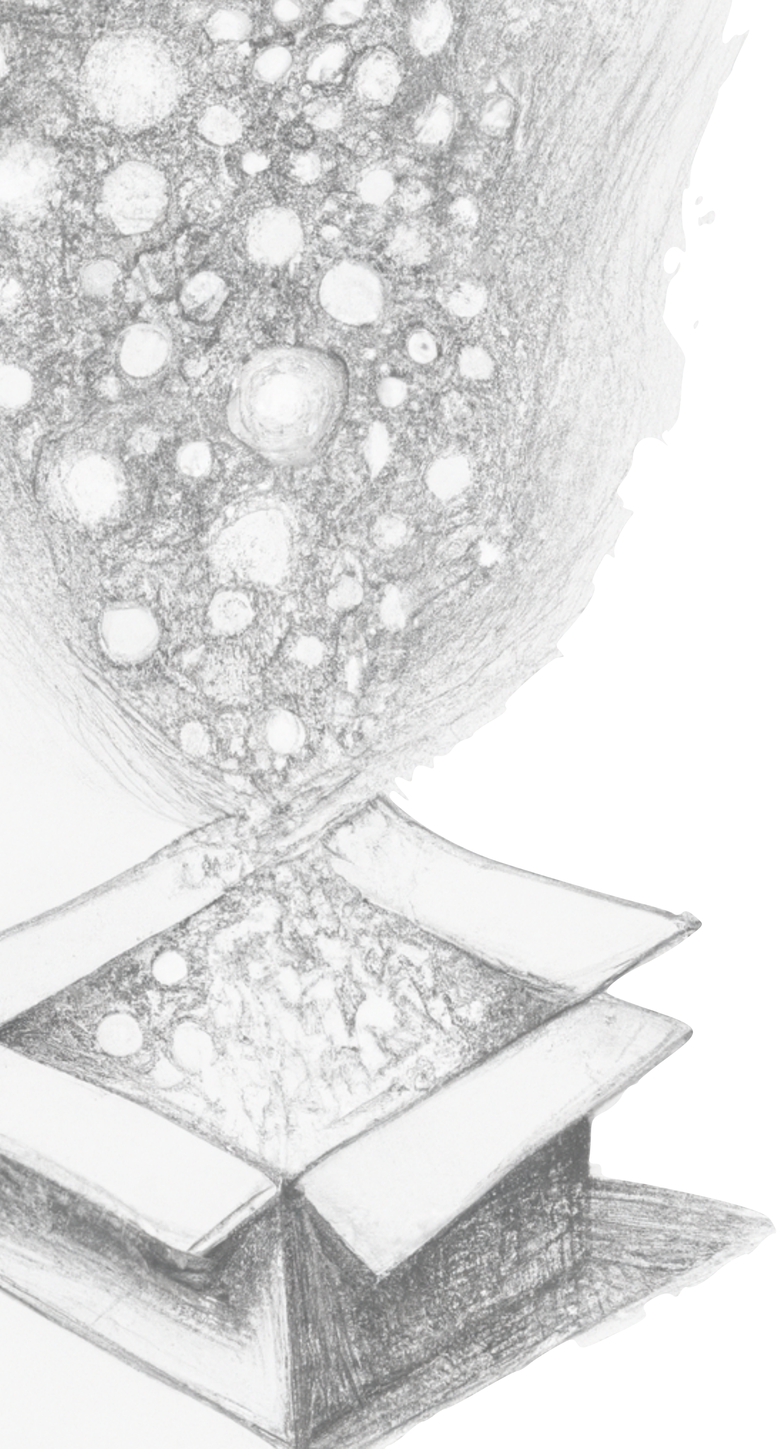


Generative modeling

Siddharth Mishra-Sharma

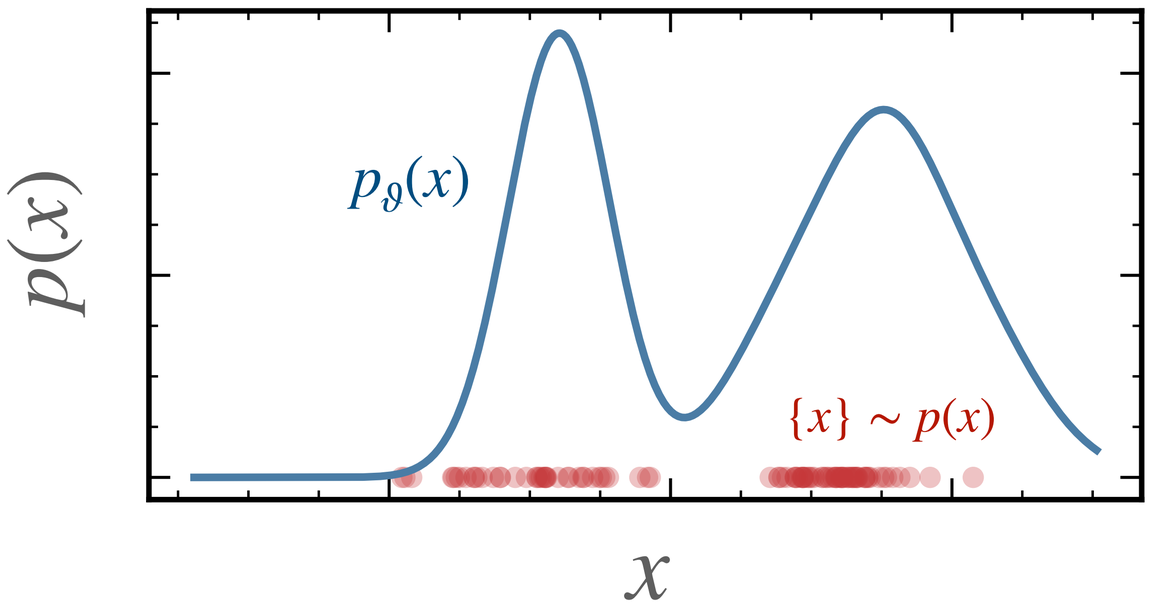
CDS DS 595

https://smsharma.io/teaching/ds595-ai4science.html

Generative models

Generative models are simulators of the data

Goal: learn a probability distribution $p_\vartheta(x)$ that is as close as possible to the true underlying data distribution $p(x)$



1. Sampling

$$x \sim p_\vartheta(x)$$

2. Density estimation

$$\log p_\vartheta(x)$$

Evolution of deep generative models

Variational autoencoders
(from Kingma et al 2013)

