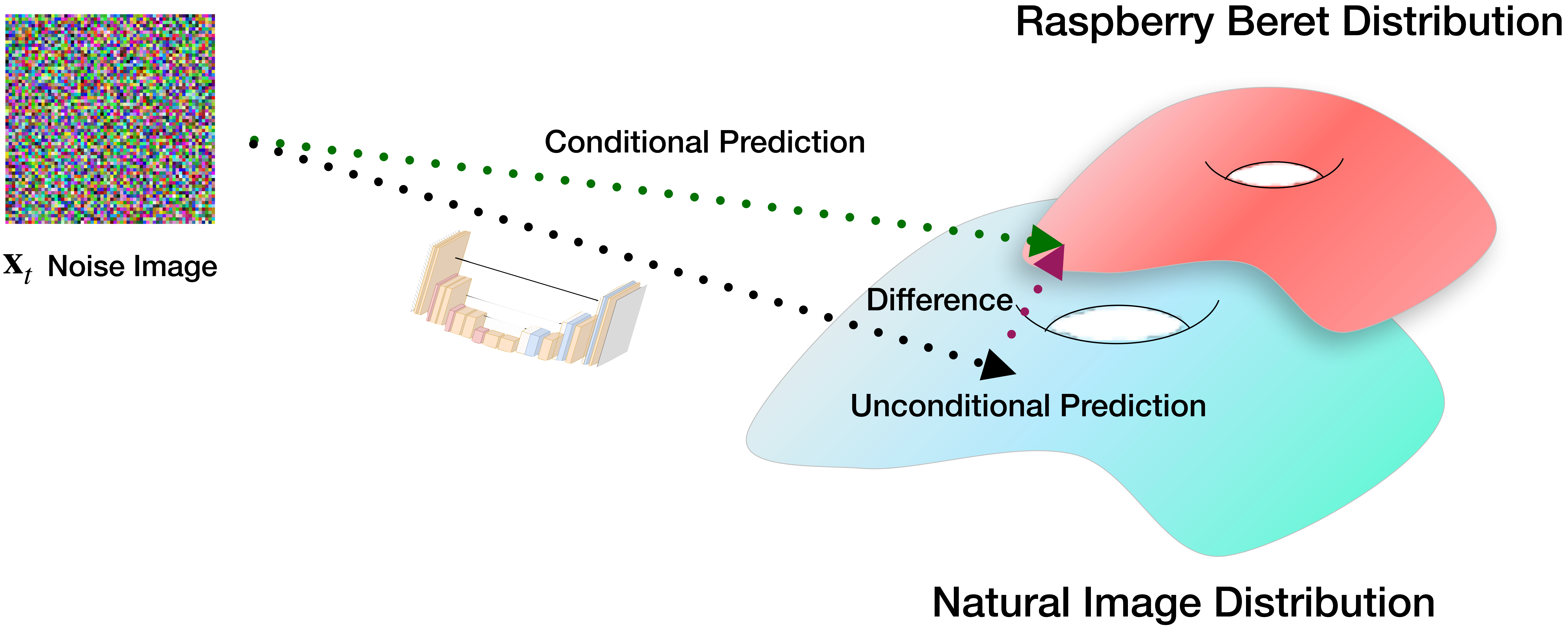



# Model is trained to output both

- $\epsilon_\phi(\mathbf{x}_t; c)$
- $\epsilon_\phi(\mathbf{x}_t; \varnothing)$

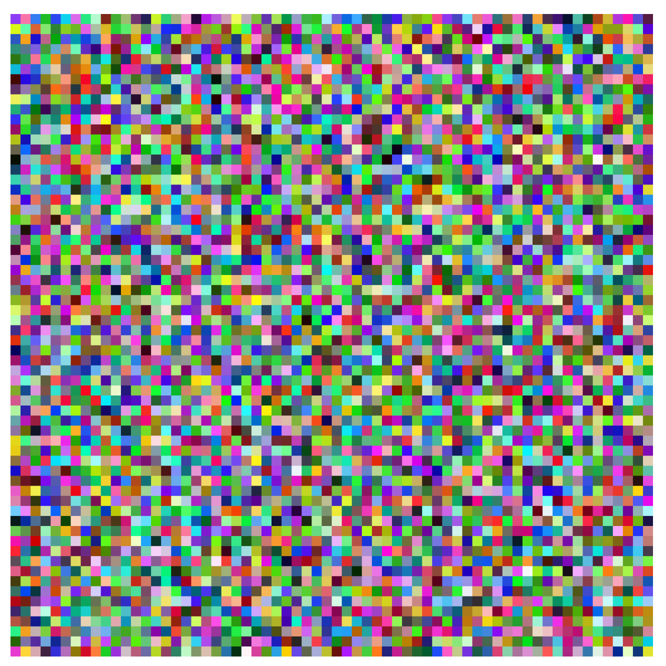
# Conditional Diffusion

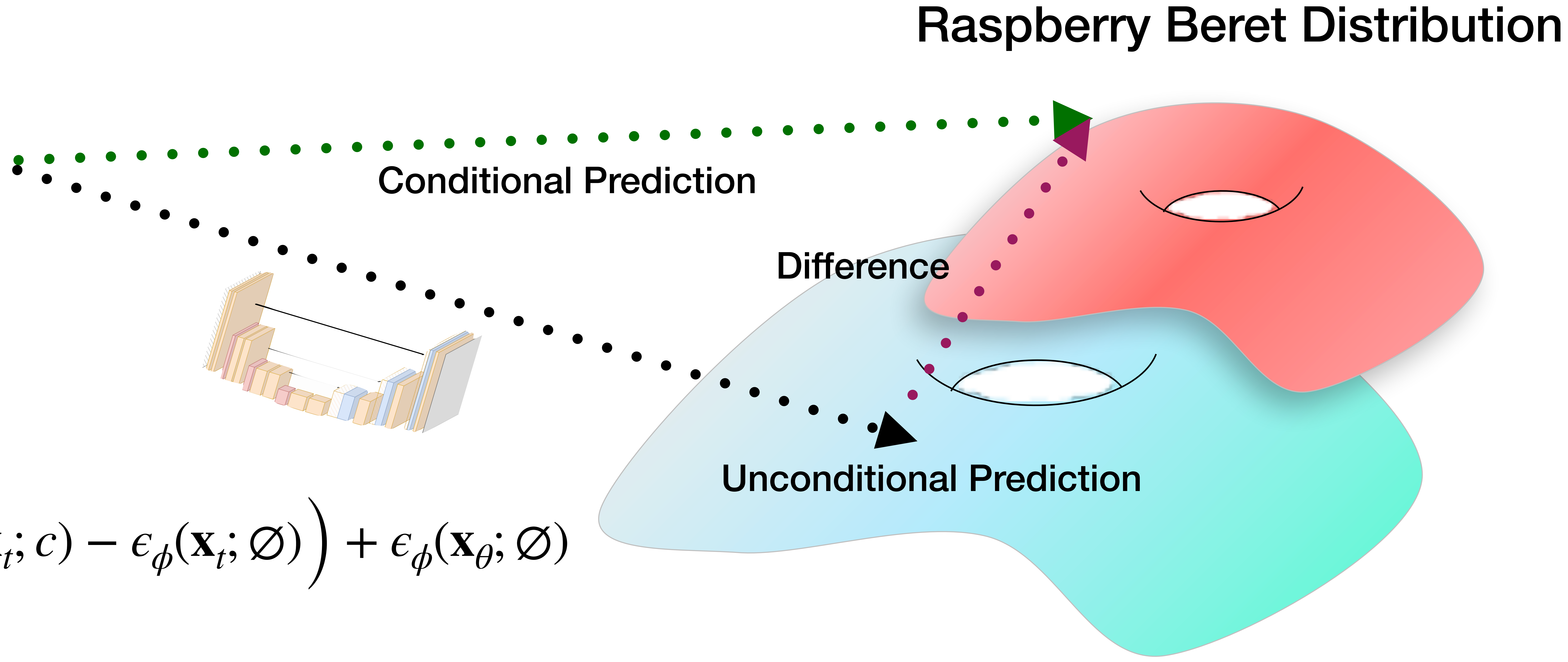
## Model Knows Unconditional, Text-Conditioned Distributions



# Classifier-Free Guidance

## Model Knows Unconditional, Text-Conditioned Distributions




$$\epsilon_{CFG} = s \cdot \Big( \epsilon_\phi(\mathbf{x}_t; c) - \epsilon_\phi(\mathbf{x}_t; \varnothing) \Big) + \epsilon_\phi(\mathbf{x}_\theta; \varnothing)$$

