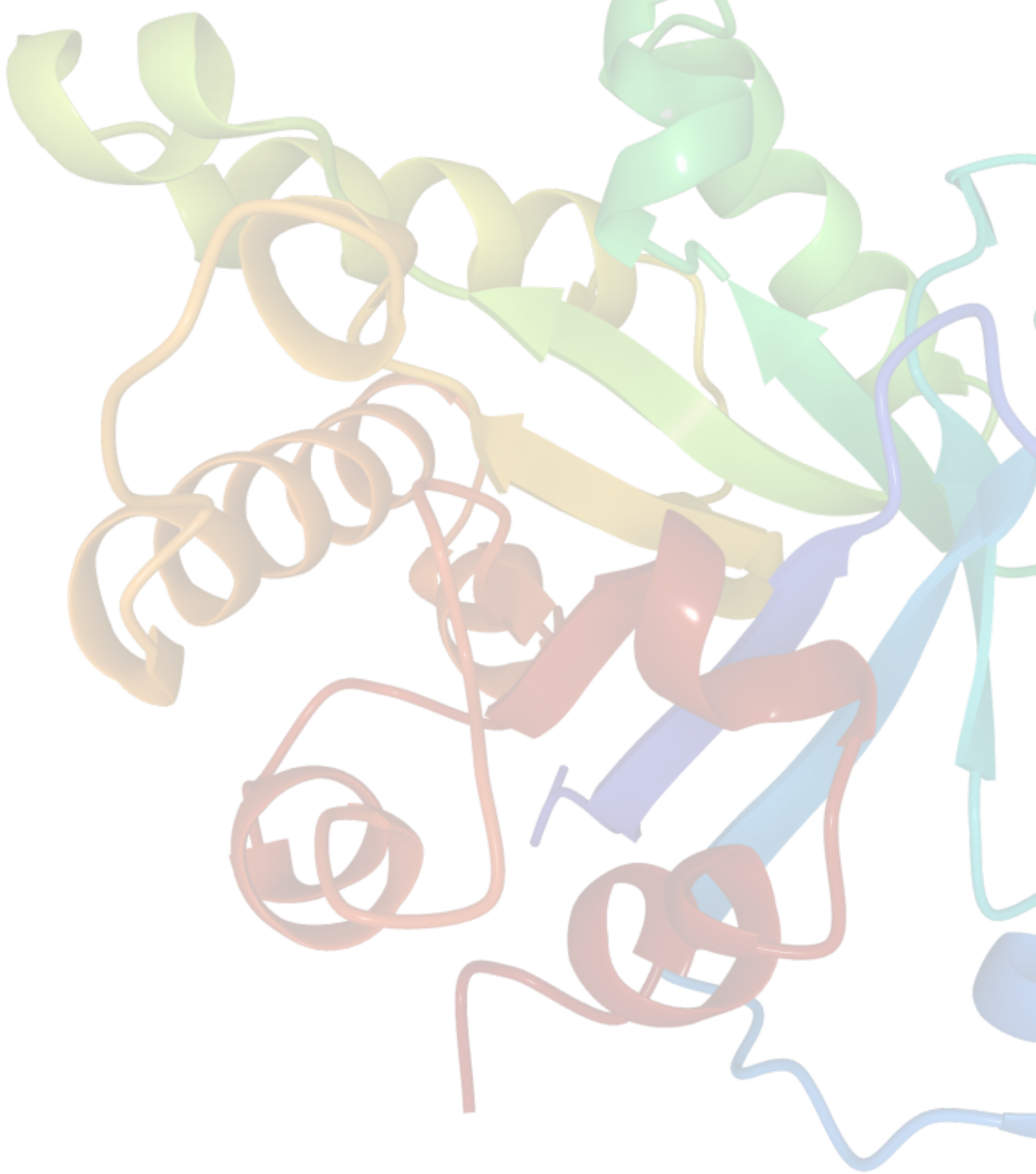
Flow Matching

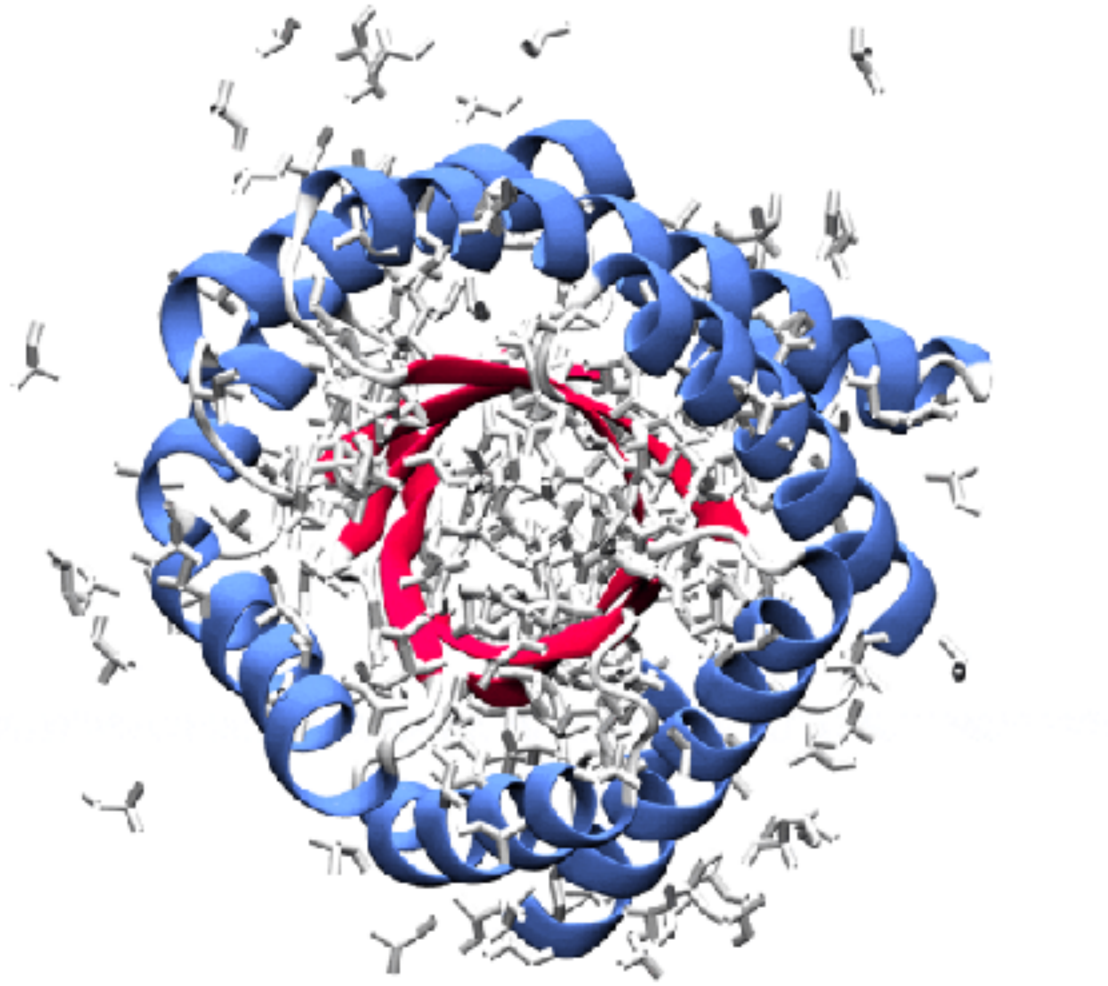
Alex Tong

Probabilistic AI Summer School 2025



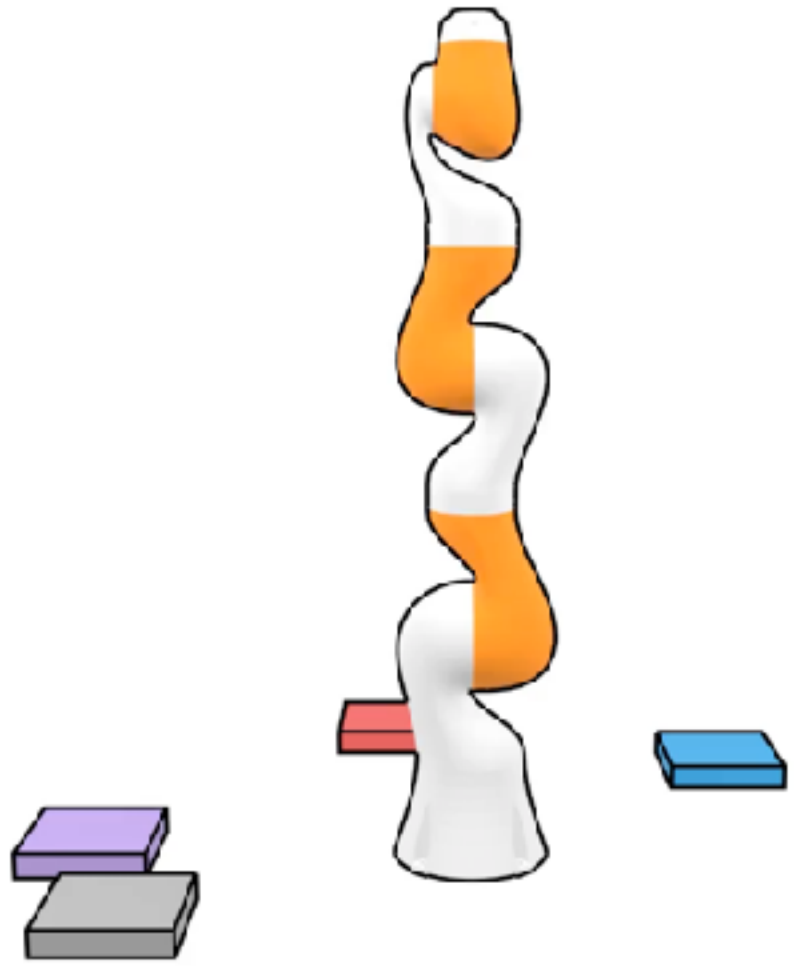
Generative Models Beyond Images and Text

Scientific Data



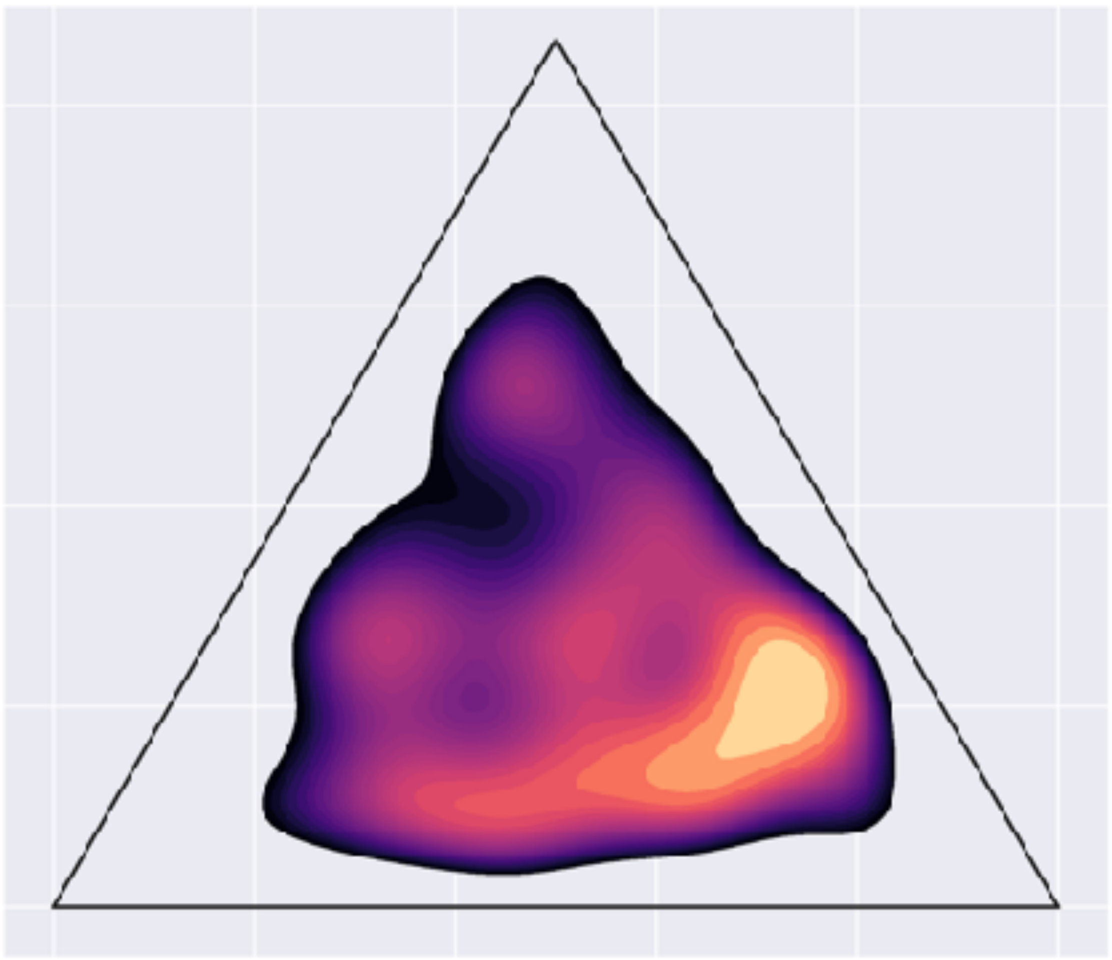
SE(3) invariant
Protein structure generation