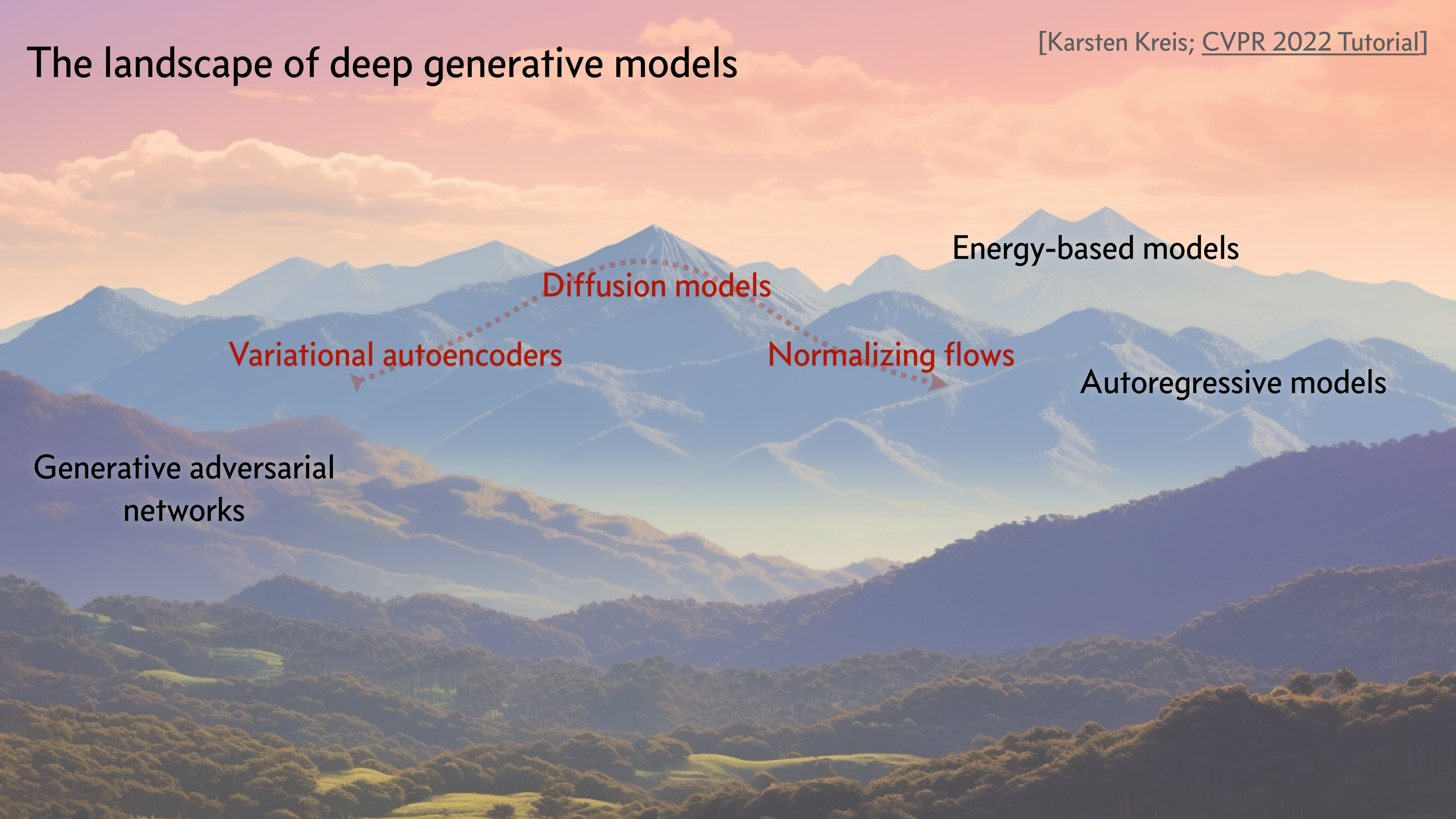
Diffusion models
(*Midjourney* 2023)

Seedance 2.0 (Feb 2026)

https://x.com/lexx_aura/status/2022994297881280557?s=20

The landscape of deep generative models

[Karsten Kreis; CVPR 2022 Tutorial]