$\epsilon_\phi$ Predicted Noise

$\varnothing$ Null Prompt

$c$ Target Prompt

$s$ CFG Scale

# Classifier-Free Guidance

TLDR: Amplify Delta Between Conditional, Unconditional Predictions

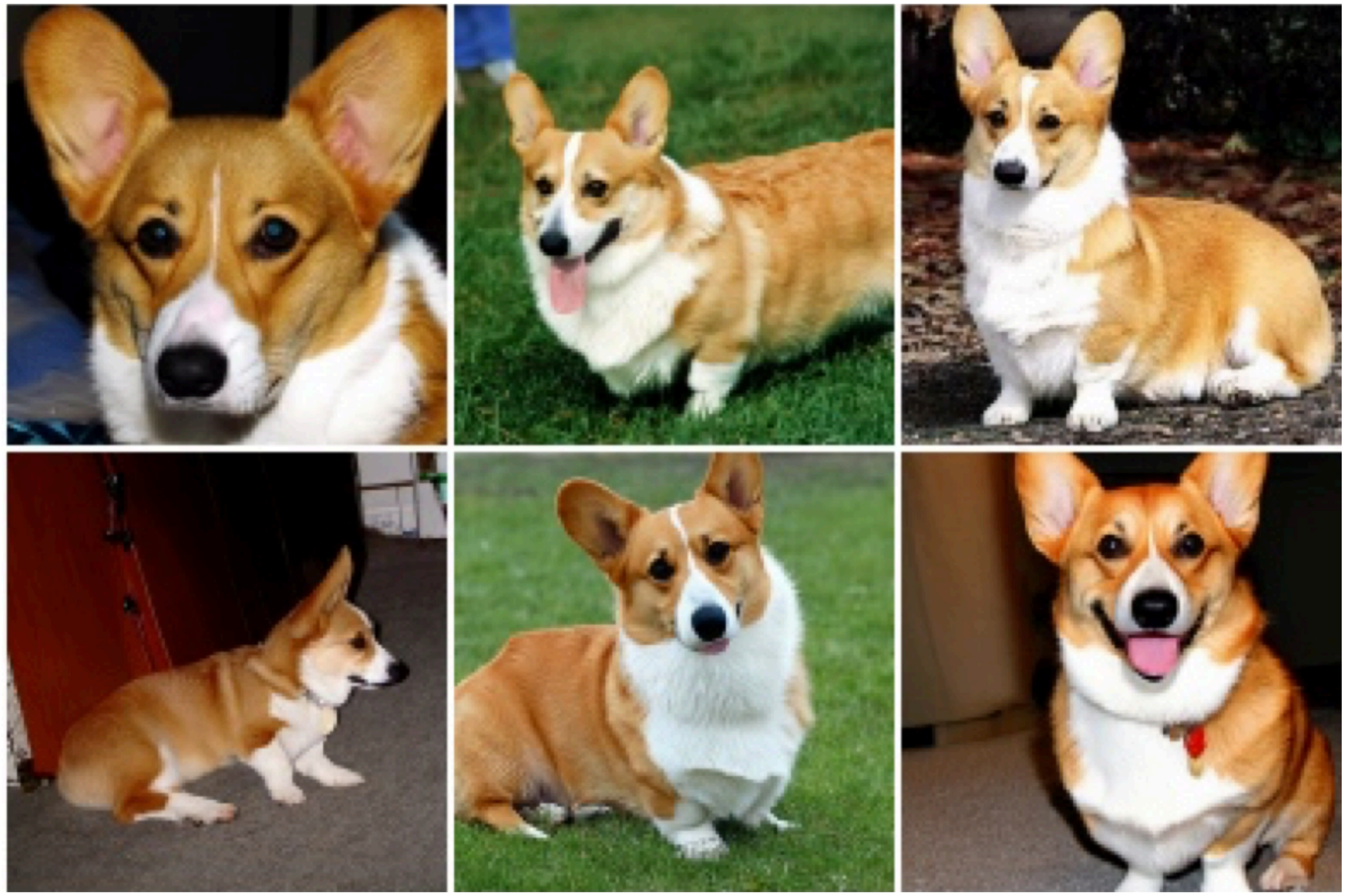
$$\epsilon_{\phi,CFG} = s \cdot \Big(\epsilon_\phi(\mathbf{x}_t; c) - \epsilon_\phi(\mathbf{x}_t; \varnothing)\Big) + \epsilon_\phi(\mathbf{x}_\theta; \varnothing)$$

| | |
|---|---|
| $\epsilon_\phi$ | Predicted Noise |
| $\varnothing$ | Null Prompt |
| $c$ | Target Prompt |
| $s$ | CFG Scale |