Robotics



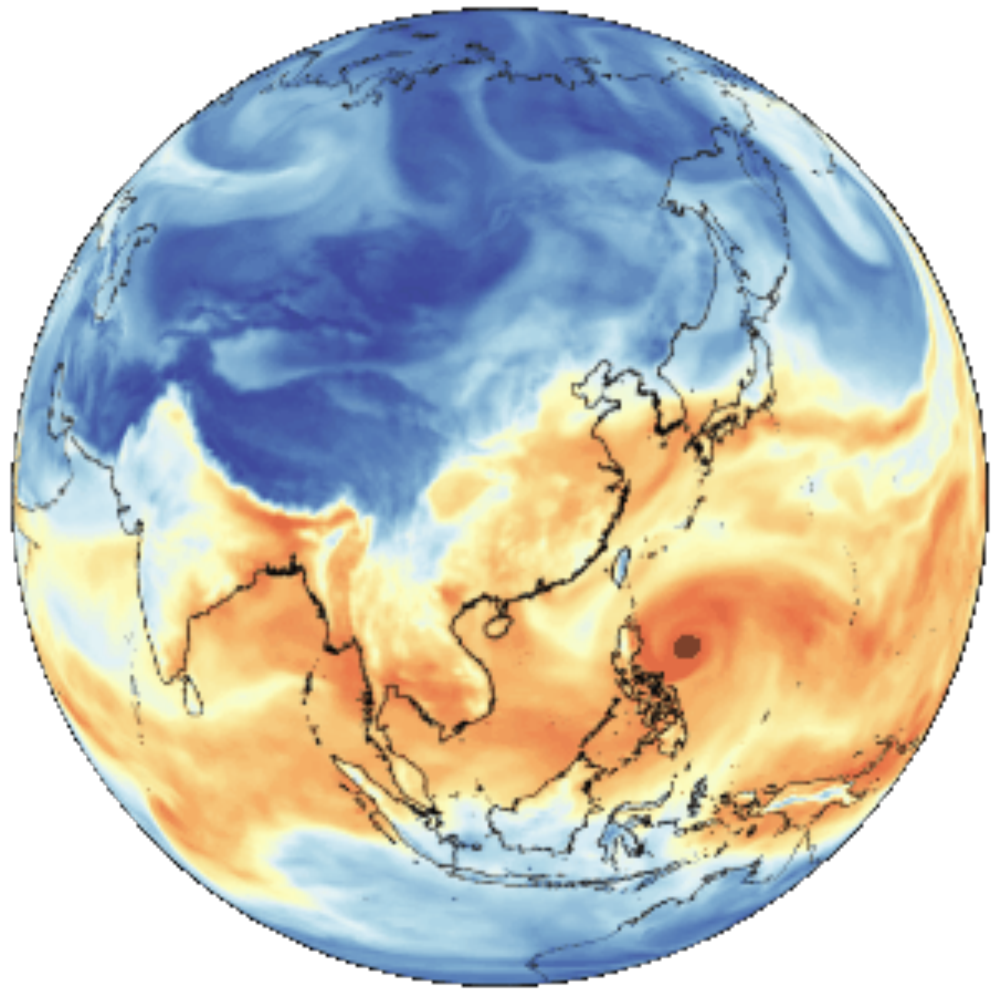
SO(2) invariant
Block stacking

Information Geometry



Fisher-Rao geometry
On the probability Simplex

Climate Modeling



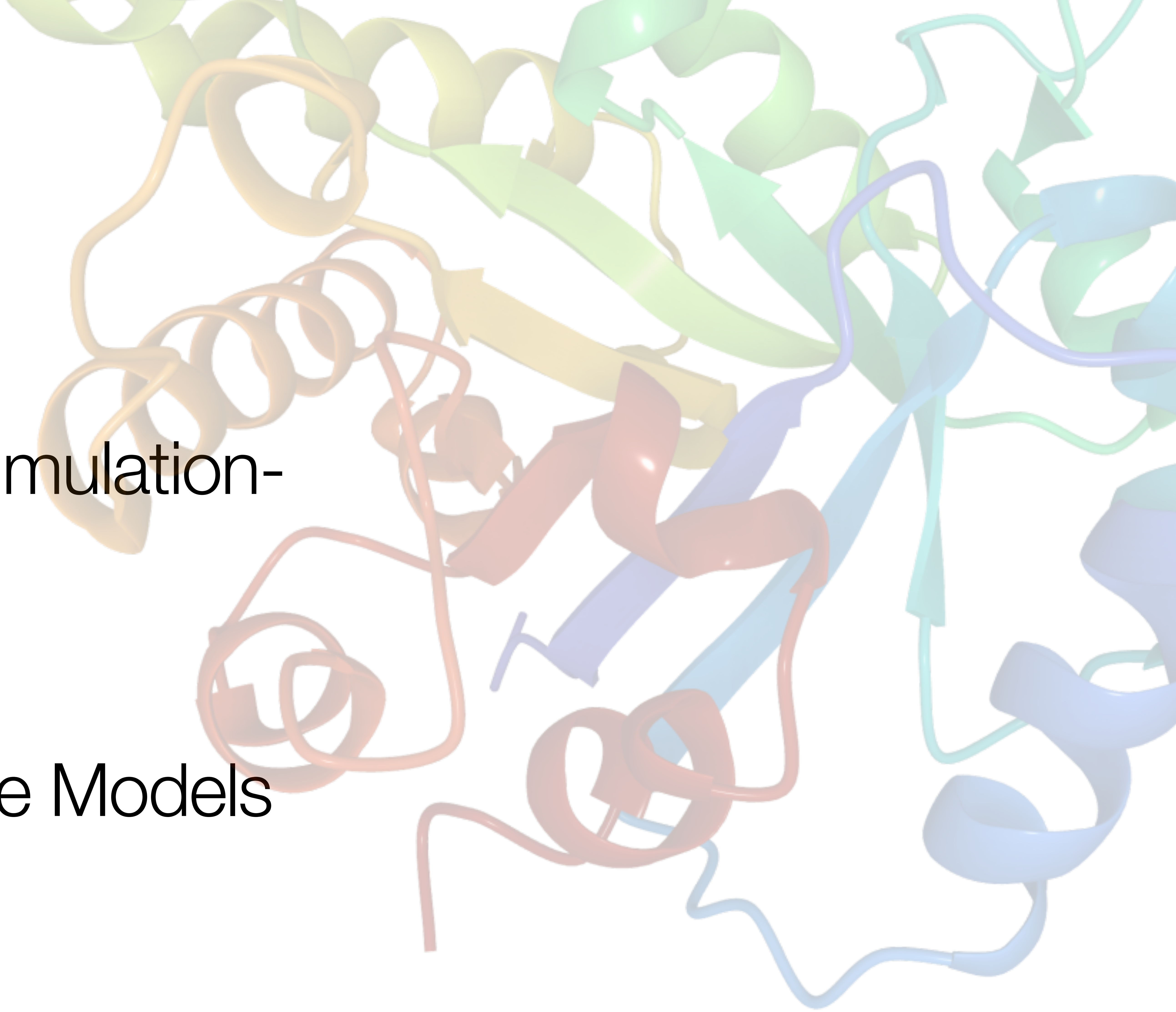
Spherical Geometry $\mathbb{S}^2$

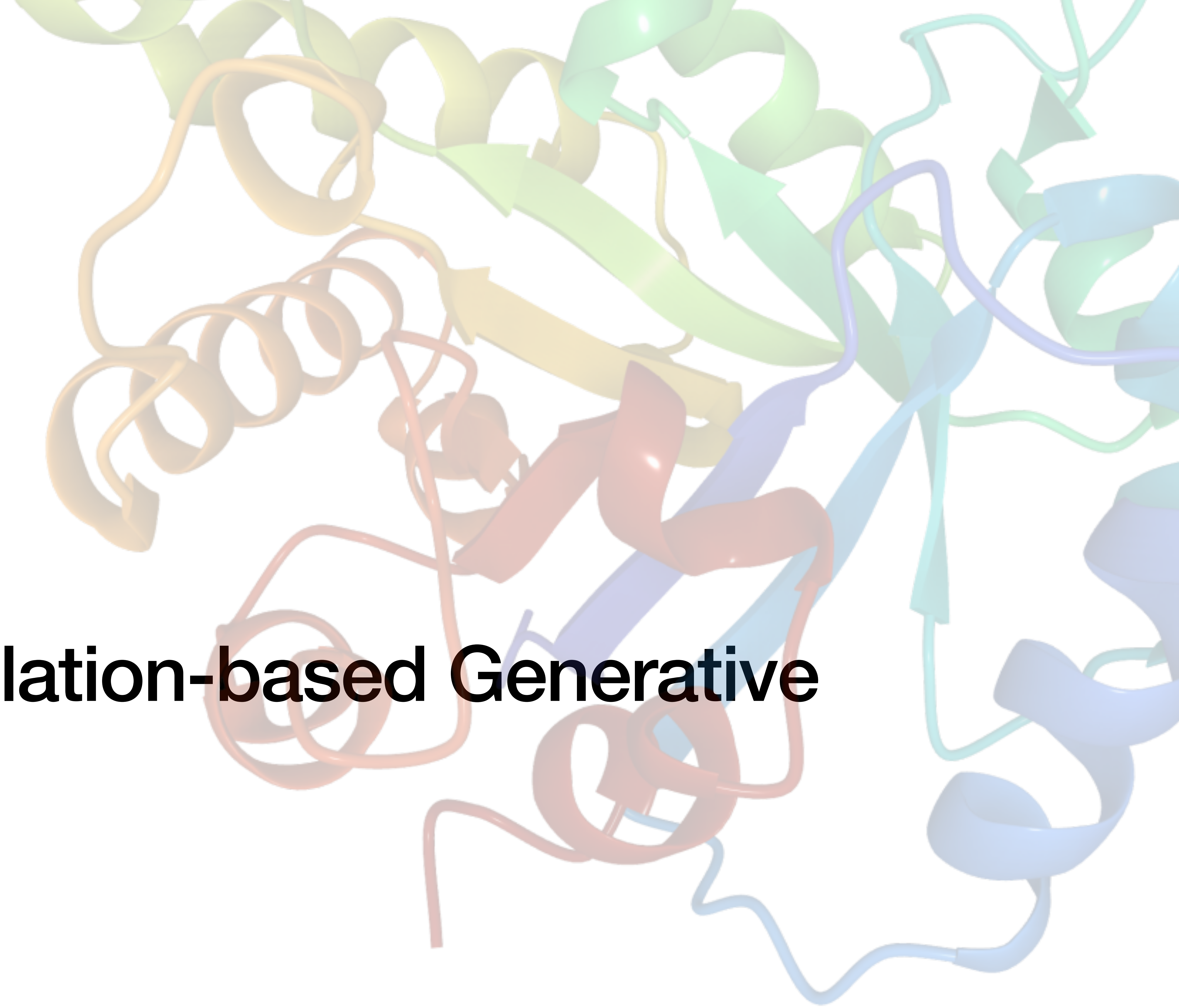
Outline:

Part I: Dynamical Systems and Simulation-based Generative Models

Part II: Simulation-Free Generative Models

Part III: Applications and Tricks



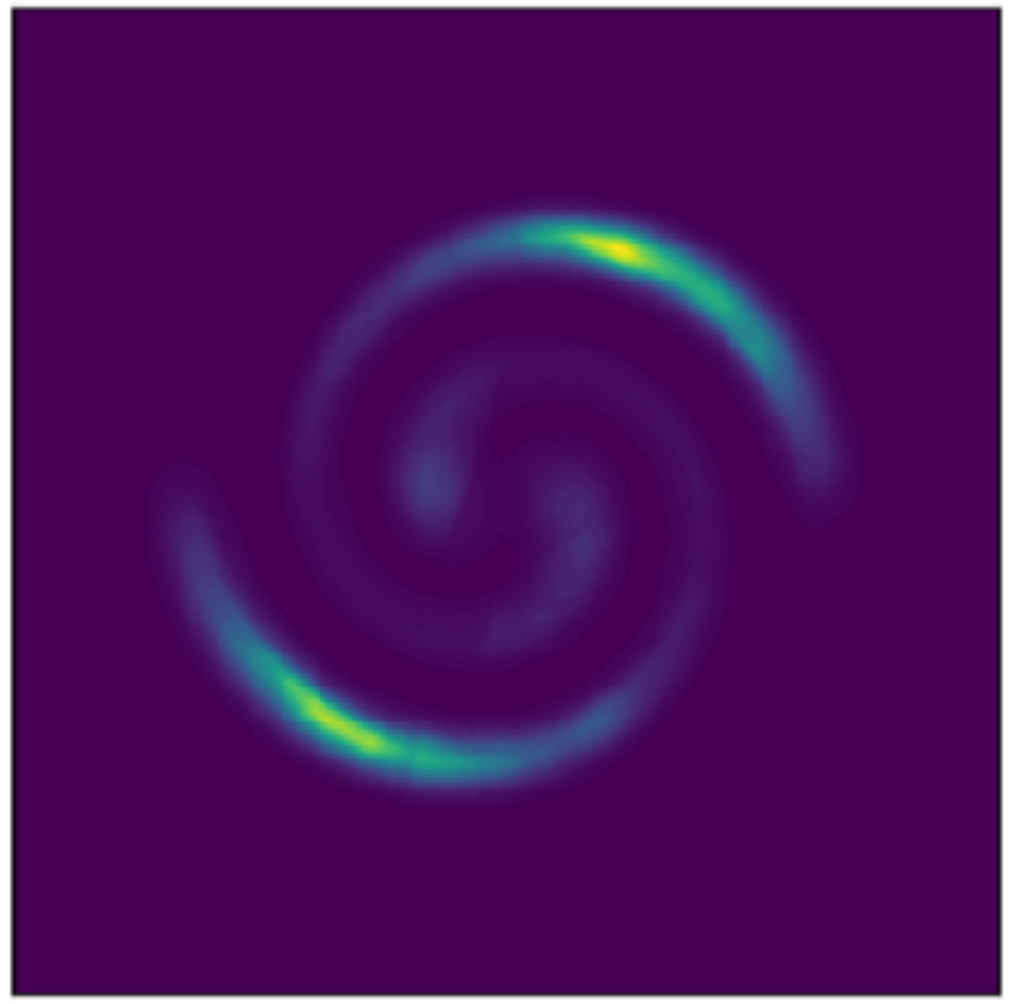


Part I:

Dynamical Systems and Simulation-based Generative Models

Problem Setting: Generative Modeling

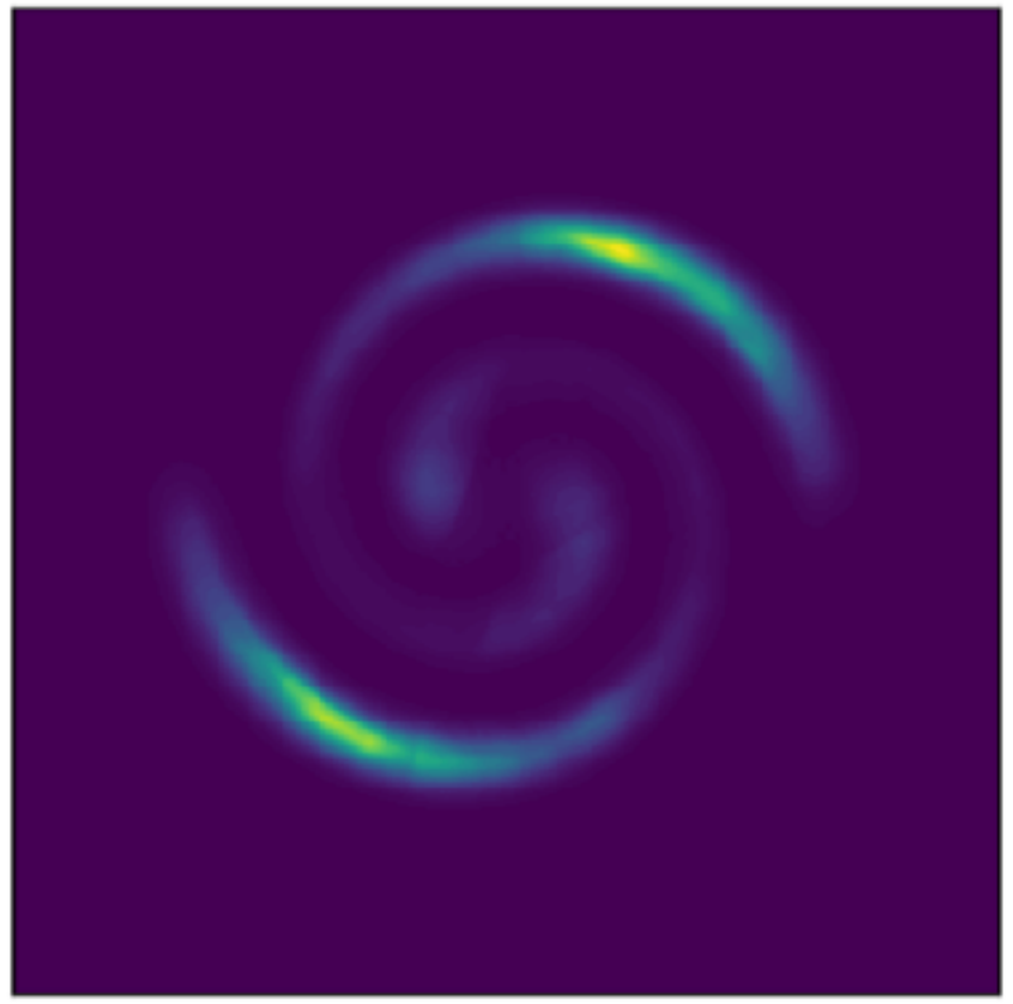
- Unknown: data distribution q



Problem Setting: Generative Modeling

- Unknown: data distribution q

- Given: samples $x_1 \sim q$

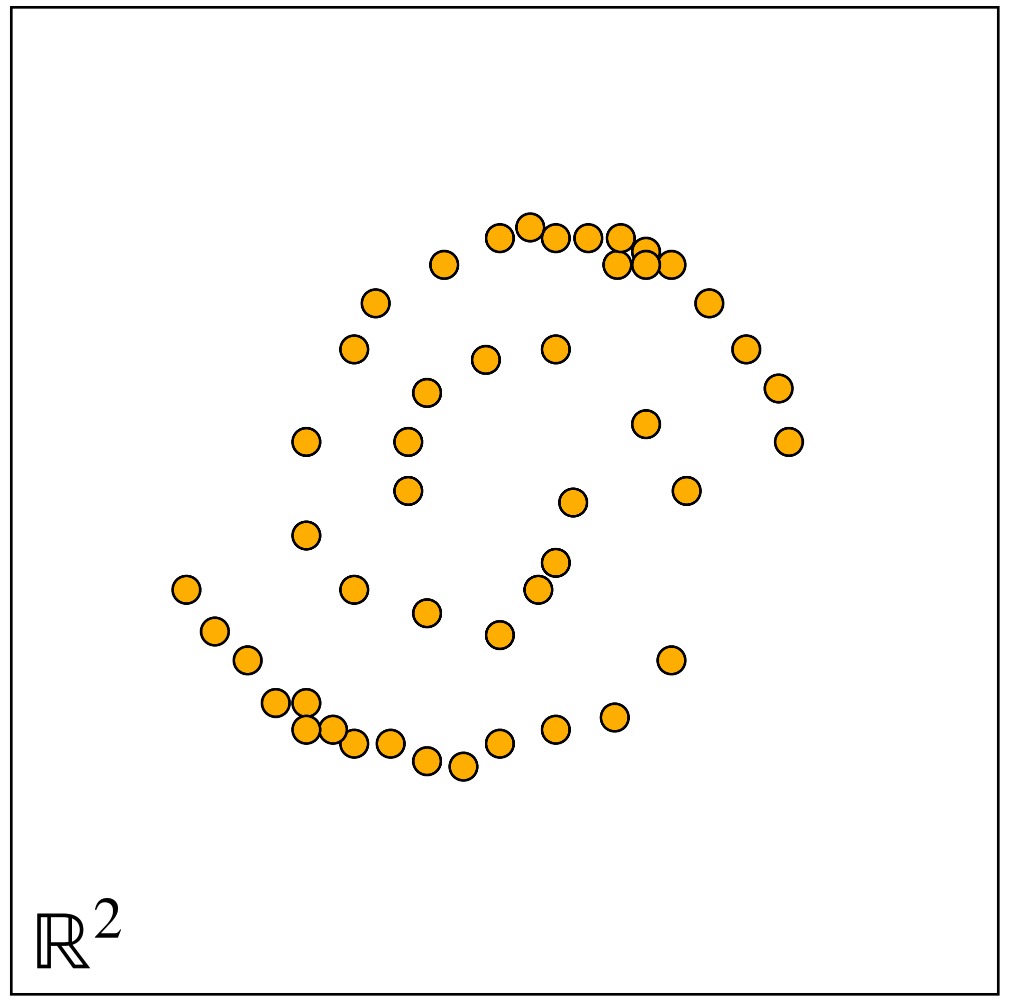


Problem Setting: Generative Modeling

- Unknown: data distribution q

- Given: samples $x_1 \sim q$

$x_1 \sim q$

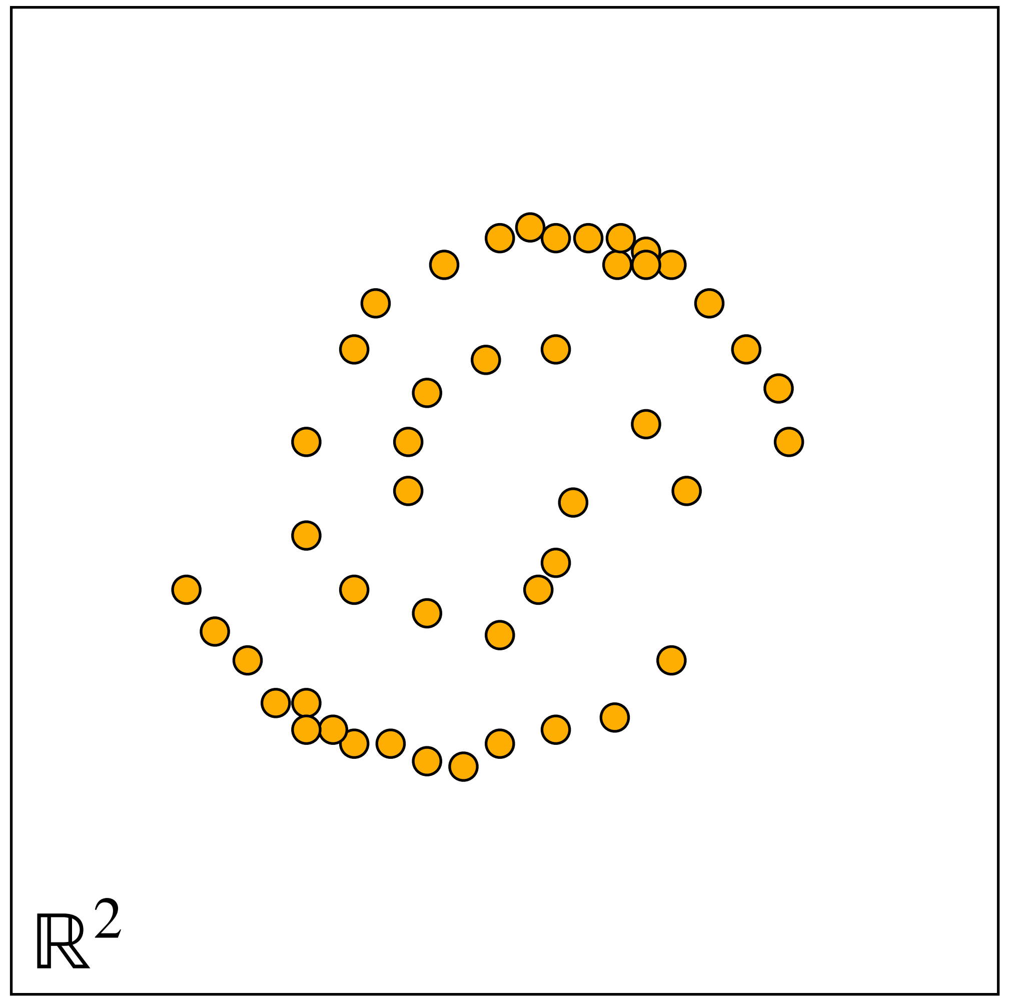


Problem Setting: Generative Modeling

- Unknown: data distribution q

- Given: samples $x_1 \sim q$

$x_1 \sim q$



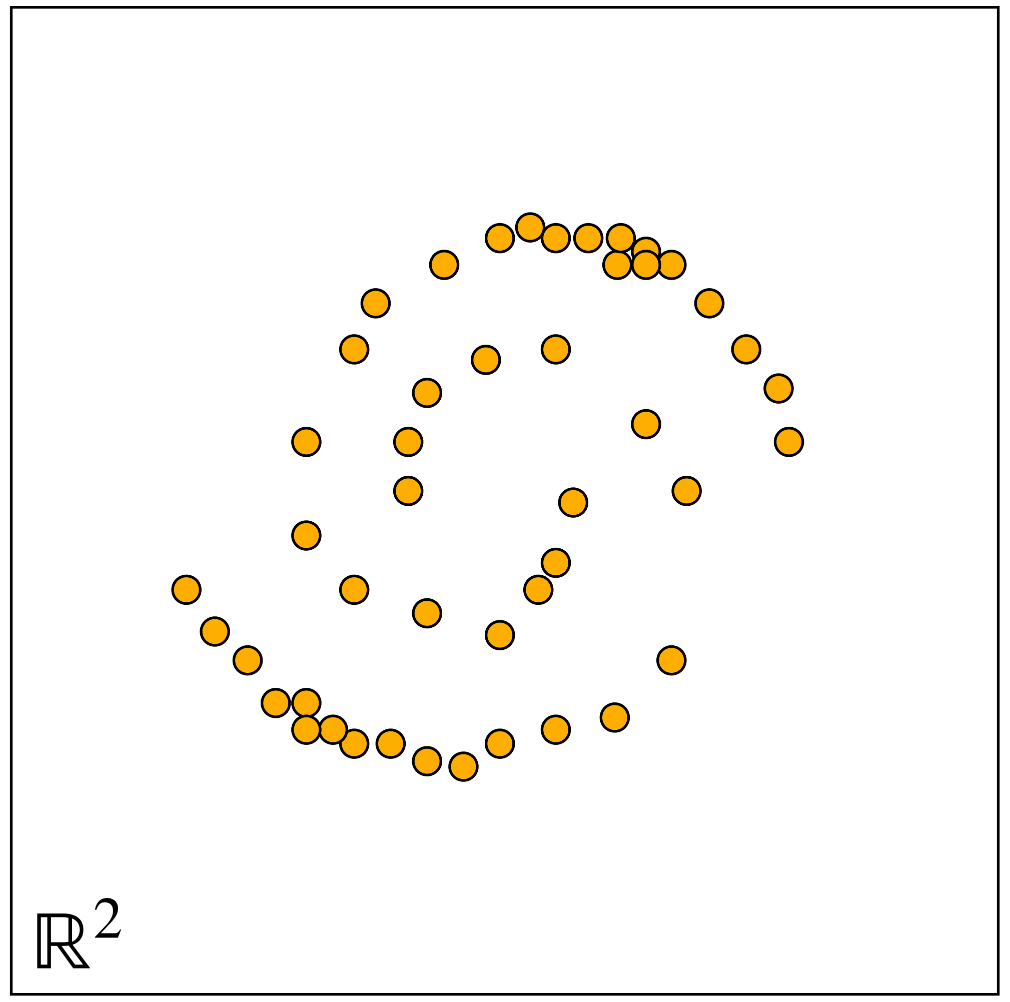
Goal: learn *a sampler* from the unknown q