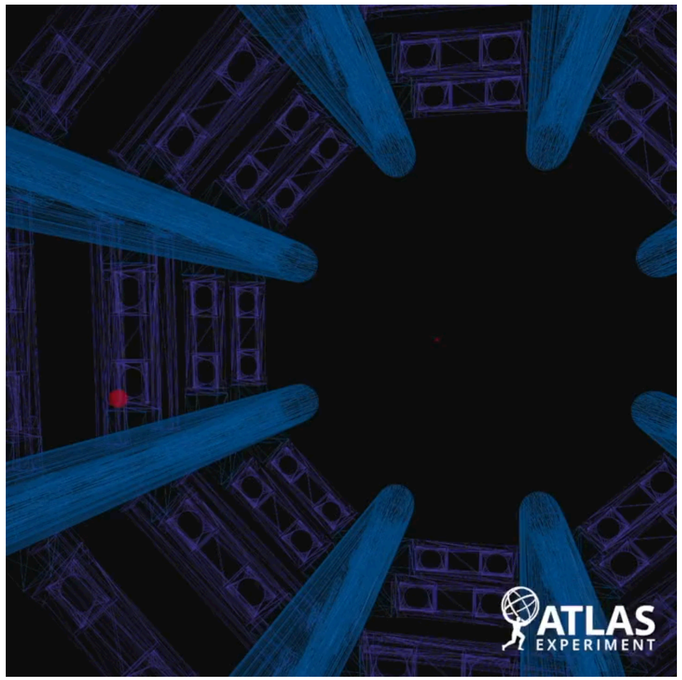
Simulators

$$x \sim p(x)$$

Simulators are ubiquitous: *they prescribe a way to sample from the data distribution*

Collider data

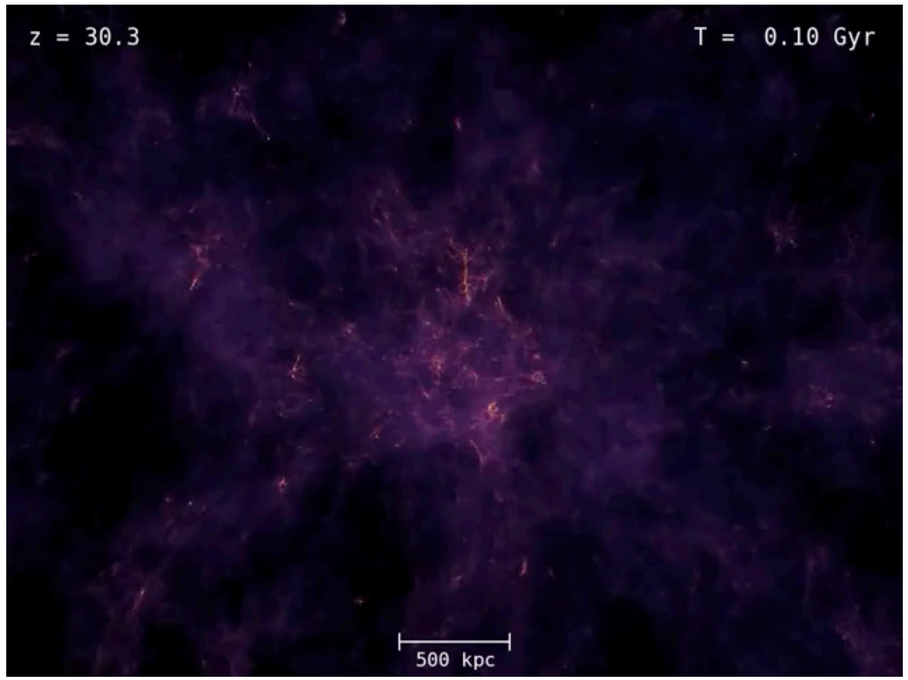
particles $\sim p$(particles)



[C. Cesarotti with ATLAS]

Cosmology data

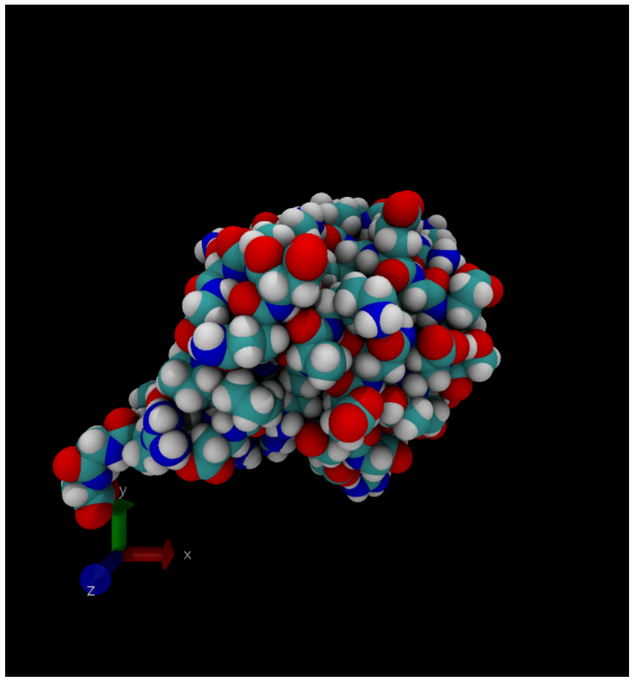
particles $\sim p$(particles)