# Classifier-Free Guidance

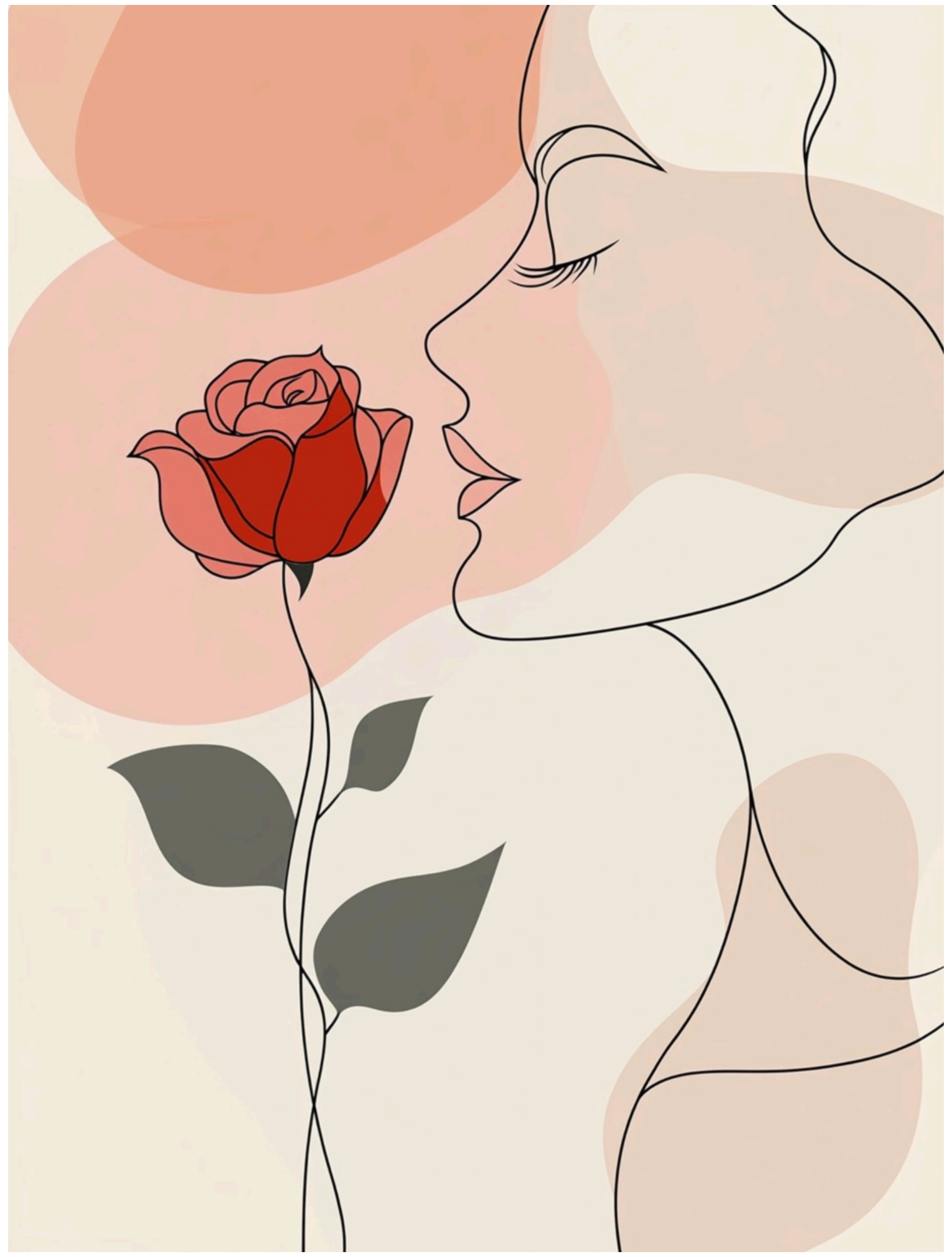
## Drives Generations Toward Conditional Mode

s = 1 (Conditional Generation)



s = 3 (Conditional Generation)

"Classifier-Free Diffusion Guidance" Ho et al. (2022)



# Text-Conditioned Diffusion Samples

**(Midjourney)**

Text-Conditioned Diffusion Sample

(Veo 2)

$$p(x \mid c)$$

*Guidance Lets Us Sample a Conditional Distribution*

$$p(x \mid c)$$

*Guidance Lets Us Sample a Conditional Distribution*

**What if we're creative about the conditioning?**


Condition on the First frame and Audio, Generate the Video

OmniHuman-1: Rethinking the Scaling-Up of One-Stage Conditioned Human Animation Models. Lin et al. (2025)