Deep Generative Modeling

- Unknown: data distribution q

- Given: samples $x_1 \sim q$

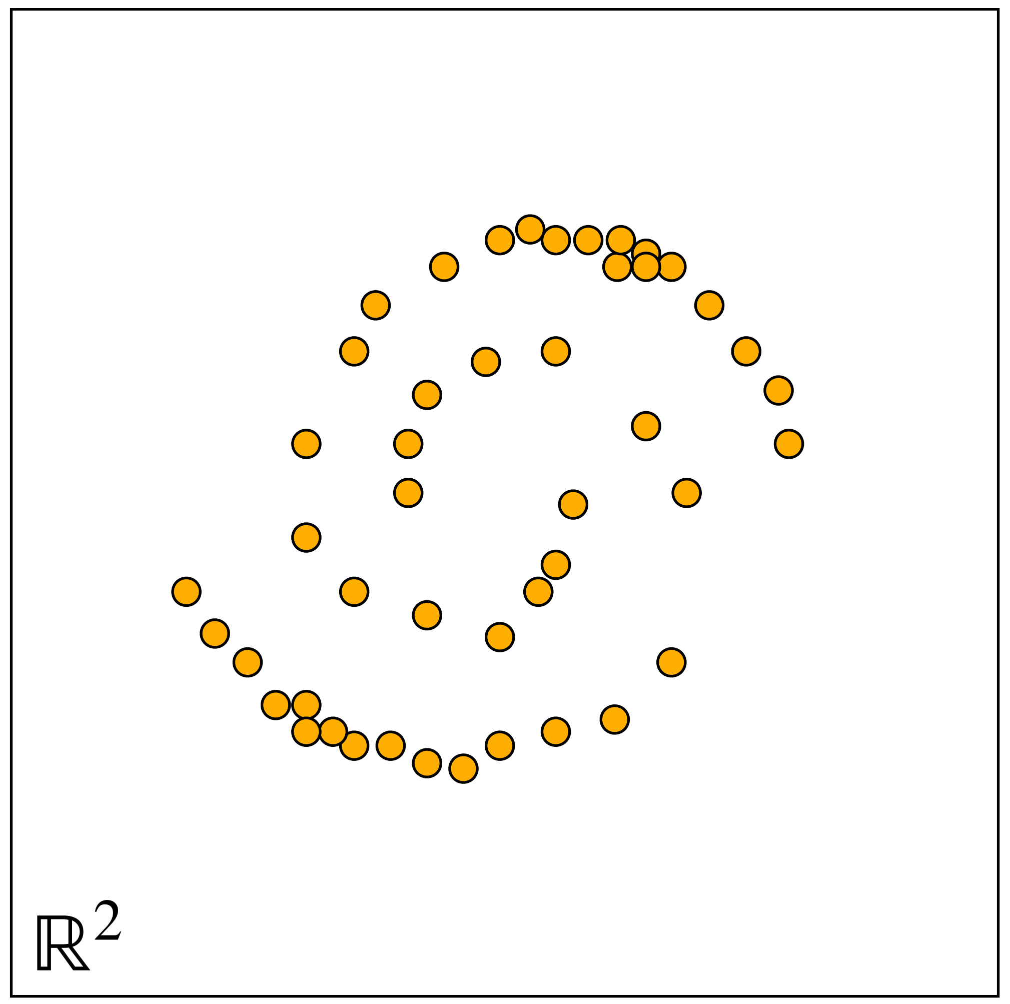
$x_1 \sim q$



Deep Generative Modeling

- Unknown: data distribution q
- Given: samples $x_1 \sim q$
- Learn: neural network with parameters θ

$x_1 \sim q$



Generative Model

$$(\psi_\theta , p_\theta)$$

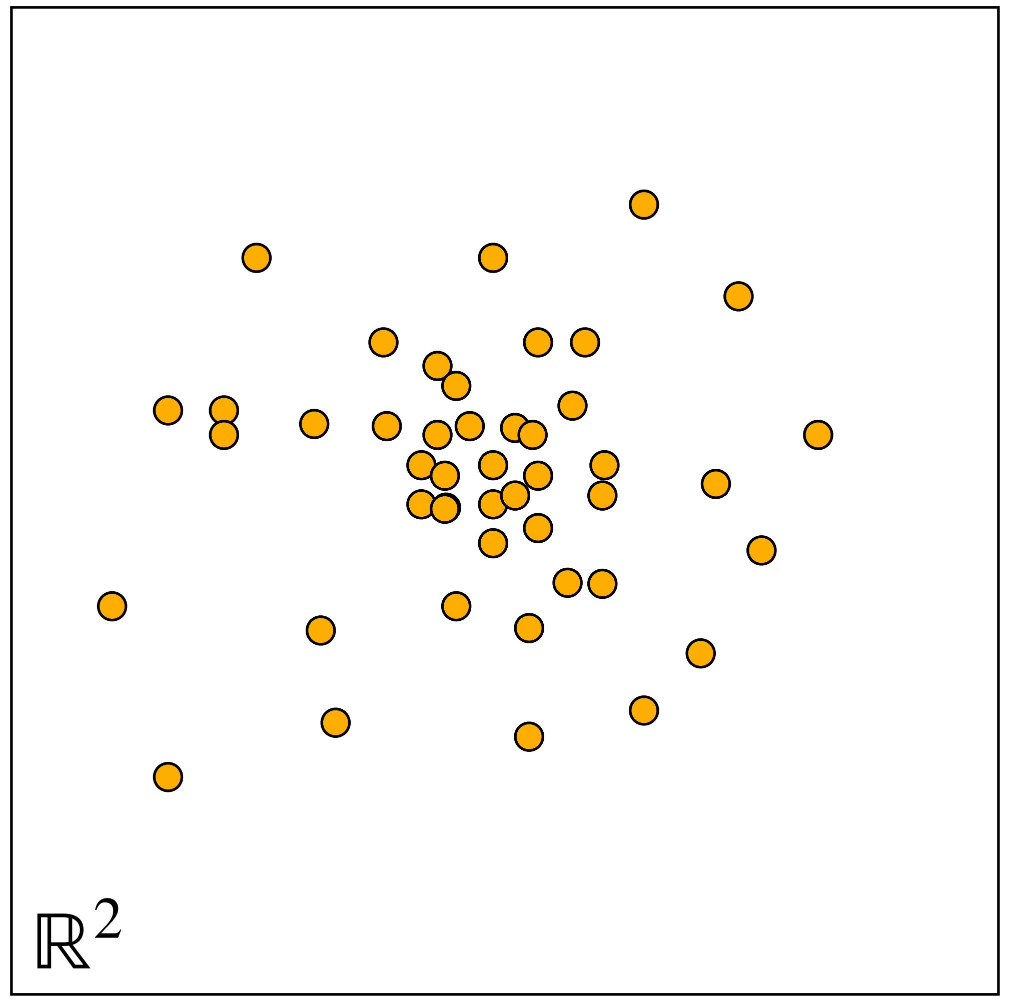
Generator   Underlying Density

Goal: find parameters θ s.t. $p_\theta \approx q$

Deep Generative Modeling

Easy to sample from

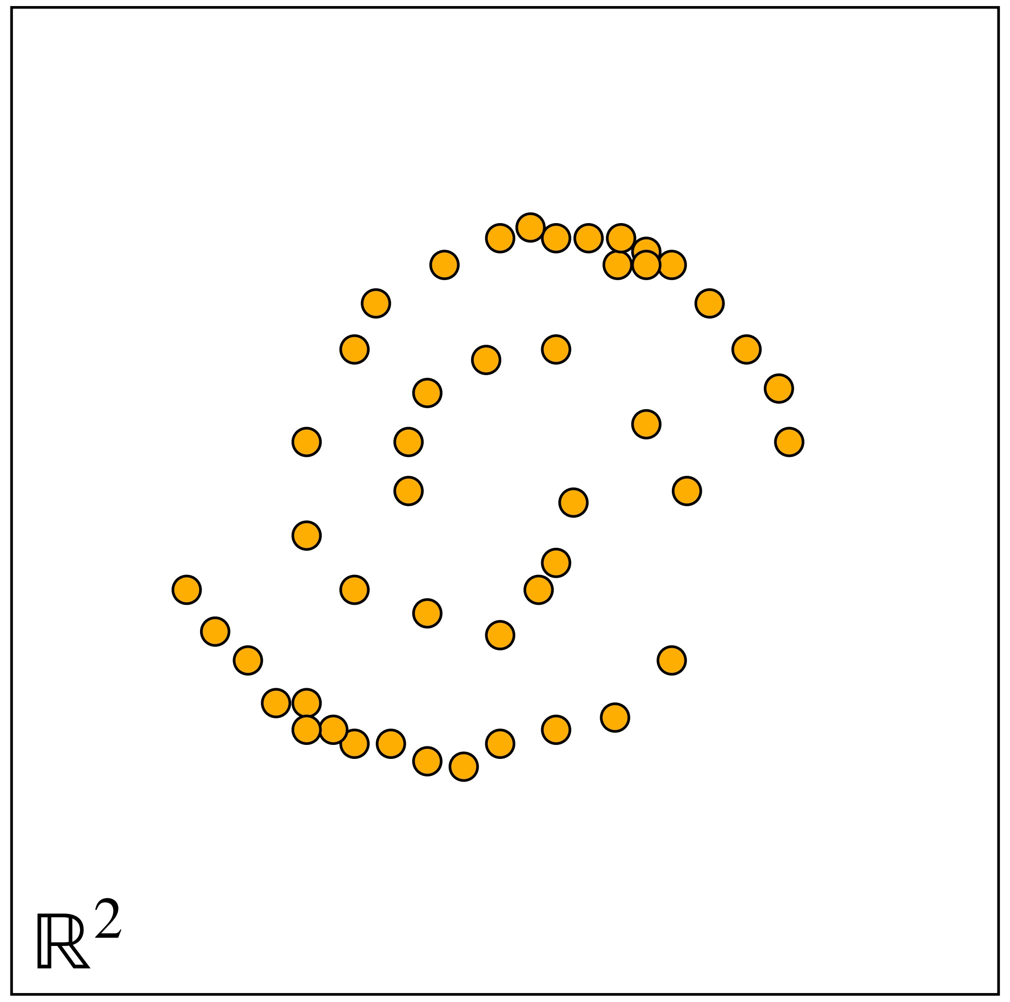
$x_0 \sim p$



Generator

ψ_θ

$x_1 \sim q$

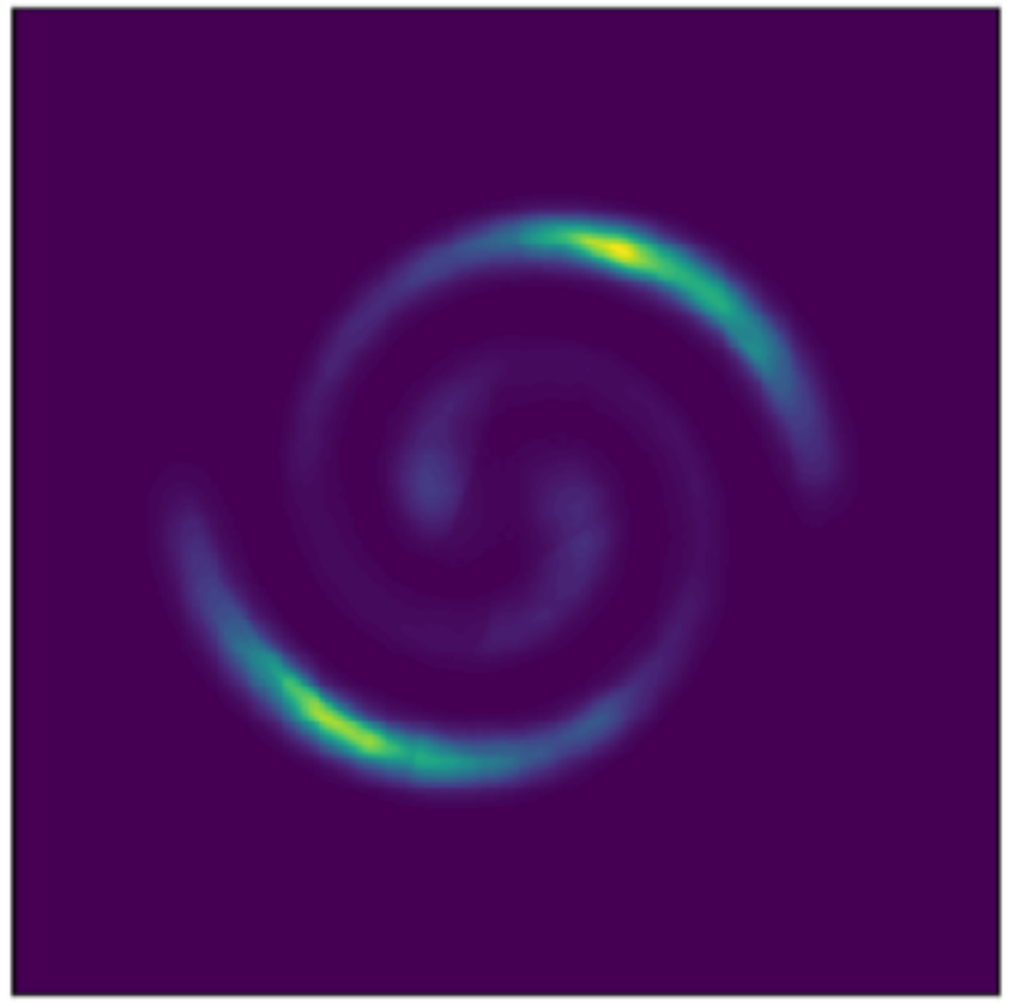


Sampling

$x_0 \sim p$

$\psi_\theta(x_0) \sim q$

How to model $\boldsymbol{\psi_\theta}$?