[Aquarius simulation]

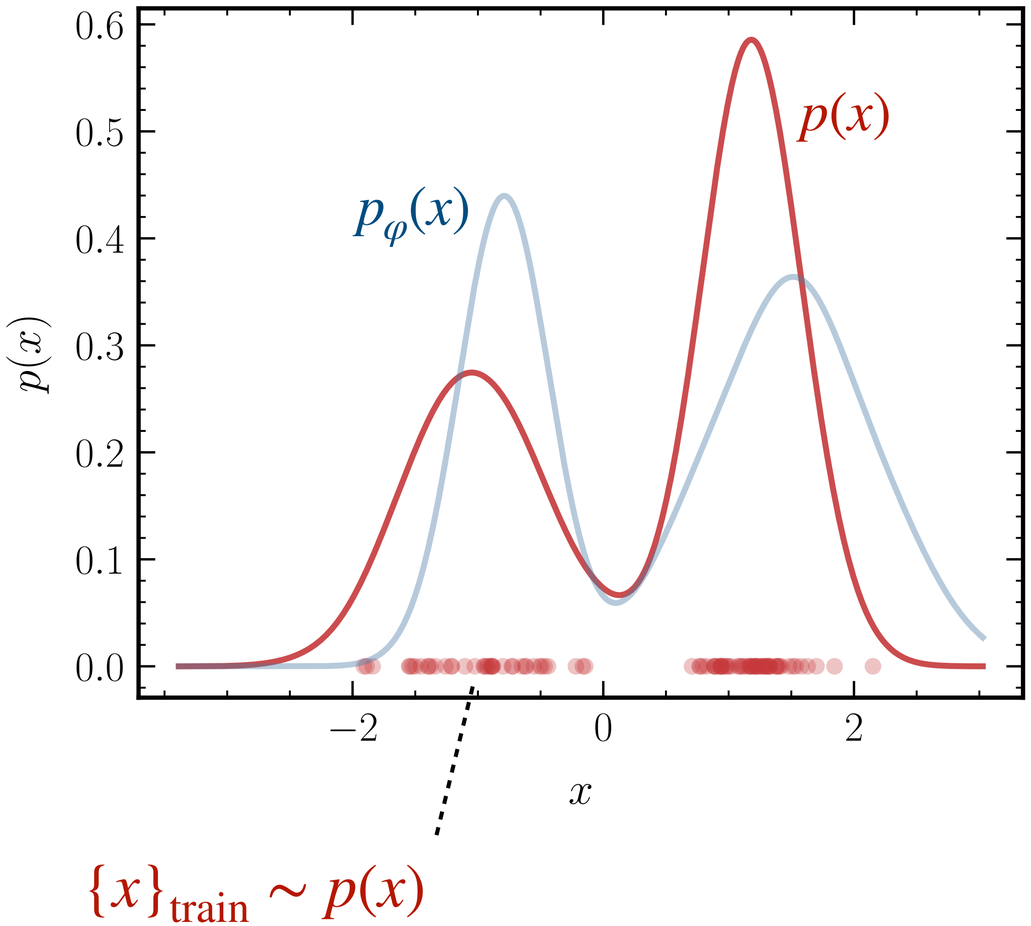
Molecular dynamics

configurations $\sim p$(configurations)



[E. Cances et al]

Learning the data distribution



1. Ingredients:

- A parameterized distribution $p_\varphi(x)$
- Samples from the data distribution $\{x\}_{\text{train}} \sim p(x)$

(empirical or simulated)