Condition on the First Frame and Actions, Generate an Interactive Environment

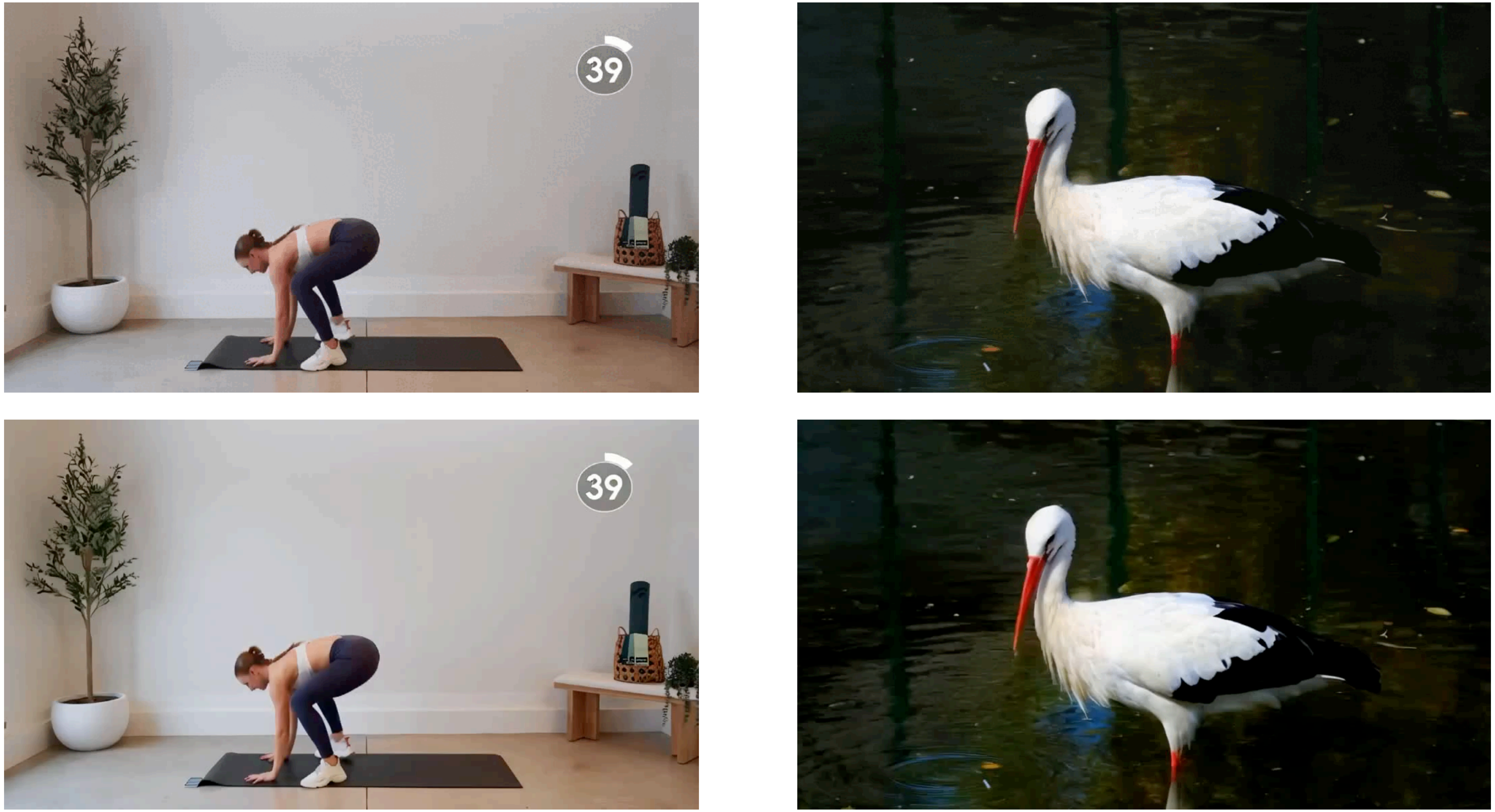
Genie: Generative Interactive Environments. Jake et al. (2024)

Condition on Video Pixels, Generate Depth Map

"Generating Consistent Long Depth Sequences for Open-world Videos" Hu et al. (2025)

Conditioning

Samples




# Condition on Start and End Frame, Generate Intermediate Frames

"Explorative Inbetweening of Time and Space" Feng et al. (2024)

Condition on Image + Camera, Generate Novel Views

"CAT3D: Create Anything in 3D with Multi-View Diffusion Models" Gao et al. (2024)


# Condition on Image + Camera, Generate Novel Views

Stable Virtual Camera: Generative View Synthesis with Diffusion Models (Zhou*, Gao* et al.)


Condition on Ego View, Generate Human Poses

"EgoAllo" Yi et al. (2025)