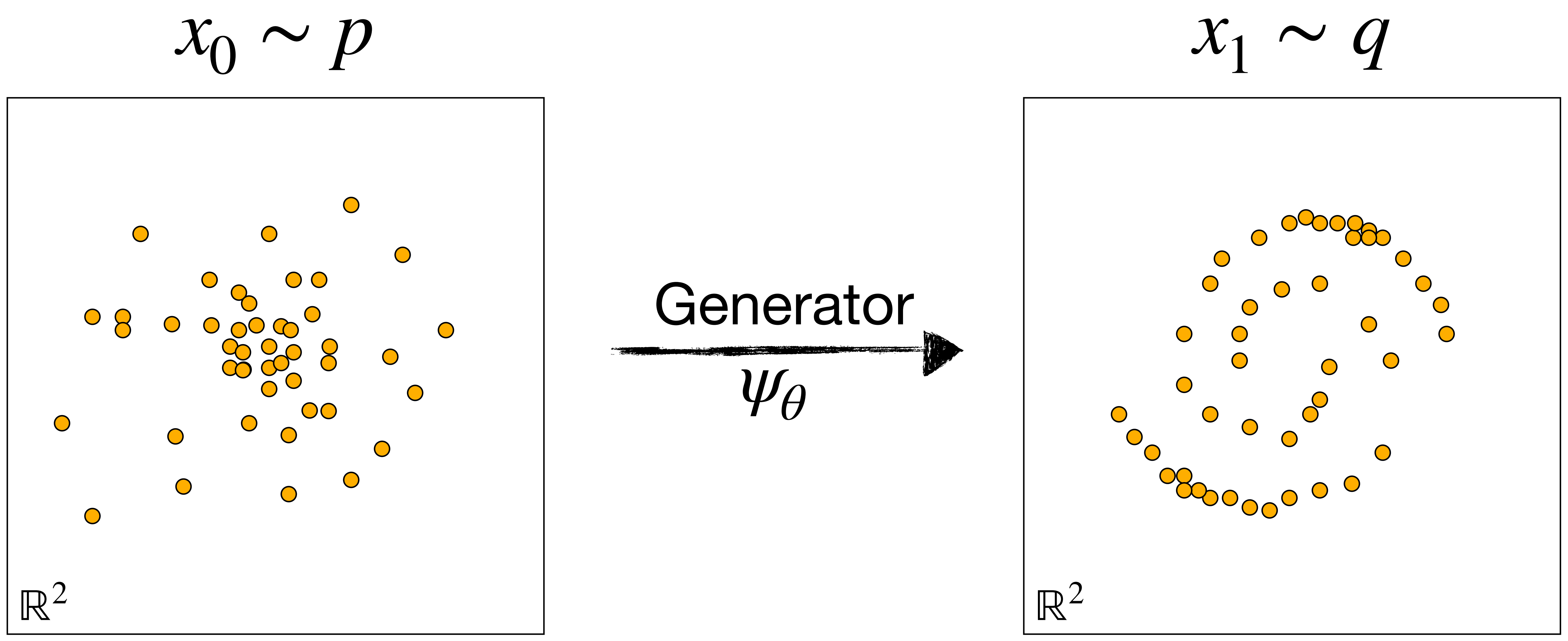


Density Estimation

$p_\theta \approx q$

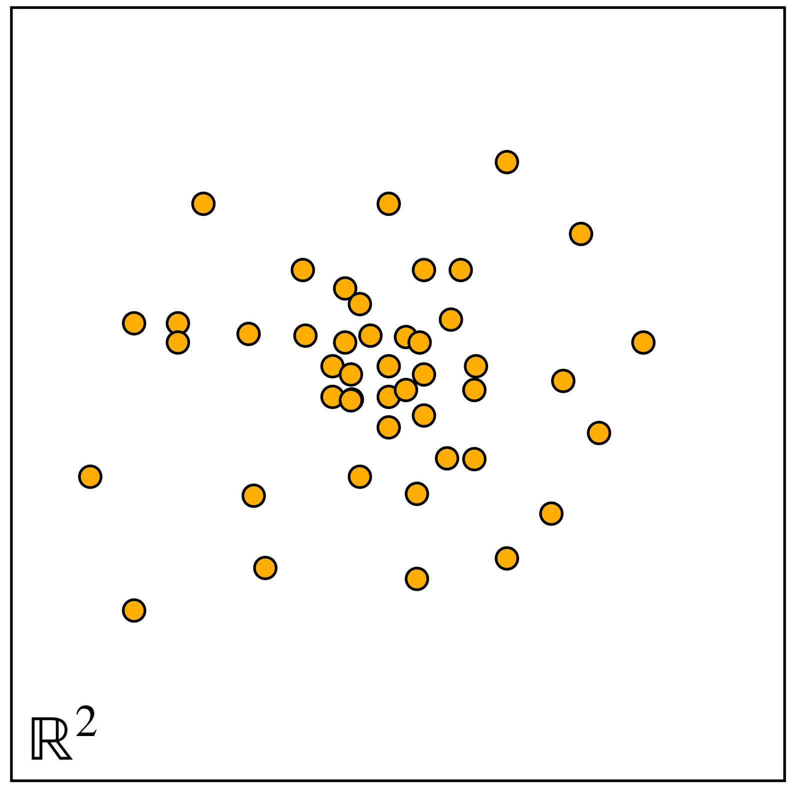
GANs, VAEs, Autoregressive

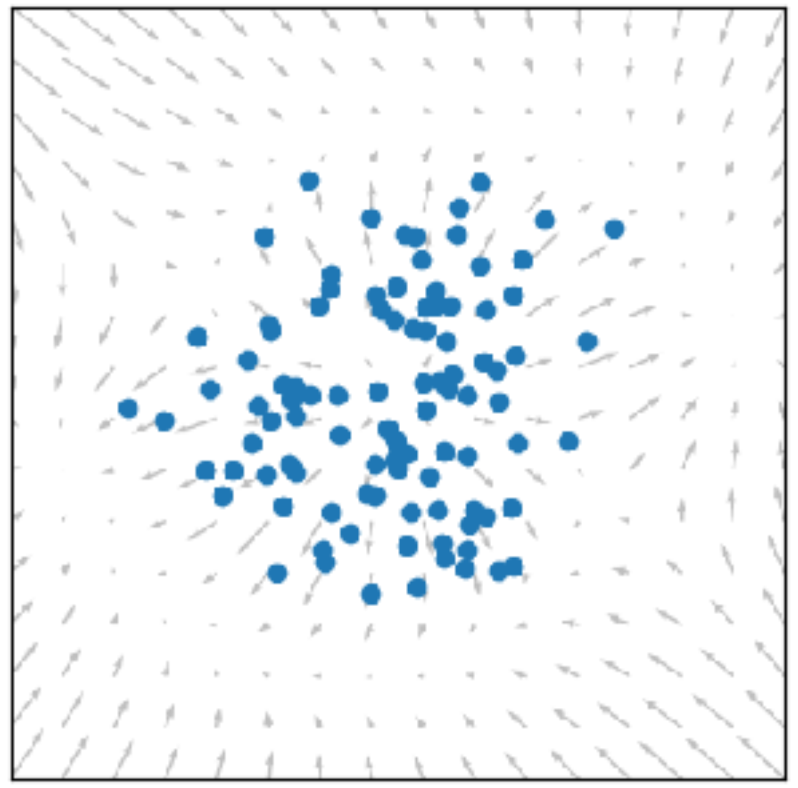
Focus: Generative Models as Dynamical Systems



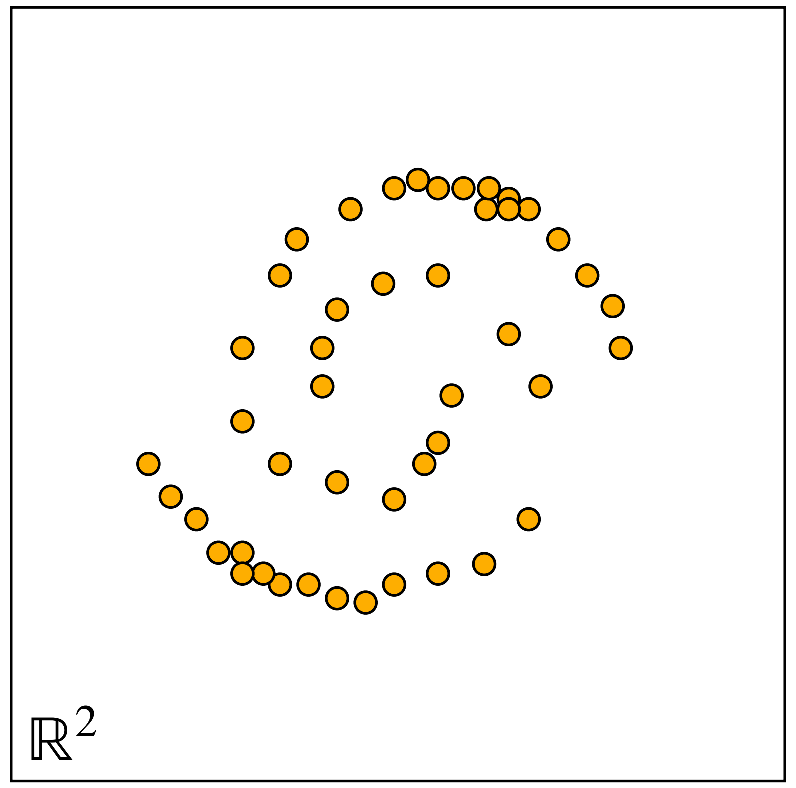
Focus: Dynamical Systems as Generative Models

$x_0 \sim p$





$x_1 \sim q$



$$\psi_t : [0,1] \times \mathbb{R}^2 \to \mathbb{R}^2$$

Time-Dependent Generator

Sampling = Simulating

$x_0 \sim p$ $\rightarrow$ $\text{simulate}(x_0, t) = \psi_t(x_0)$

Deterministic and Stochastic Dynamics

Flows

ODE $$dx_t = u_t(x_t)dt$$

Velocity field

Diffusion

SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$

Drift

Diffusion Coefficient

Brownian Motion

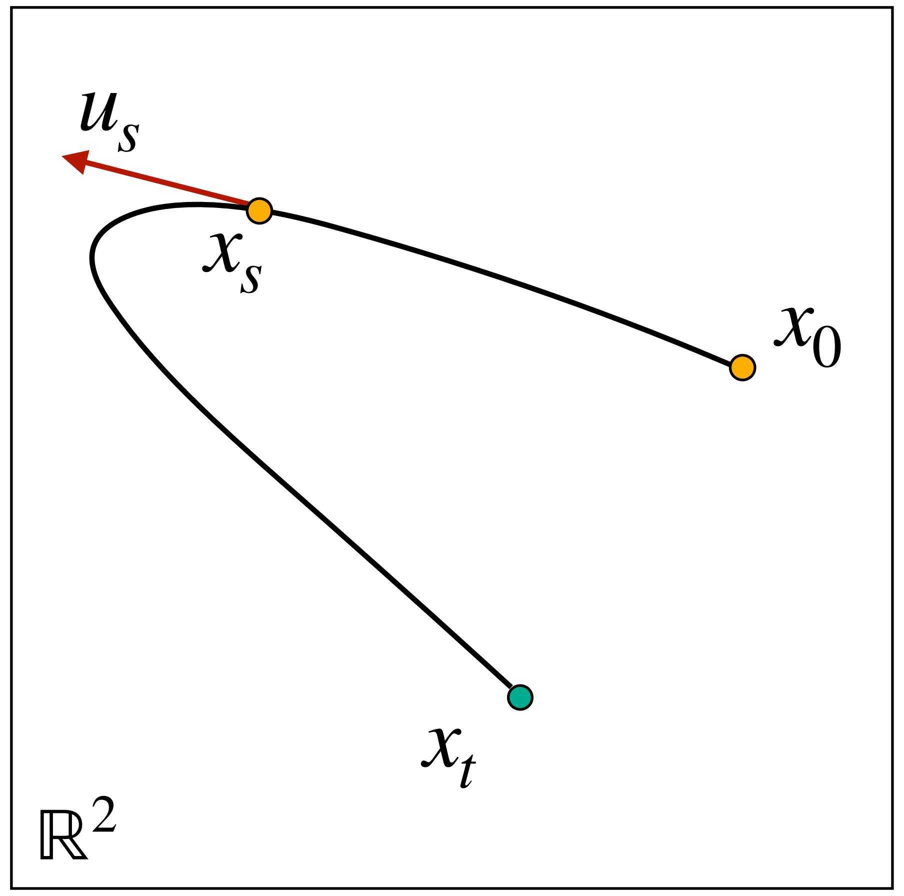
Deterministic and Stochastic Dynamics

Flows

ODE $$dx_t = u_t(x_t)dt$$

Velocity field

Deterministic



$$x_t = x_0 + \int_0^t u_s(x_s)ds$$

Diffusion

SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$

Drift

Diffusion Coefficient

Brownian Motion

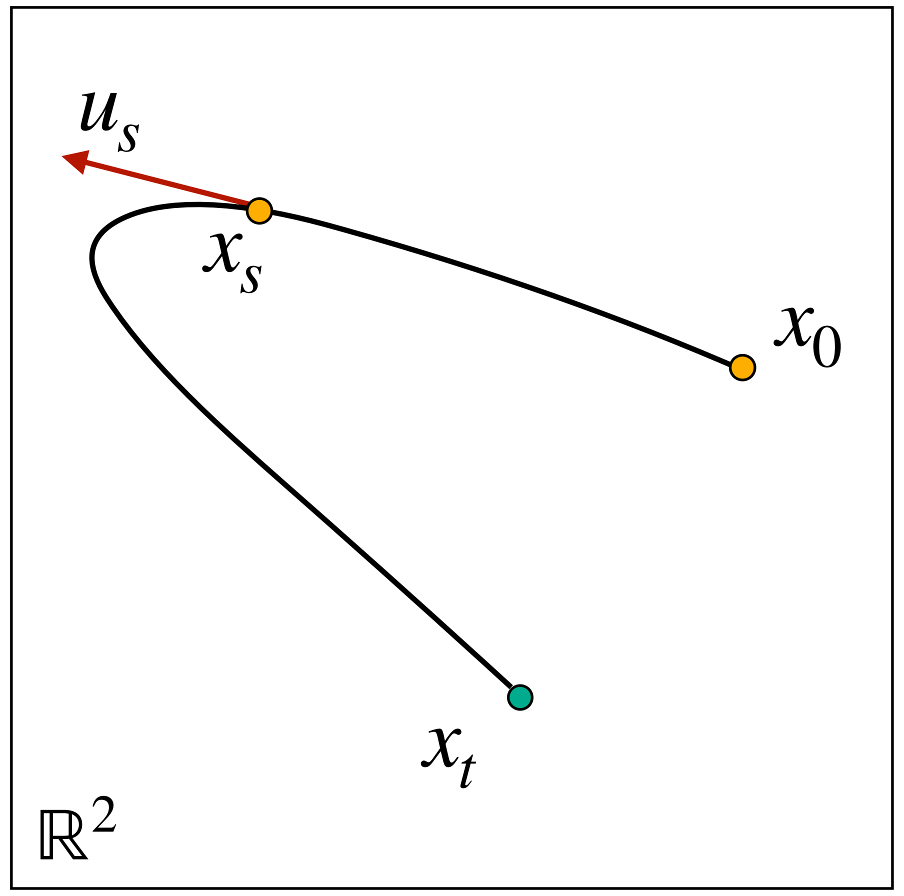
Deterministic and Stochastic Dynamics

Flows

ODE $$dx_t = u_t(x_t)dt$$

Velocity field

Deterministic



$$x_t = x_0 + \int_0^t u_s(x_s)ds$$

Diffusion

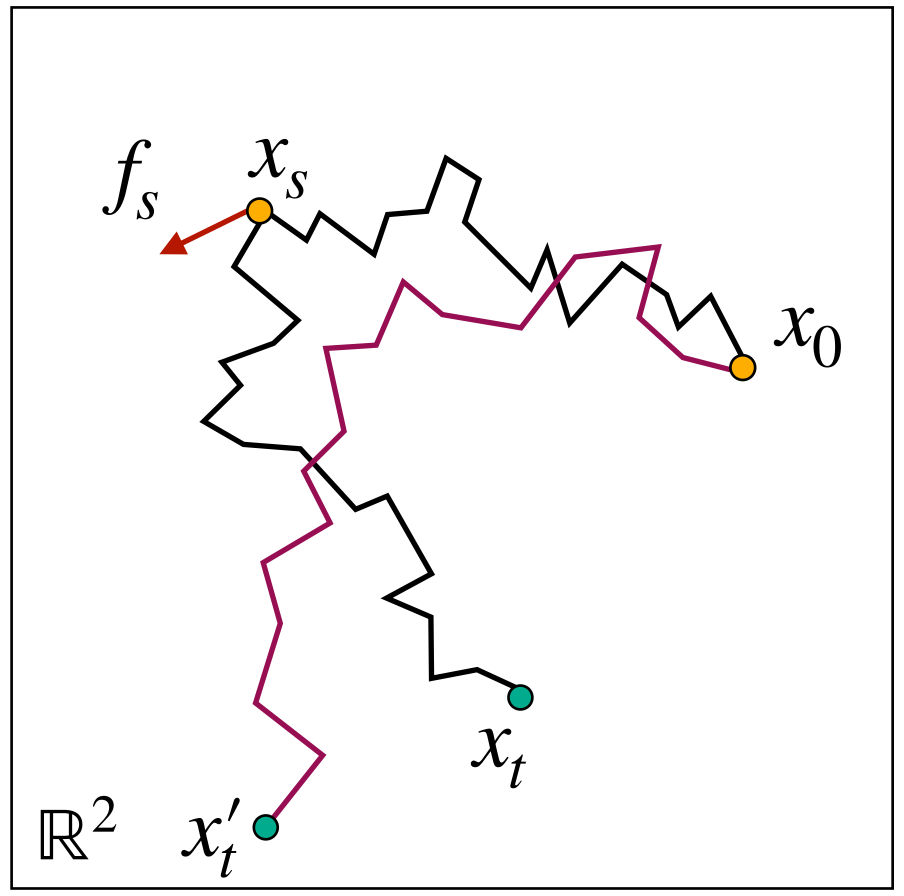
SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$

Drift

Diffusion Coefficient

Brownian Motion

Stochastic



$$x_t = x_0 + \int_0^t f_s(x_s)ds + \int_0^t g_s dw_s$$

Where are the probabilities?

Flows

ODE $$dx_t = u_t(x_t)dt$$

Velocity field

The Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Diffusion

SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$

Drift

Diffusion Coefficient

Brownian Motion

The Fokker-Planck Equation

$$\partial_t p_t = -\operatorname{div}(p_t f_t) + \frac{1}{2} g_t^2 \nabla^2 p_t$$

Where are the probabilities?

Flows

ODE $dx_t = u_t(x_t)dt$

Velocity field

The Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Yes!

Diffusion

SDE $dx_t = f_t(x_t)dt + g_t dw_t$

Drift

Diffusion Coefficient

Brownian Motion

The Fokker-Planck Equation

$$\partial_t p_t = -\operatorname{div}(p_t f_t) + \frac{1}{2} g_t^2 \nabla^2 p_t$$

Need one more thing…

Can we build a generative model with these?

Diffusion and Score Models

SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$

Drift

Diffusion Coefficient

Brownian Motion

The Fokker-Planck Equation

$$\partial_t p_t = -\operatorname{div}(p_t f_t) + \frac{1}{2} g_t^2 \nabla^2 p_t$$

Need one more thing…

Diffusion and Score Models

Forward SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$

Data → Noise

The Fokker-Planck Equation

$$\partial_t p_t = -\operatorname{div}(p_t f_t) + \frac{1}{2} g_t^2 \nabla^2 p_t$$

Need one more thing…

Diffusion and Score Models

Forward SDE $dx_t = f_t(x_t)dt + g_t dw_t$ Data → Noise

Reverse SDE $d\bar{x}_t = (f_t(x_t) - g_t^2 \boxed{\nabla \log p_t})dt + g_t d\bar{w}_t$ Noise → Data

The Fokker-Planck Equation

$$\partial_t p_t = -\operatorname{div}(p_t f_t) + \frac{1}{2} g_t^2 \nabla^2 p_t$$

Need one more thing… The Score!

Diffusion and Score Models

Forward SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$ Data → Noise

Reverse SDE $$d\bar{x}_t = (f_t(x_t) - g_t^2 \boxed{\nabla \log p_t})dt + g_t d\bar{w}_t$$ Noise → Data

Learn the score by regressing to conditional scores:

$$\min_\theta \mathbb{E}_{p_{data}, p_t(x|x_{data})} \left[\| s_t^\theta(x) - \nabla \log p_t(x \,|\, x_{data}) \|^2 \right]$$

Simulation-free

Known SDEs: Variance Exploding
Variance Preserving

Where are the probabilities?

Flows

ODE $dx_t = u_t(x_t)dt$

Velocity field

The Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Yes!

Diffusion

SDE $dx_t = f_t(x_t)dt + g_t dw_t$

Drift

Diffusion Coefficient

Brownian Motion

The Fokker-Planck Equation

$$\partial_t p_t = -\operatorname{div}(p_t f_t) + \frac{1}{2} g_t^2 \nabla^2 p_t$$

Learn: score $\nabla \log p_t$

- Only Gaussian source
- Solution asymptotically reaches source

Where are the probabilities?

Flows

ODE $$dx_t = u_t(x_t)dt$$

Velocity field

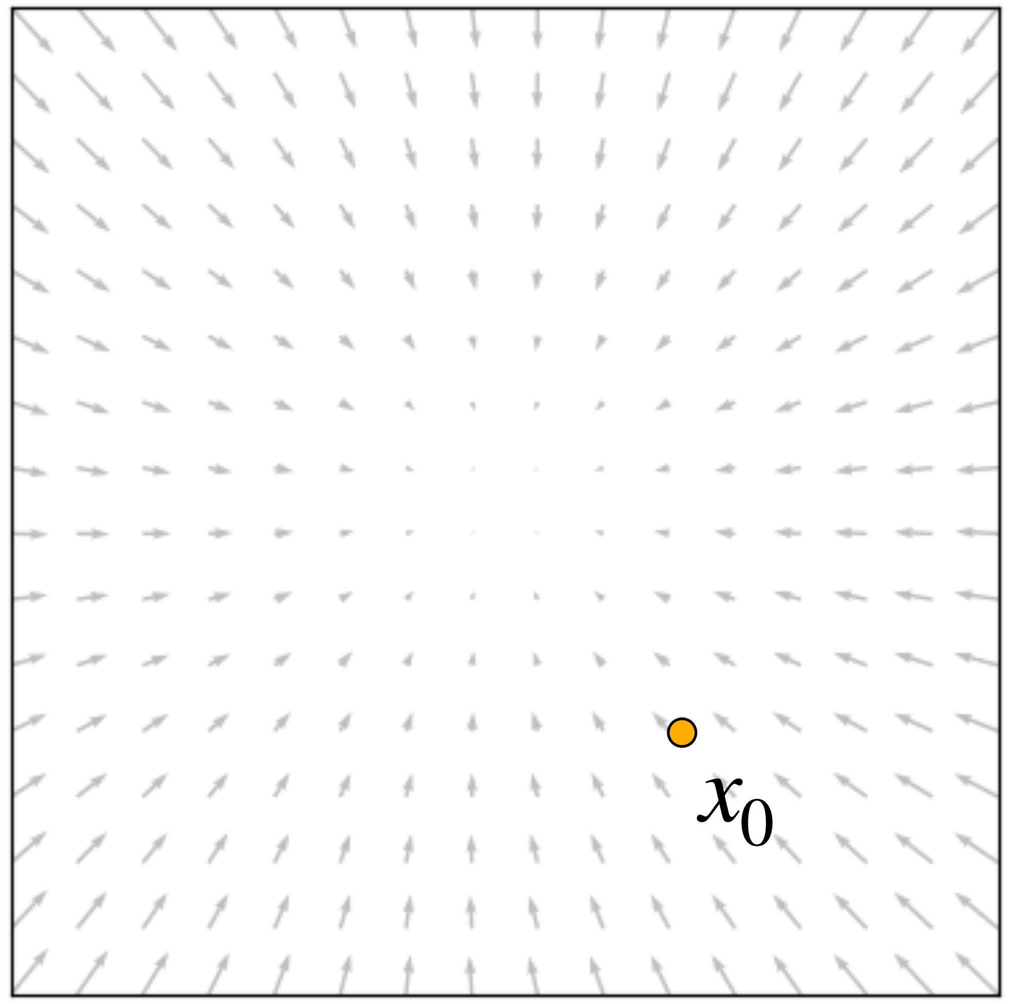
The Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Yes!