2. Maximize the likelihood of the model under the training data samples

$$\hat{\varphi} = \arg\max_{\varphi} \left[\log p_\varphi \left(\{x\}_{\text{train}} \right) \right]$$

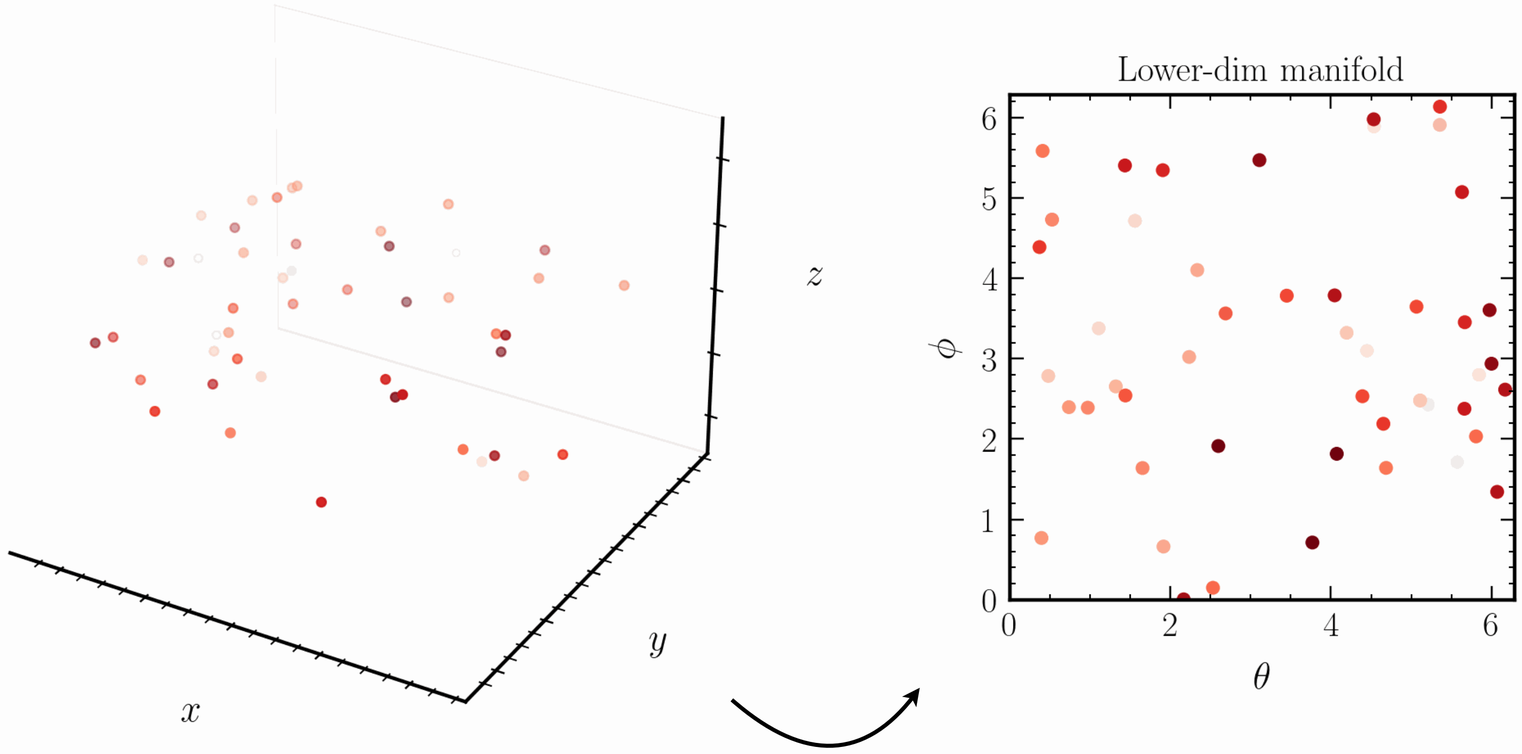
Not so fast…

The pursuit of low-dimensional structure