Flow ODE

$$\dot{\psi}_t(x_0) = u_t(\psi_t(x_0))$$



$\psi_t(x)$ is smooth with smooth **inverse** defined by $-u_t(x)$

Where are the probabilities?

Flow ODE

$$\dot{\psi}_t(x_0) = u_t(\psi_t(x_0))$$

The Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Learn: velocity field u_t

- Universal transformation between densities
- Defined on finite time interval

Diffusion

SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$

Drift

Diffusion Coefficient

Brownian Motion

The Liouville Equation

$$\partial_t p_t = -\operatorname{div}\left(p_t\left(f_t - \frac{1}{2}g_t^2 \nabla \log p_t\right)\right)$$

Learn: score $\nabla \log p_t$

- Only Gaussian source
- Solution asymptotically reaches source

Where are the probabilities?

Flow ODE

$$\dot{\psi}_t(x_0) = u_t(\psi_t(x_0))$$

The Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Learn: velocity field u_t

- Universal transformation between densities
- Defined on finite time interval

Diffusion

SDE $$\underbrace{dx_t = f_t(x_t)dt + g_t dw_t}$$

Drift

Diffusion Coefficient

Brownian Motion

The Liouville Equation

$$\partial_t p_t = -\operatorname{div}\left(p_t\left(f_t - \frac{1}{2}g_t^2 \nabla \log p_t\right)\right)$$

Learn: score $\nabla \log p_t$

- Only Gaussian source
- Solution asymptotically reaches source

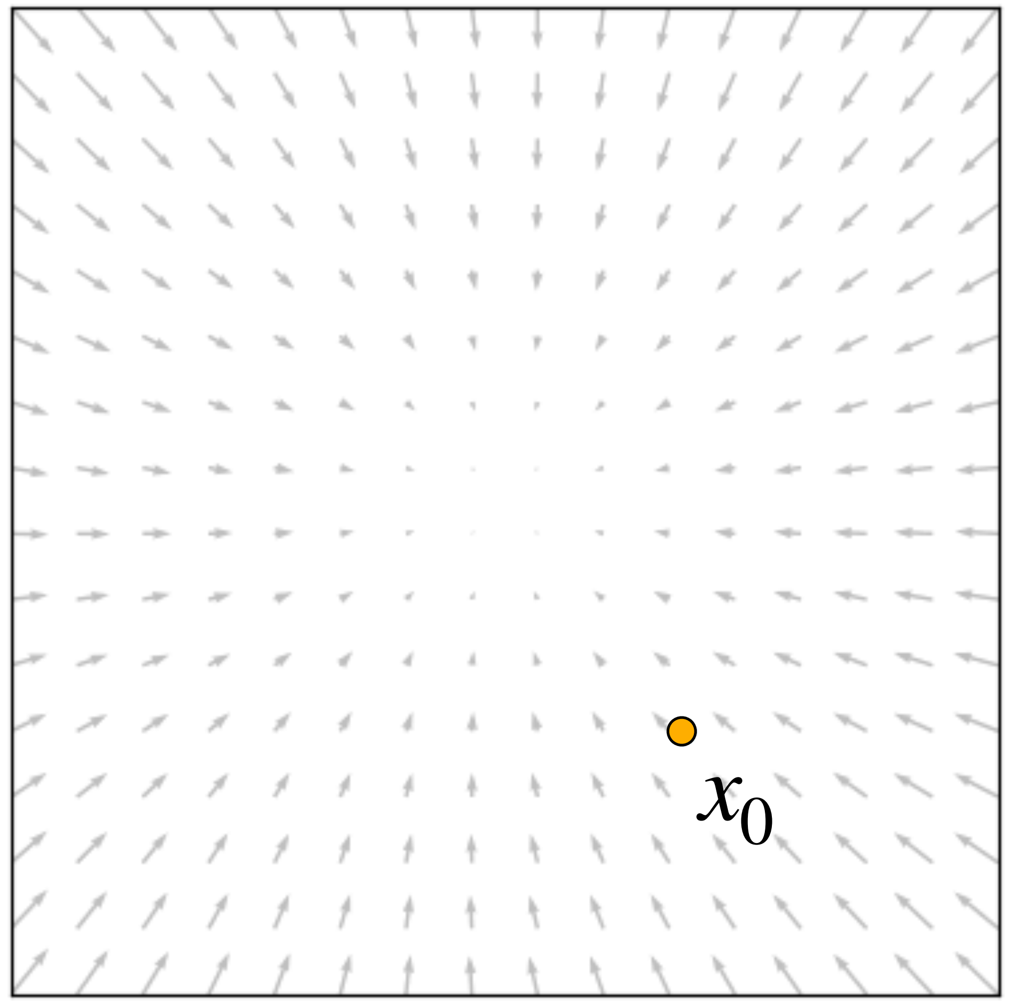
Flows as Generative Models

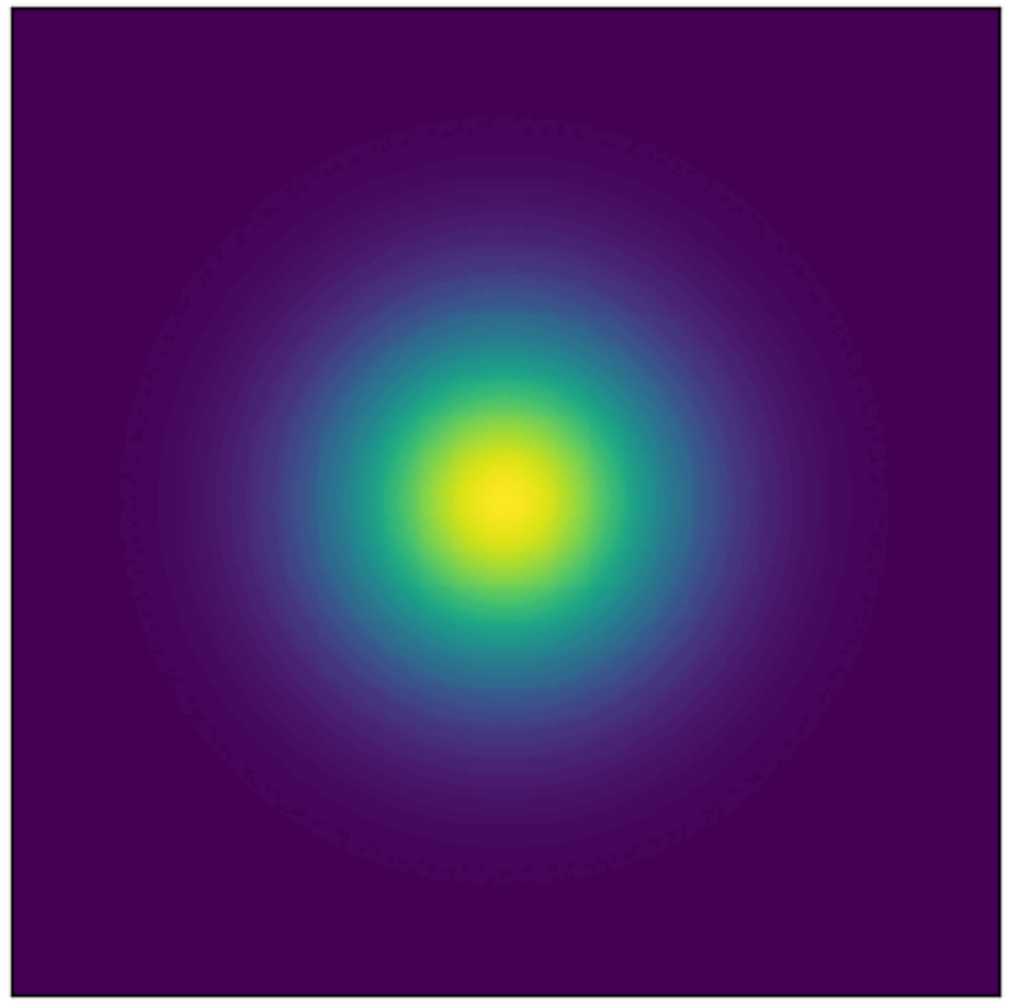
Flow ODE

$$\dot{\psi}_t(x_0) = u_t(\psi_t(x_0))$$

Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$





Learn: velocity field u_t

Goal: find velocity field u_t s.t. $p_1 \approx q$

Chen et al., *Neural Ordinary Differential Equations.* (NeurIPS 2018)

Training with Simulation

Flow ODE

$$\dot{\psi}_t(x_0) = u_t(\psi_t(x_0))$$

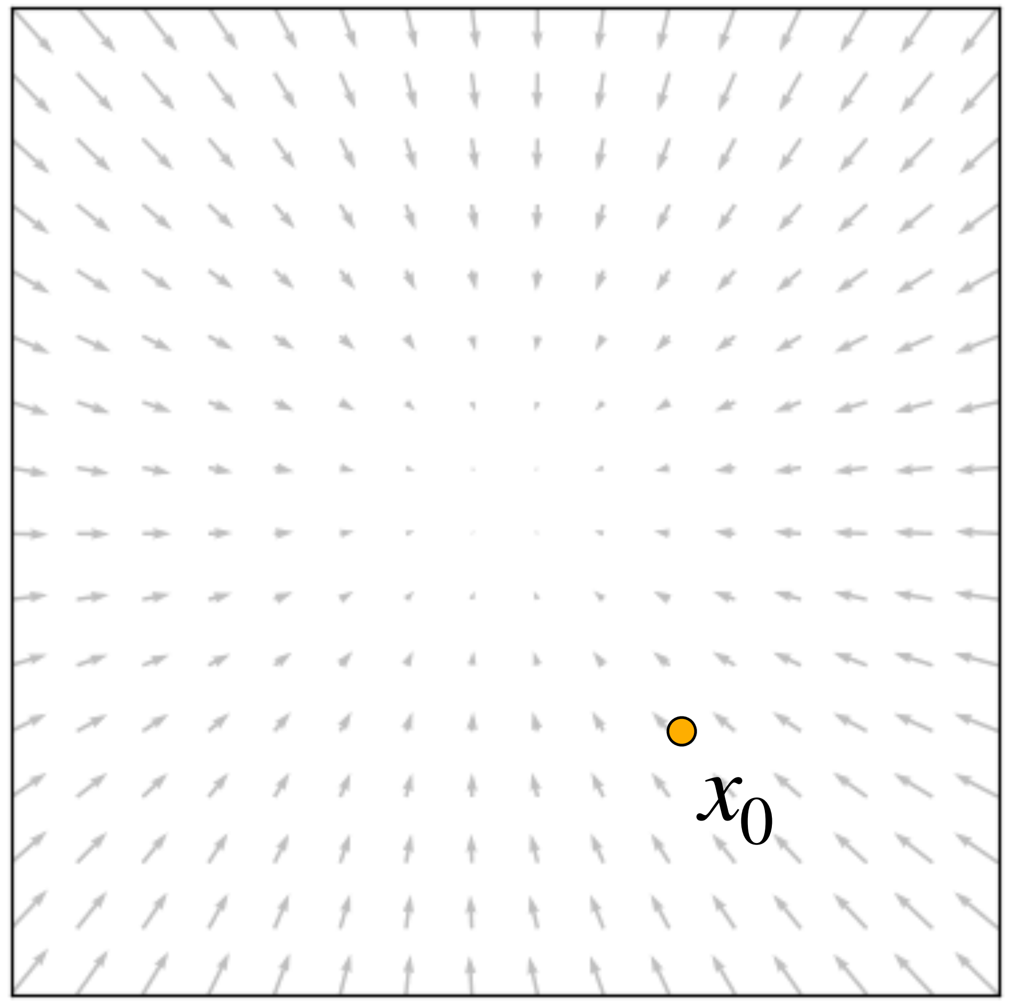
Continuity Equation

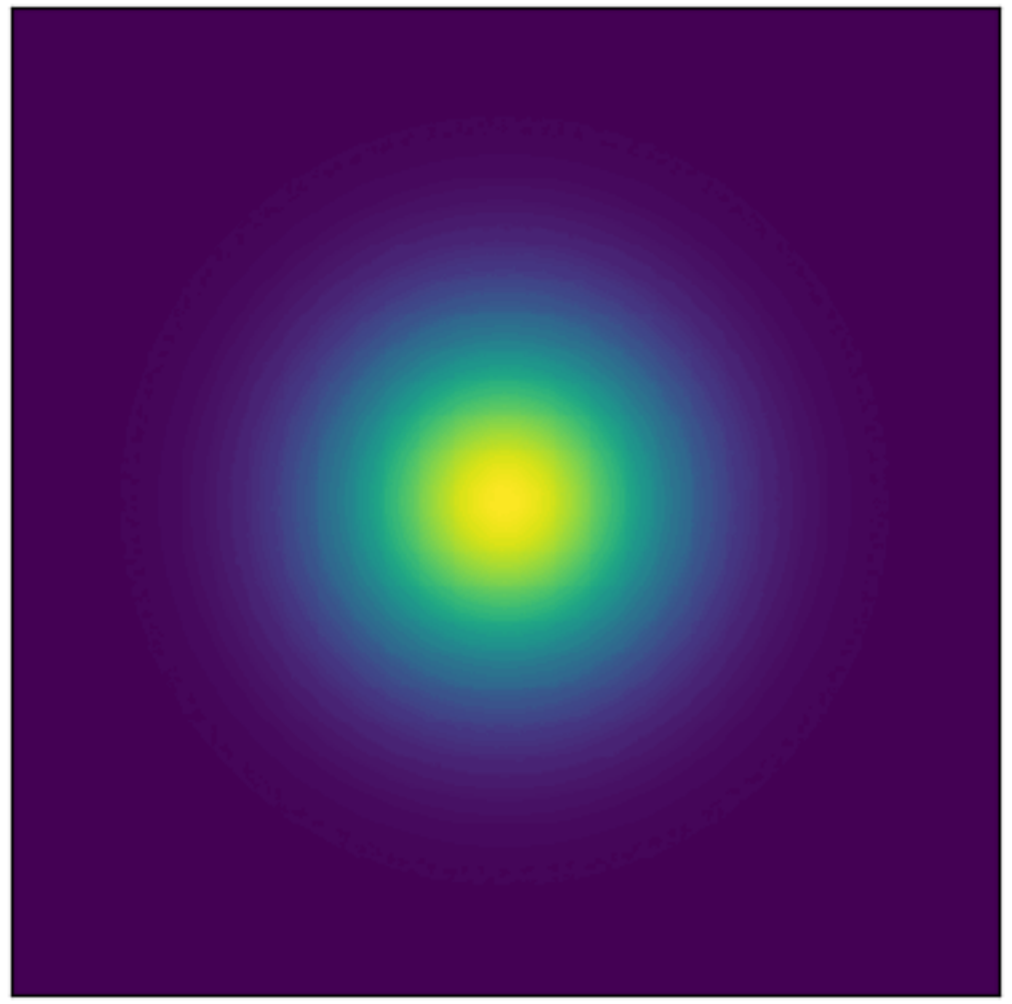
$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Log-likelihood computation

$$\log p_1(x_1) = \log p(x_0) + \int_1^0 \operatorname{div}(u_t(x_t))dt$$

$$x_t = x_1 + \int_1^t u_s(x_s)ds$$





Learn: velocity field u_t

Maximum Likelihood Objective

$$D_{\mathrm{KL}}(q \,\|\, p_1) = -\mathbb{E}_{x\sim q} \log p_1(x) + c$$

Chen et al., *Neural Ordinary Differential Equations.* (NeurIPS 2018)

Training with Simulation

Flows

ODE $$dx_t = u_t(x_t)dt$$

Velocity field

The Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Requires:

- Simulating x_t
- Backprop through simulation
- (Unbiased) estimator of $\operatorname{div}(u_t)$
- Can compute $\log p(x)$

Log-likelihood computation

$$\log p_1(x_1) = \log p(x_0) + \int_1^0 \operatorname{div}(u_t(x_t))dt$$

$$x_t = x_1 + \int_1^t u_s(x_s)ds$$

Maximum Likelihood Objective

$$D_{\mathrm{KL}}(q \,||\, p_1) = -\mathbb{E}_{x\sim q} \log p_1(x) + c$$

Chen et al., *Neural Ordinary Differential Equations.* (NeurIPS 2018)

Where are the probabilities?

Flows

ODE $$dx_t = u_t(x_t)dt$$

Velocity field

The Continuity Equation

$$\partial_t p_t = -\operatorname{div}(p_t u_t)$$

Diffusion

SDE $$dx_t = f_t(x_t)dt + g_t dw_t$$

Drift

Diffusion Coefficient

Brownian Motion

The Liouville Equation

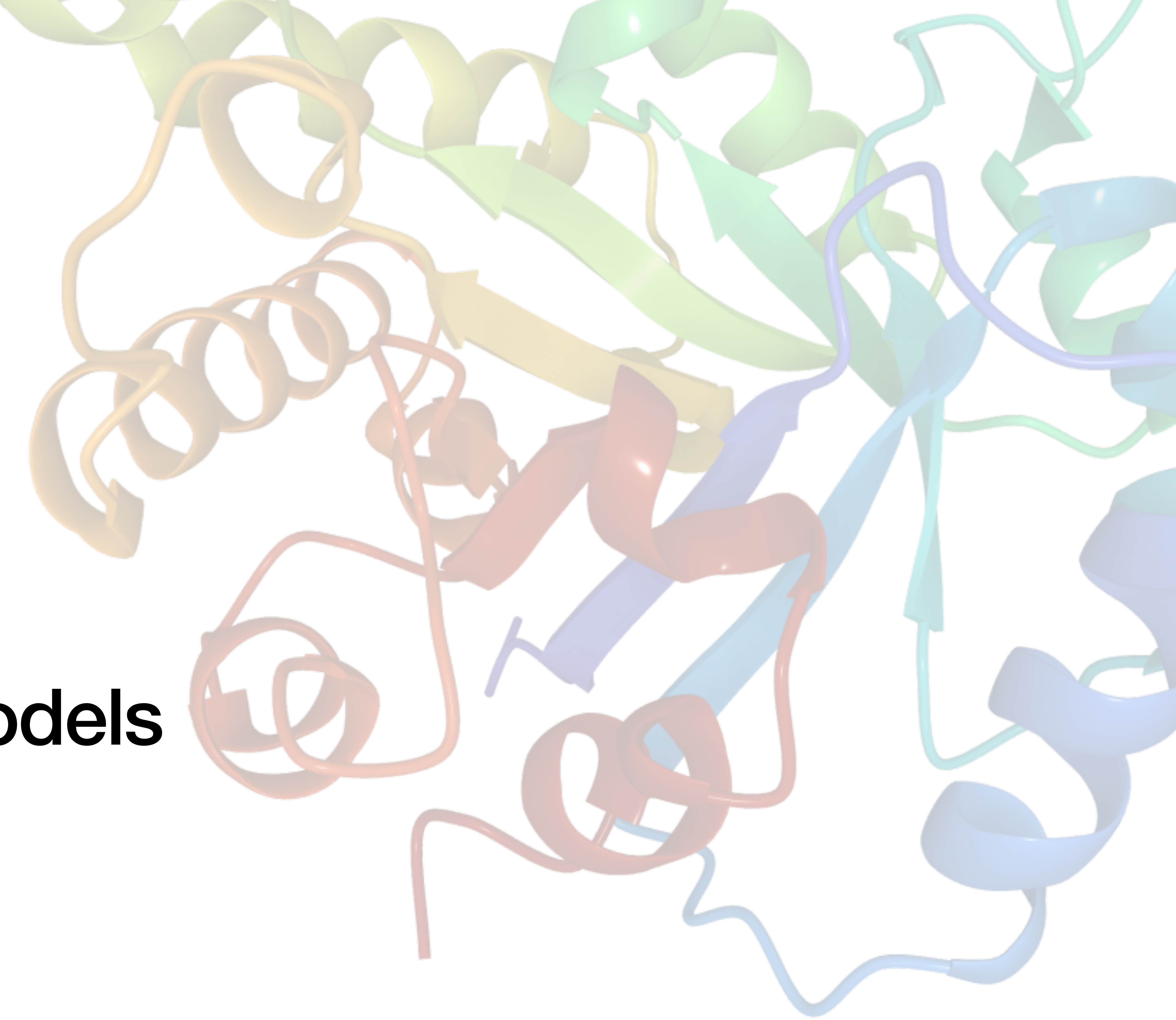
$$\partial_t p_t = -\operatorname{div}\left(p_t\left(f_t - \frac{1}{2}g_t^2 \nabla \log p_t\right)\right)$$

Core Principle:

u_t generates p_t
iff they satisfy the continuity equation

Outline: Part I

Part II:
Simulation Free Generative Models



Flow Matching

FLOW MATCHING FOR GENERATIVE MODELING

Yaron Lipman[1,2] **Ricky T. Q. Chen**[1] **Heli Ben-Hamu**[2] **Maximilian Nickel**[1] **Matt Le**[1]
[1]Meta AI (FAIR) [2]Weizmann Institute of Science

ICLR 2023

BUILDING NORMALIZING FLOWS WITH STOCHASTIC INTERPOLANTS

Michael S. Albergo
Center for Cosmology and Particle Physics
New York University
New York, NY 10003, USA
albergo@nyu.edu

Eric Vanden-Eijnden
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA
eve2@cims.nyu.edu

ICLR 2023

FLOW STRAIGHT AND FAST: LEARNING TO GENERATE AND TRANSFER DATA WITH RECTIFIED FLOW

Xingchao Liu*, **Chengyue Gong***, **Qiang Liu**
Department of Computer Science
University of Texas at Austin
{xcliu, cygong, lqiang}@cs.utexas.edu

ICLR 2023

Flow Matching

$$L_{\mathrm{FM}}(\theta) = \min \mathbb{E}_{t, p_t(x)} \| u_t^\theta(x) - u_t(x) \|^2$$

Core Principle:

u_t generates p_t
iff they satisfy the continuity equation

Construct:

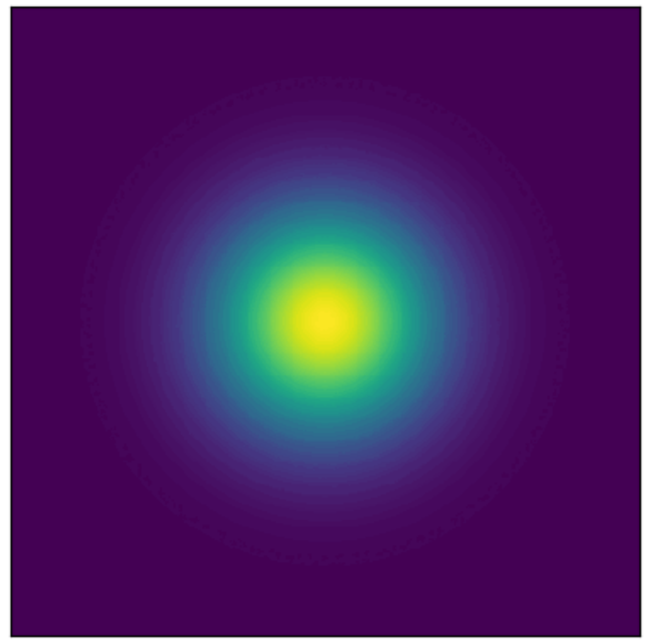
- Target probability path p_t s.t. $p_0 = p$, $p_1 \approx q$
- Generating velocity field u_t

Find a tractable optimization objective

Conditional Probability Paths

Law of total probability

$$p_t(x) = \int p_t(x \mid x_1) q(x_1) dx_1$$

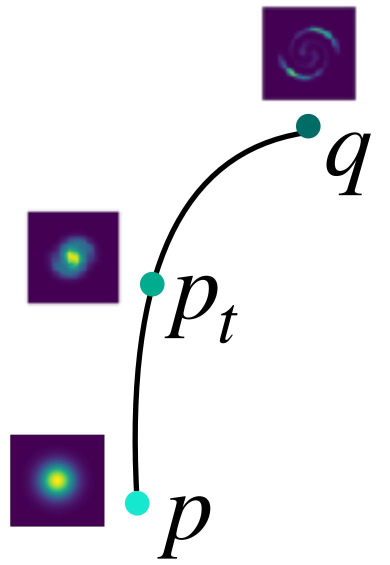


Marginal path

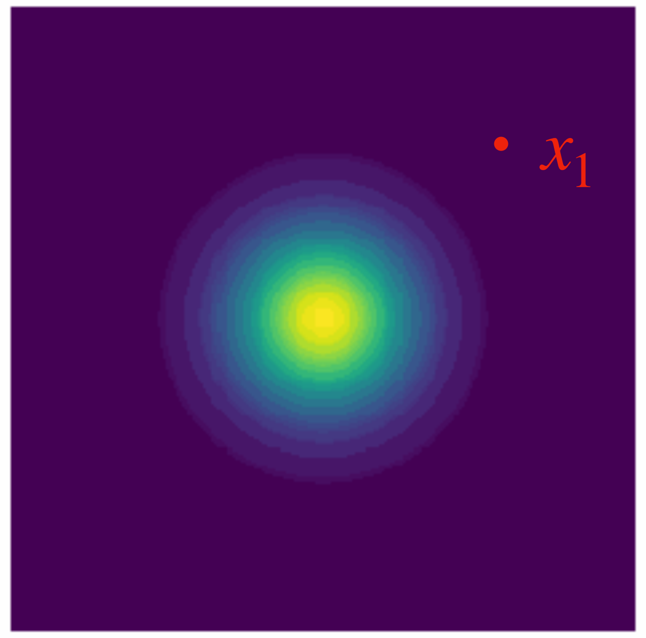
Boundary conditions:

$$p_0 = p$$

$$p_1 = q$$



$$p_t(x \mid x_1)$$



Conditional path

$$p_0(\cdot \mid x_1) = p$$

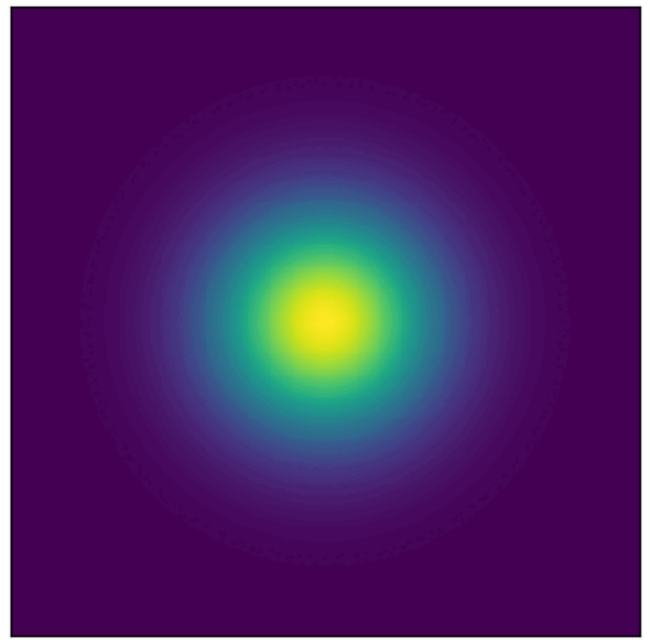
$$p_1(\cdot \mid x_1) = \delta_{x_1}$$

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

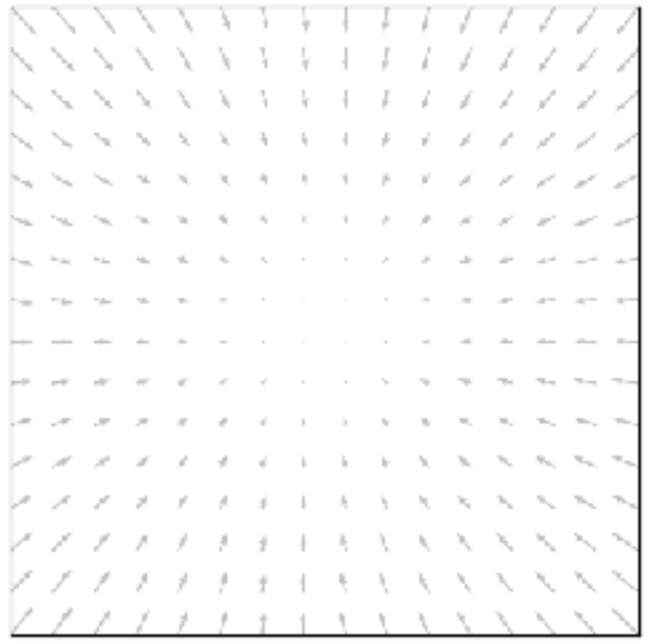
The marginalization “trick”

Marginal path

$$p_t(x) = \int p_t(x|x_1)q(x_1)dx_1$$

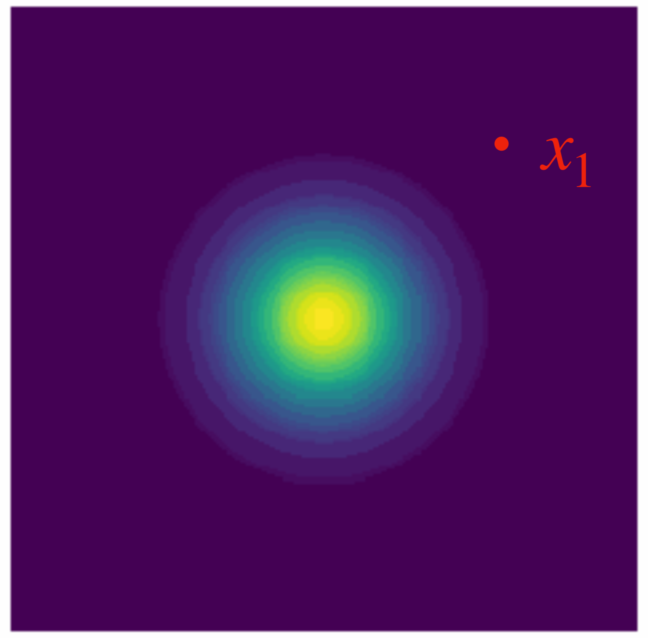


$$u_t(x) = \int u_t(x|x_1)\frac{p_t(x|x_1)q(x_1)}{p_t(x)}dx_1$$

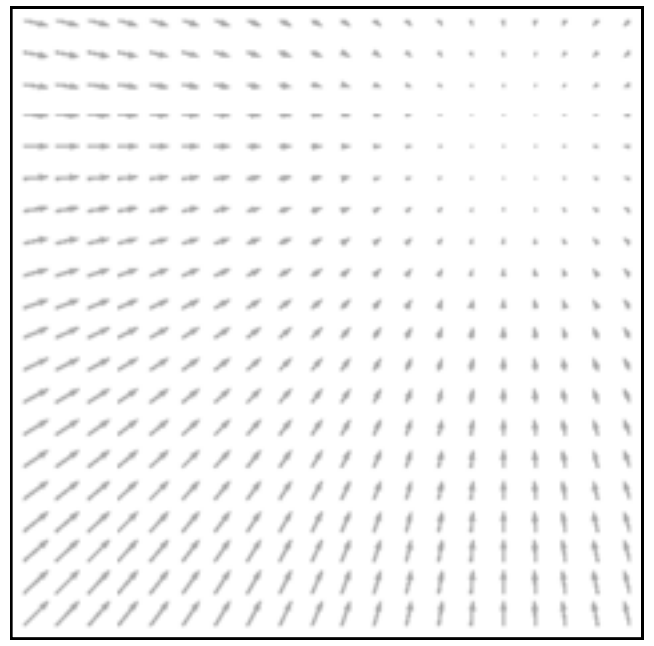


Conditional path

$$p_t(x|x_1)$$



$$u_t(x|x_1)$$



Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Flow Matching

$$L_{\mathrm{FM}}(\theta) = \min \mathbb{E}_{t,\, p_t(x)} \| u_t^\theta(x) - u_t(x) \|^2$$

$$u_t(x) = \int u_t(x \mid x_1) \frac{p_t(x \mid x_1) q(x_1)}{p_t(x)} dx_1$$

Construct:

- Target probability path p_t s.t. $p_0 = p$, $p_1 \approx q$
- Generating velocity field u_t

✔

Find a tractable optimization objective

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Flow Matching

$$L_{\mathrm{FM}}(\theta) = \min \mathbb{E}_{t,\, p_t(x)} \| u_t^{\theta}(x) - u_t(x) \|^2$$

$$u_t(x) = \int u_t(x \mid z) \frac{p_t(x \mid z) q(z)}{p_t(x)} dz$$

Useful example:

$z = (x_0, x_1)$

Construct:

- Target probability path p_t s.t. $p_0 = p$, $p_1 \approx q$
- Generating velocity field u_t

✔

Find a tractable optimization objective

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Conditional Flow Matching Loss

$$L_{\mathrm{FM}}(\theta) = \min \mathbb{E}_{t,\, p_t(x)} \| u_t^\theta(x) - u_t(x) \|^2$$

$$\Updownarrow$$

$$L_{\mathrm{CFM}}(\theta) = \min \mathbb{E}_{t,\, q(z),\, p_t(x|z)} \| u_t^\theta(x) - u_t(x \mid z) \|^2$$

The gradients of losses coincide:

$$\nabla_\theta L_{\mathrm{FM}} = \nabla_\theta L_{\mathrm{CFM}}$$

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Flow Matching

$$L_{\mathrm{FM}}(\theta) = \min \mathbb{E}_{t, p_t(x)} \| u_t^\theta(x) - u_t(x) \|^2$$

$$\Updownarrow$$

$$L_{\mathrm{CFM}}(\theta) = \min \mathbb{E}_{t, q(z), p_t(x|z)} \| u_t^\theta(x) - u_t(x \mid z) \|^2$$

Construct:

- Target probability path p_t s.t. $p_0 = p$, $p_1 \approx q$
- Generating velocity field u_t

✔

Find a tractable optimization objective

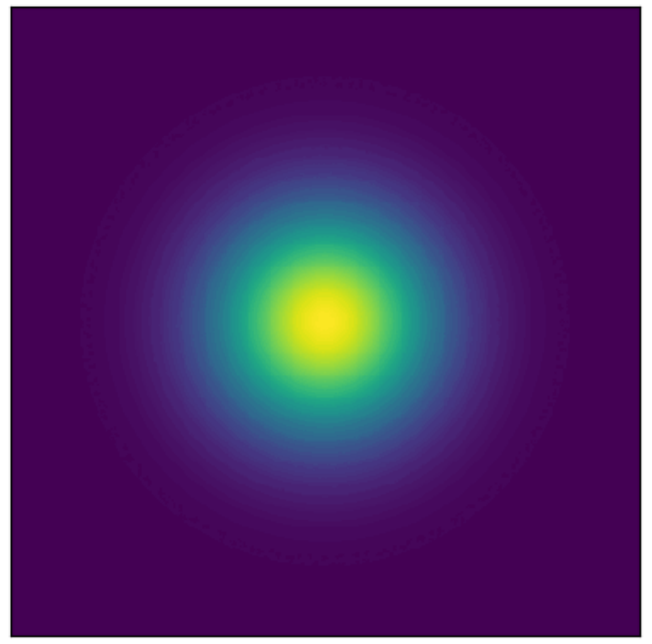
✔

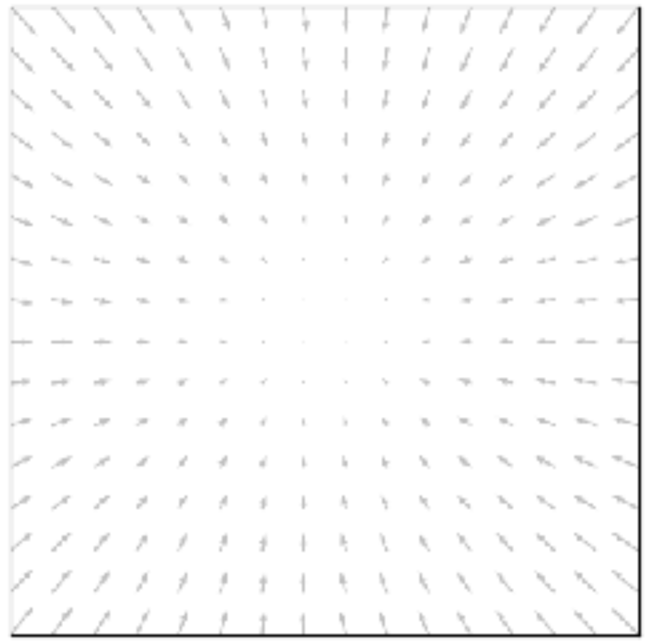
Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Conditional Flows

$$p_t(x) = \int p_t(x|x_1)q(x_1)dx_1 \qquad u_t(x) = \int u_t(x|x_1)\frac{p_t(x|x_1)q(x_1)}{p_t(x)}dx_1$$

Marginal path



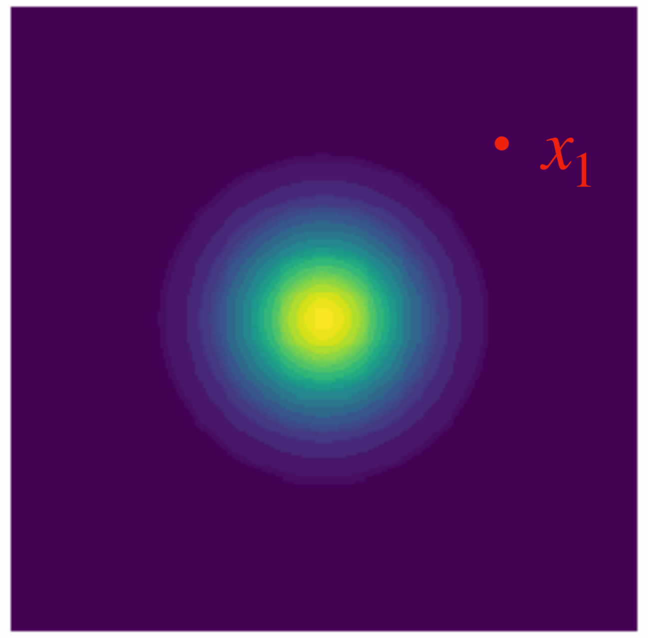


Construct a conditional flow s.t.

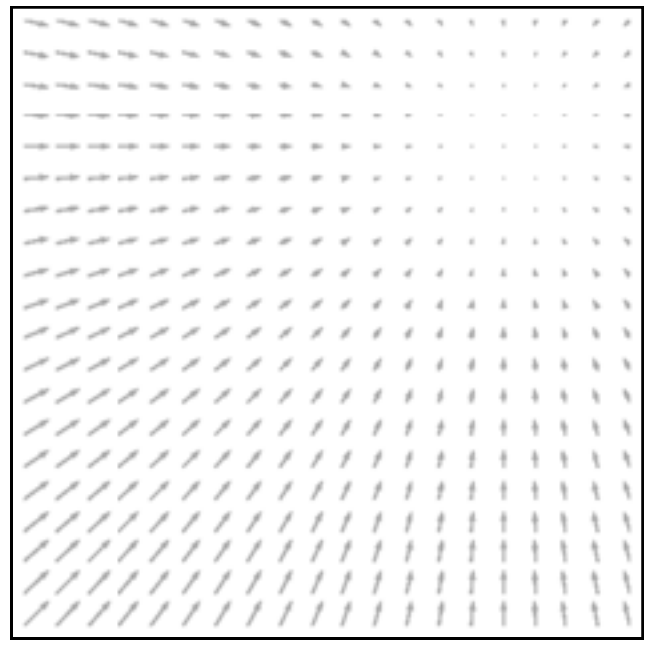
$$\psi_0(x|x_1) = x \ , \ \psi_1(x|x_1) = x_1$$

Conditional path

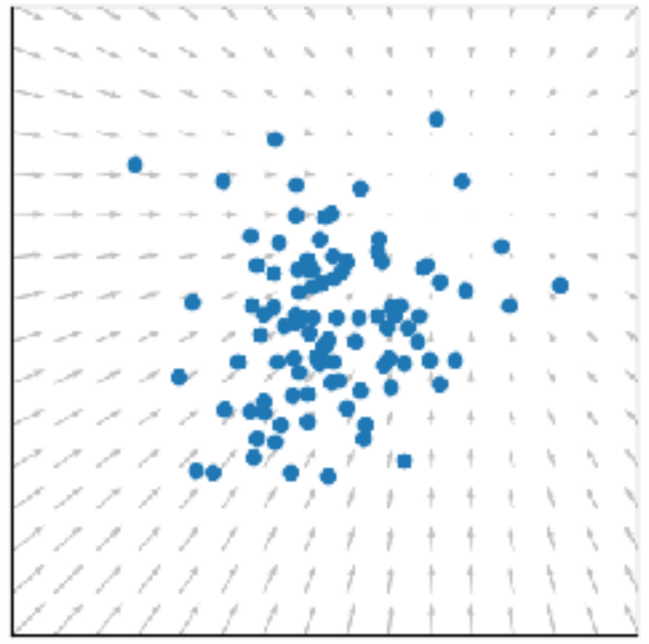
$p_t(x|x_1)$



$u_t(x|x_1)$



$\psi_t(x_0|x_1)$

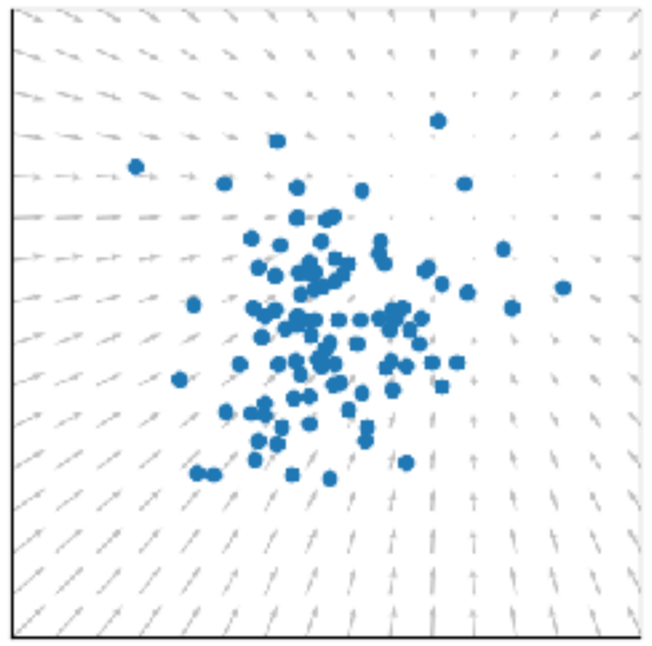


Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Affine Conditional Flows

Construct a conditional flow
s.t.
$\psi_0(x | x_1) = x \quad , \quad \psi_1(x | x_1) = x_1$

$\psi_t(x_0 | x_1)$



Affine conditional flow: $\psi_t(x_0 | x_1) = x_t = \alpha_t x_1 + \sigma_t x_0$

Conditional velocity field: $\frac{d}{dt}\psi_t(x_0 | x_1) = u_t(\psi_t(x_0 | x_1) | x_1)$

$$u_t(x_t | x_1) = \dot{\alpha}_t x_1 + \dot{\sigma}_t x_0$$

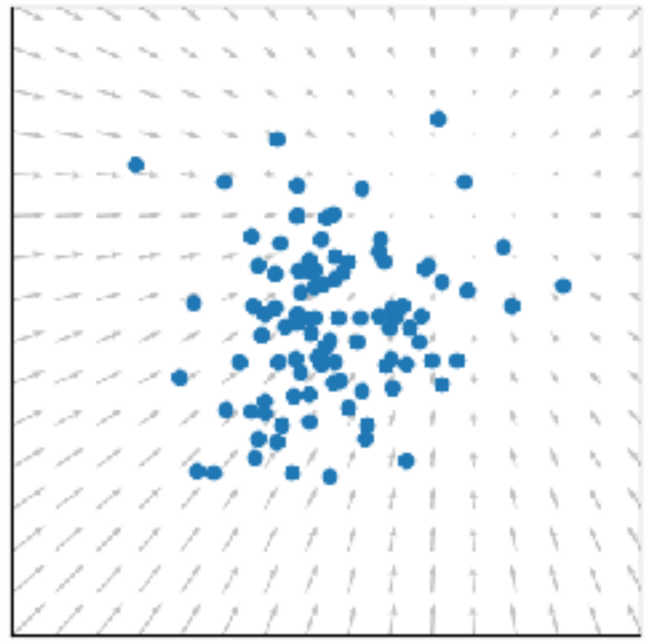
$$u_t(x | x_1) = \frac{\dot{\sigma}_t}{\sigma_t} x + (\dot{\alpha}_t - \alpha_t \frac{\dot{\sigma}_t}{\sigma_t}) x_1$$

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Affine Conditional Flows

Construct a conditional flow
s.t.
$\psi_0(x|x_1) = x \;\; , \;\; \psi_1(x|x_1) = x_1$

$\psi_t(x_0|x_1)$



Affine conditional flow: $\psi_t(x_0|x_1) = x_t = \alpha_t x_1 + \sigma_t x_0$

Conditional velocity field: $\frac{d}{dt}x_t = u_t(x_t|x_1)$

$$u_t(x_t|x_1) = \dot{\alpha}_t x_1 + \dot{\sigma}_t x_0$$

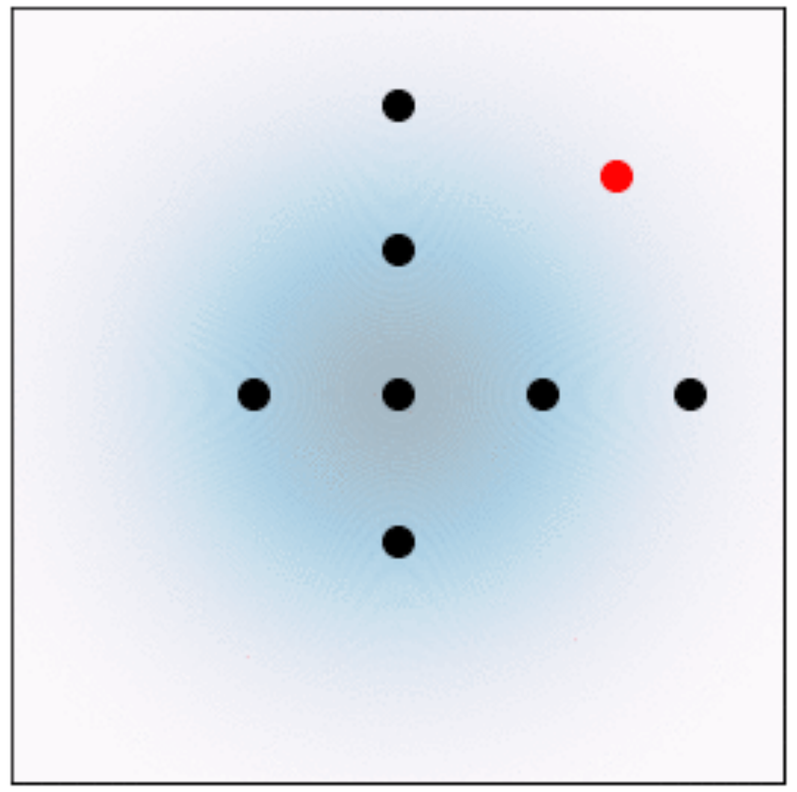
$$u_t(x_t|x_1) = \frac{\dot{\sigma}_t}{\sigma_t}x + (\dot{\alpha}_t - \alpha_t\frac{\dot{\sigma}_t}{\sigma_t})x_1$$

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Conditional Optimal Transport Flows

Construct a conditional flow
s.t.
$\psi_0(x|x_1) = x \quad , \quad \psi_1(x|x_1) = x_1$

$\psi_t(x_0|x_1)$



Cond-OT flow coefficients: $\alpha_t = t \quad , \quad \sigma_t = 1 - t$

Cond-OT flow: $\psi_t(x_0|x_1) = x_t = tx_1 + (1-t)x_0$

$$u_t(x_t|x_1) = x_1 - x_0$$

$$u_t(x|x_1) = \frac{x_1 - x}{1 - t}$$

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Gaussian Affine Conditional Flows

Construct a conditional flow
s.t.
$\psi_0(x|x_1) = x \;\;,\;\; \psi_1(x|x_1) = x_1$

Gaussian source: $x_0 \sim \mathcal{N}(0,I)$

Affine conditional flow: $\psi_t(x_0|x_1) = x_t = \alpha_t x_1 + \sigma_t x_0$

$$x_t \sim \mathcal{N}(\alpha_t x_1, \sigma_t^2 I)$$

Popular Diffusion Paths:

Variance Exploding: $x_t \sim \mathcal{N}(x_1, \sigma_t^2 I)$

Variance Preserving: $x_t \sim \mathcal{N}(\alpha_t x_1, (1-\alpha_t)_t I)$

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

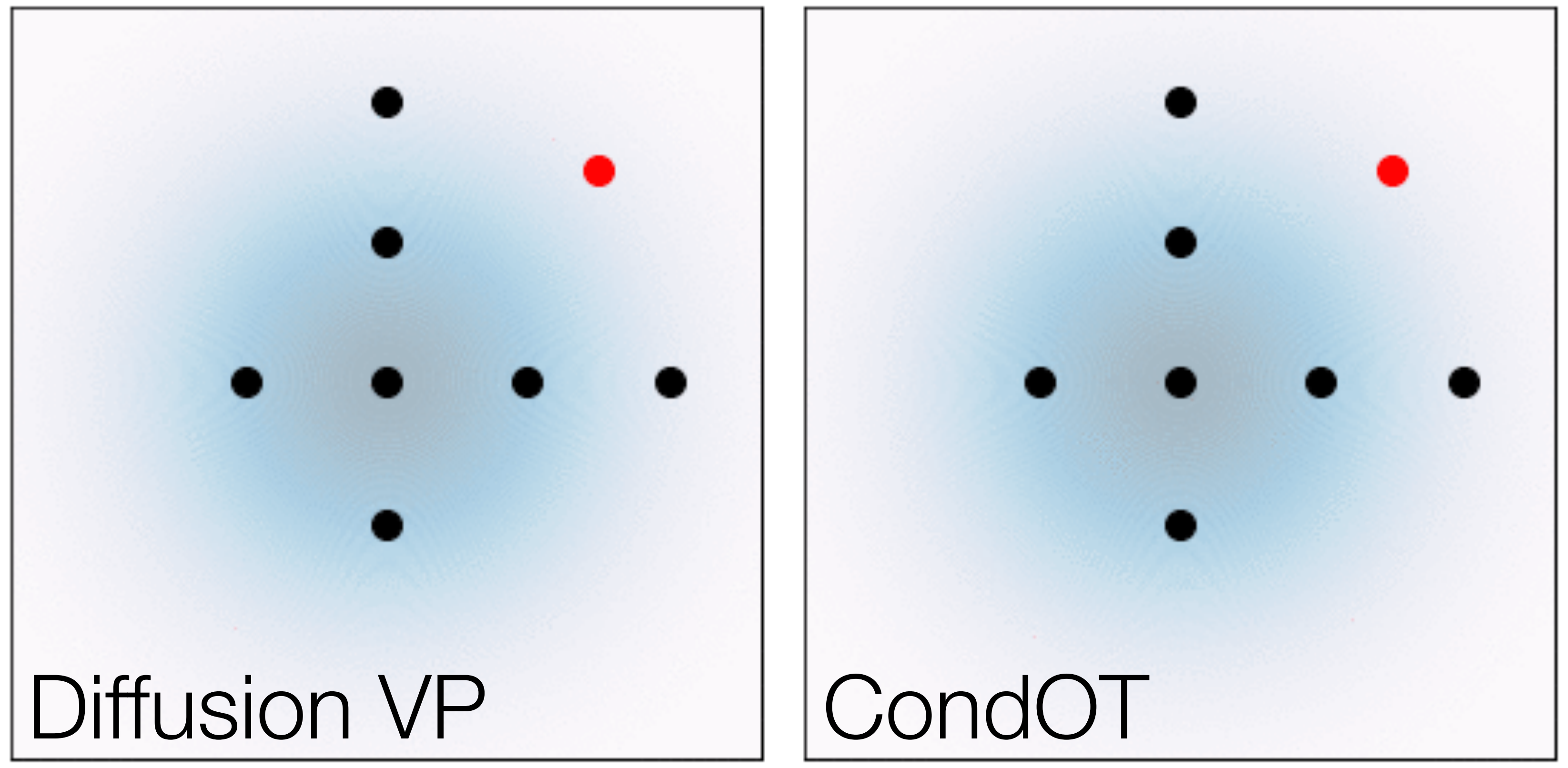
Gaussian Affine Conditional Flows

Construct a conditional flow
s.t.
$\psi_0(x\,|\,x_1) = x \quad , \quad \psi_1(x\,|\,x_1) = x_1$

Gaussian source: $x_0 \sim \mathcal{N}(0,I)$

Affine conditional flow: $\psi_t(x_0\,|\,x_1) = x_t = \alpha_t x_1 + \sigma_t x_0$

$$x_t \sim \mathcal{N}(\alpha_t x_1, \sigma_t^2 I)$$



Popular Diffusion Paths:

Variance Exploding: $x_t \sim \mathcal{N}(x_1, \sigma_t^2 I)$

Variance Preserving: $x_t \sim \mathcal{N}(\alpha_t x_1, (1-\alpha_t)_t I)$

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

Flow Matching

$$L_{\mathrm{FM}}(\theta) = \min \mathbb{E}_{t, p_t(x)} \| u_t^\theta(x) - u_t(x) \|^2$$

Add slide on types of prediction

$$\Updownarrow$$

$$L_{\mathrm{CFM}}(\theta) = \min \mathbb{E}_{t, q(z), p_t(x|z)} \| u_t^\theta(x) - u_t(x \mid z) \|^2$$

Construct:

- Target probability path p_t s.t. $p_0 = p$, $p_1 \approx q$
- Generating velocity field u_t

✔

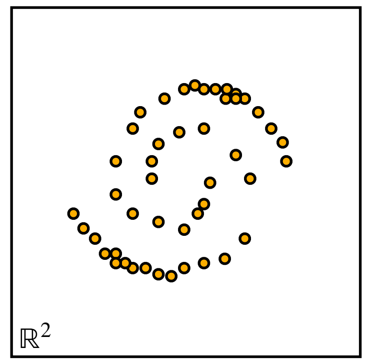
Find a tractable optimization objective

✔

Lipman et al., *Flow Matching for Generative Modeling.* (ICLR 2023)

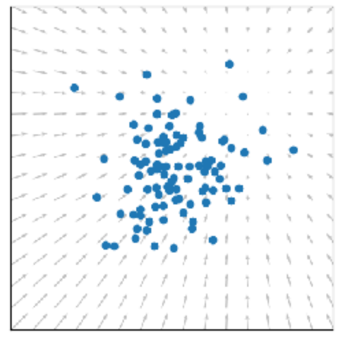
Recipe: Flow Matching

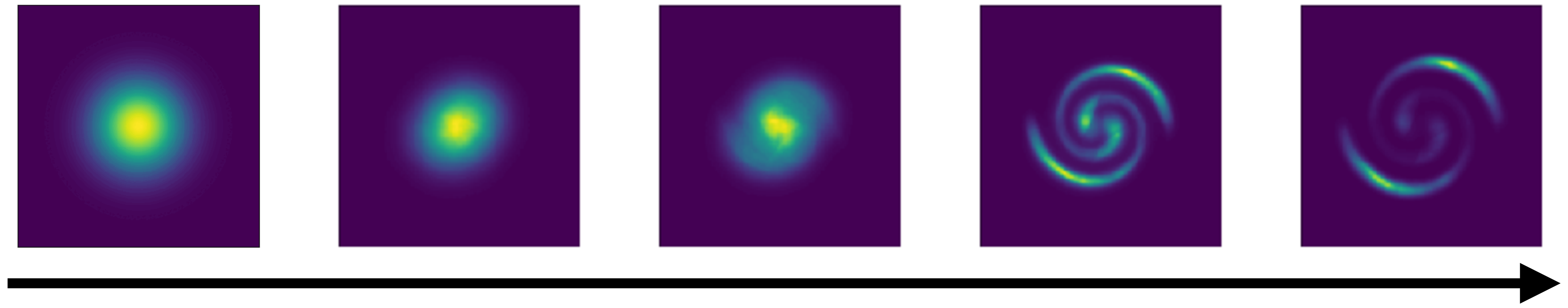
- Given: samples $x_1 \sim q$



- Construct: p_t s.t. $p_0 = p \ , p_1 \approx q$

  via conditional flows $\psi_t(x \mid z)$

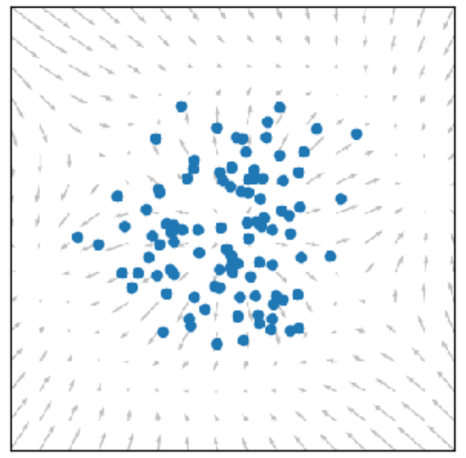




- Learn: velocity field u_t with CFM loss

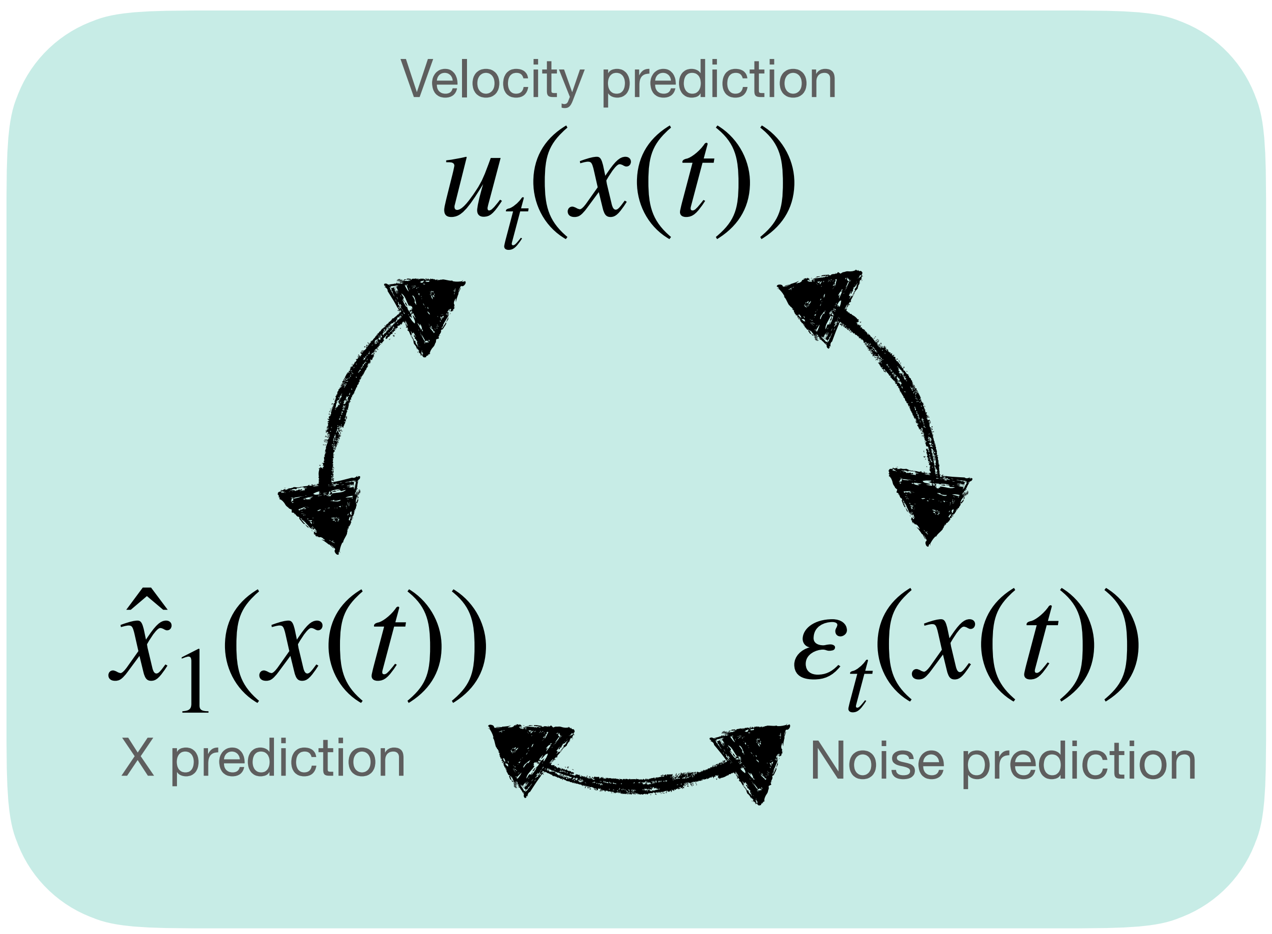
  s.t. $\psi_t(x_0) \sim p_t$ where $x_0 \sim p$

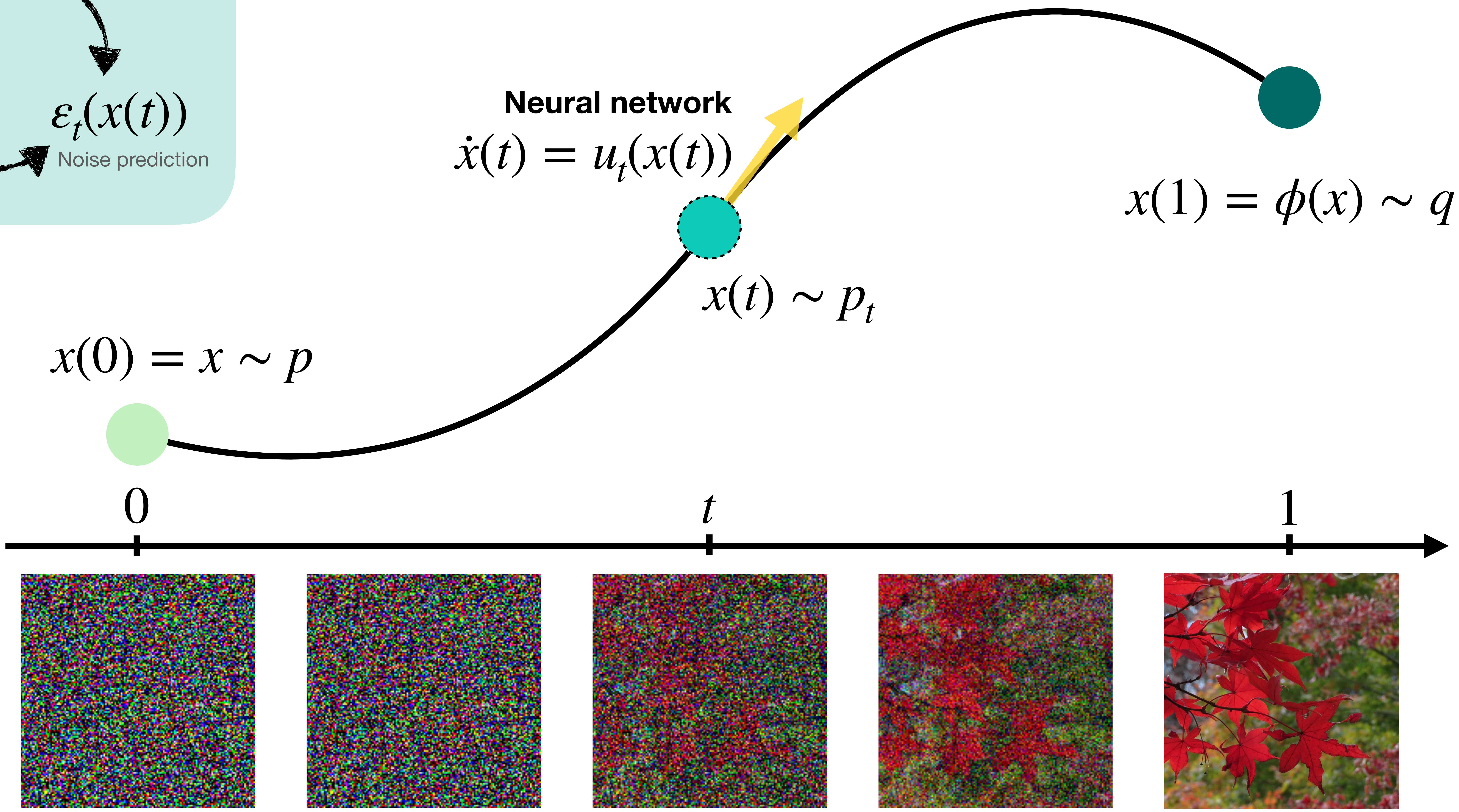
$$L_{\text{CFM}}(\theta) = \min \mathbb{E}_{t,\, q(z),\, p_t(x|z)} \| u_t^\theta(x) - u_t(x \mid z) \|^2$$



$\psi_t : [0,1] \times \mathbb{R}^2 \to \mathbb{R}^2$

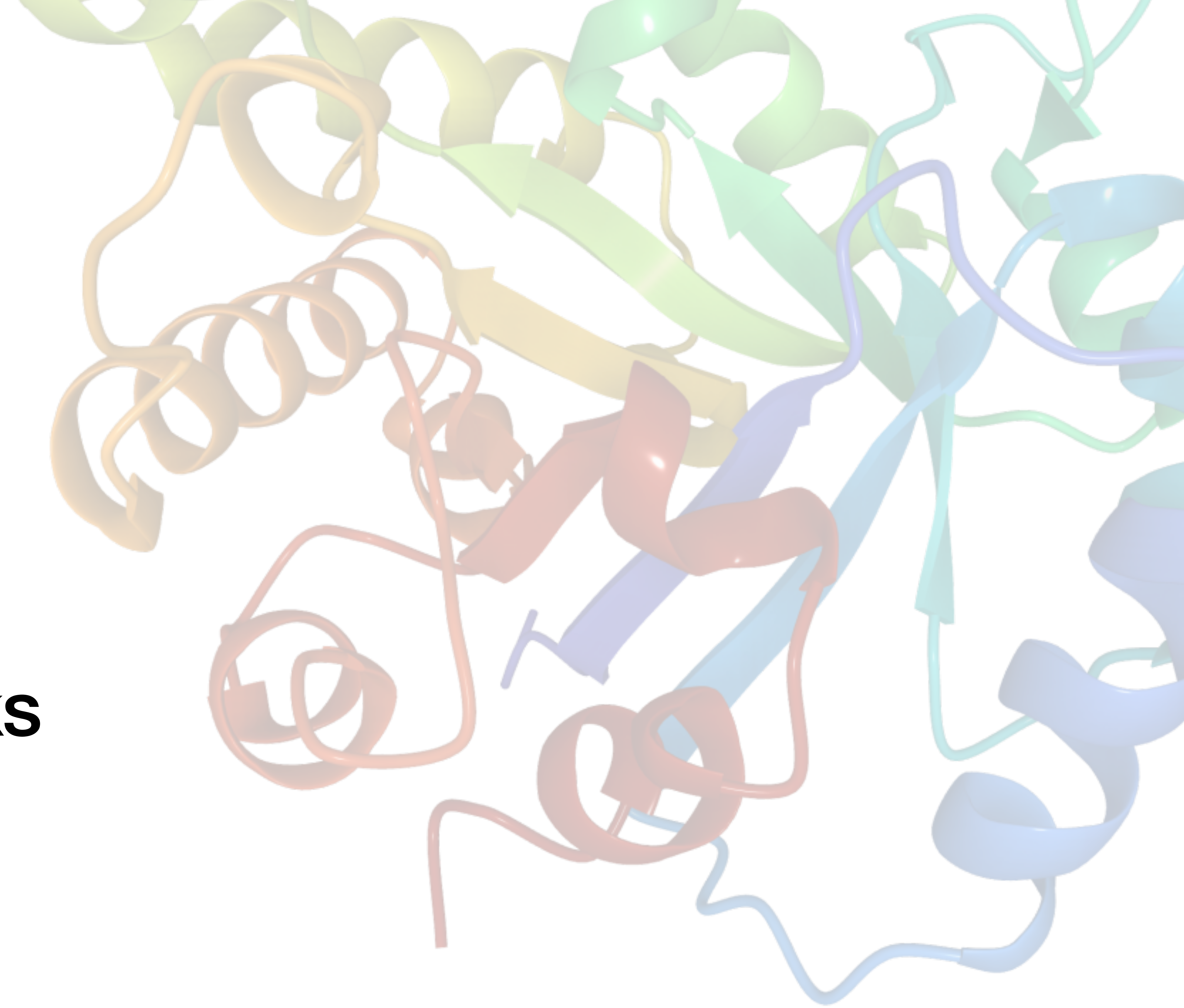
Flow/Diffusion Models





Ho J., Jain A., Abbeel P. *Denoising Diffusion Probabilistic Models* (NeurIPS 2020)
Lipman Y., Chen RTQ., **BH**, Nickel M., Le M. *Flow Matching for Generative Modeling* (ICLR 2023)

Part III: Applications and Tricks



Outline

- Images
- Image-to-image translation
- Cells

Slide Template

Slide text

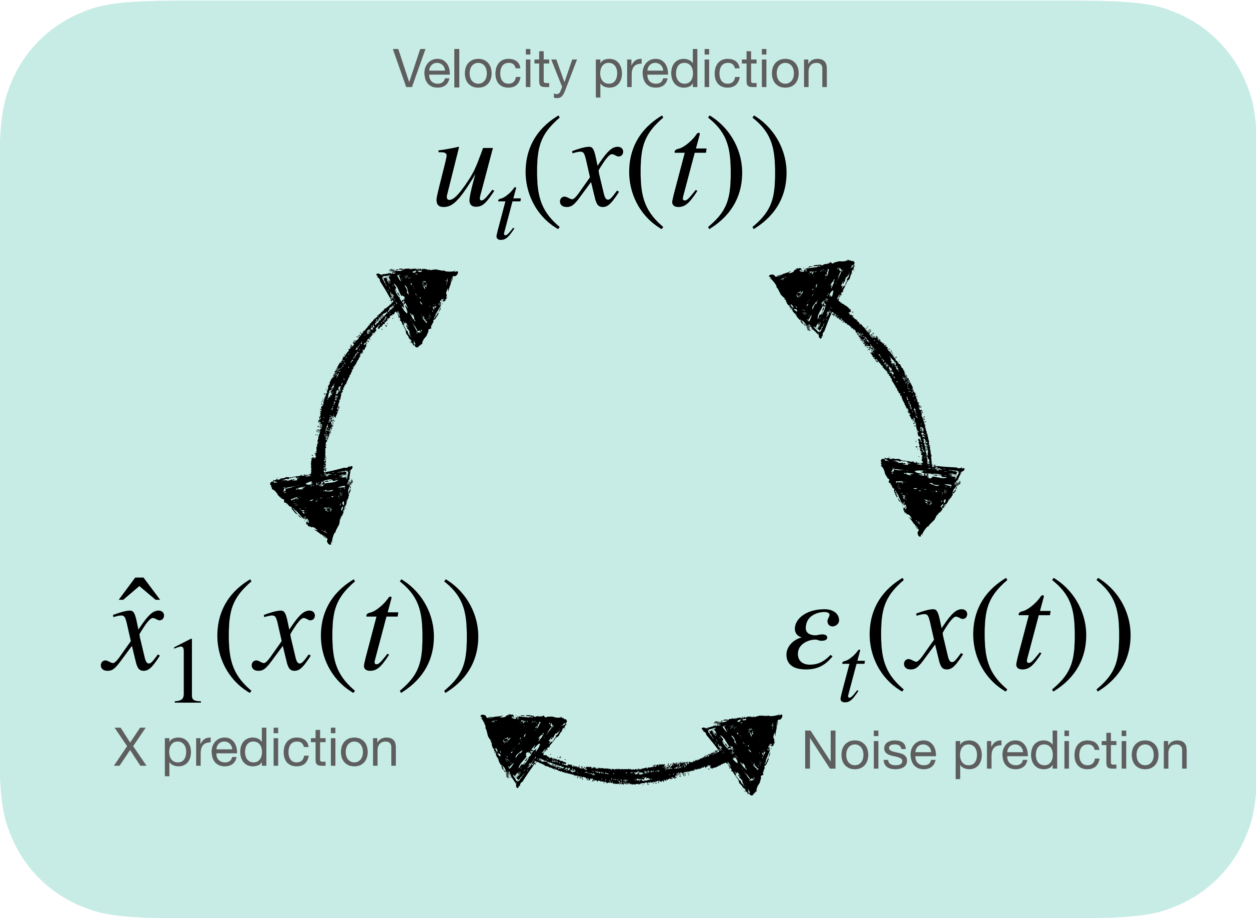
Colors

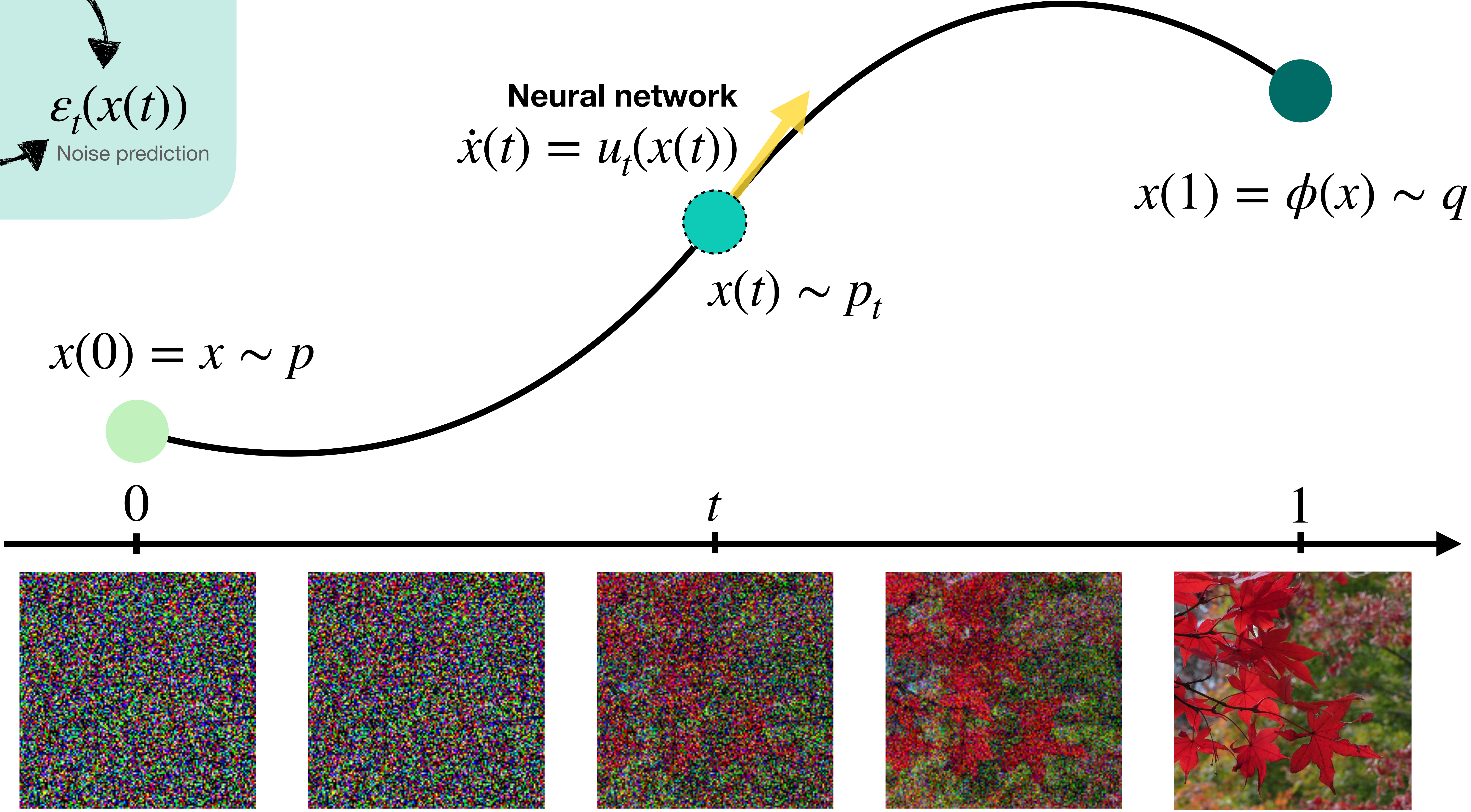
Colors

Highlight boxes

Exploration/discussion boxes

Flow/Diffusion Models





Ho J., Jain A., Abbeel P. *Denoising Diffusion Probabilistic Models* (NeurIPS 2020)
Lipman Y., Chen RTQ., **BH**, Nickel M., Le M. *Flow Matching for Generative Modeling* (ICLR 2023)