Real-world datasets often live in structured low-dimensional manifolds

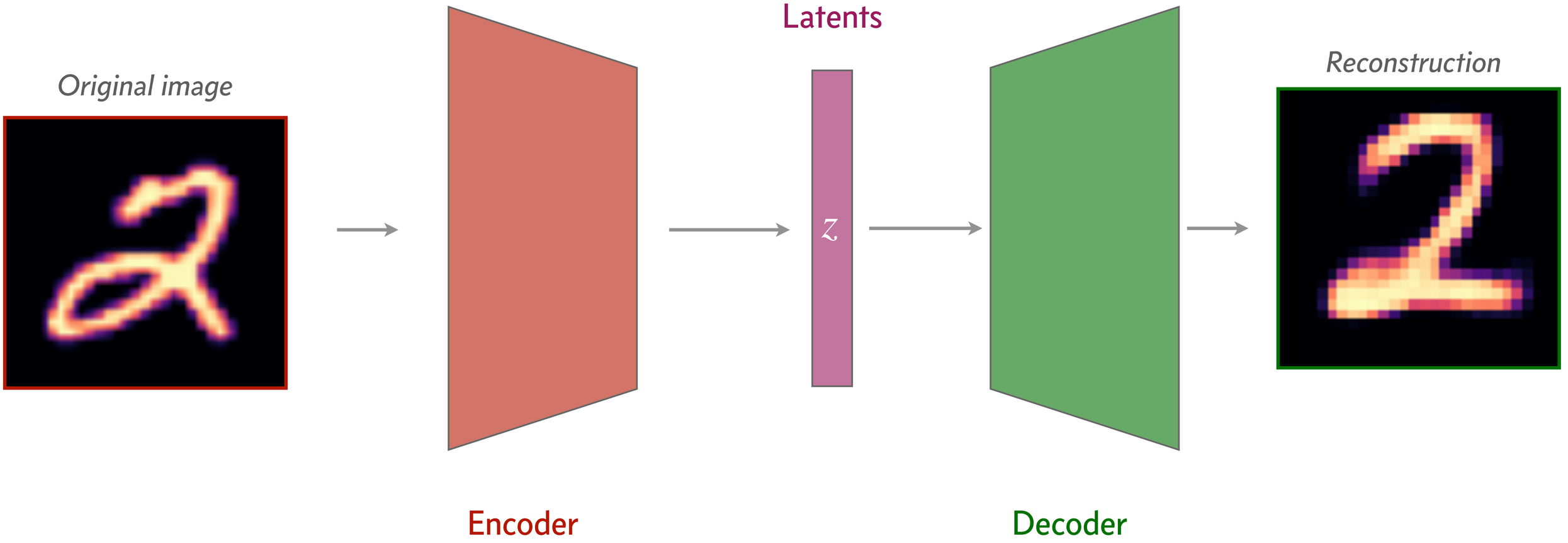
"Difficult to model" x

"Easy to model" z



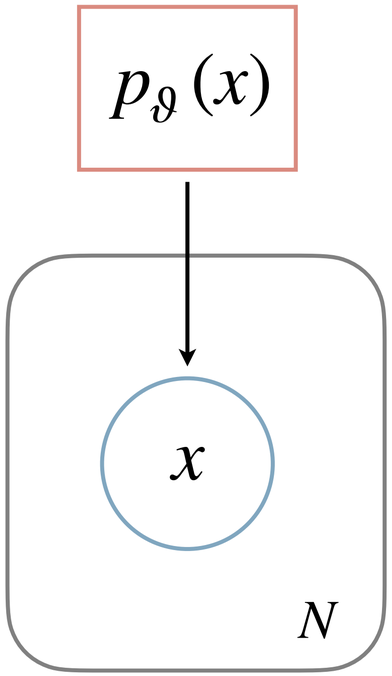
$$p(z \mid x)$$

Autoencoders



Latent-variable modeling

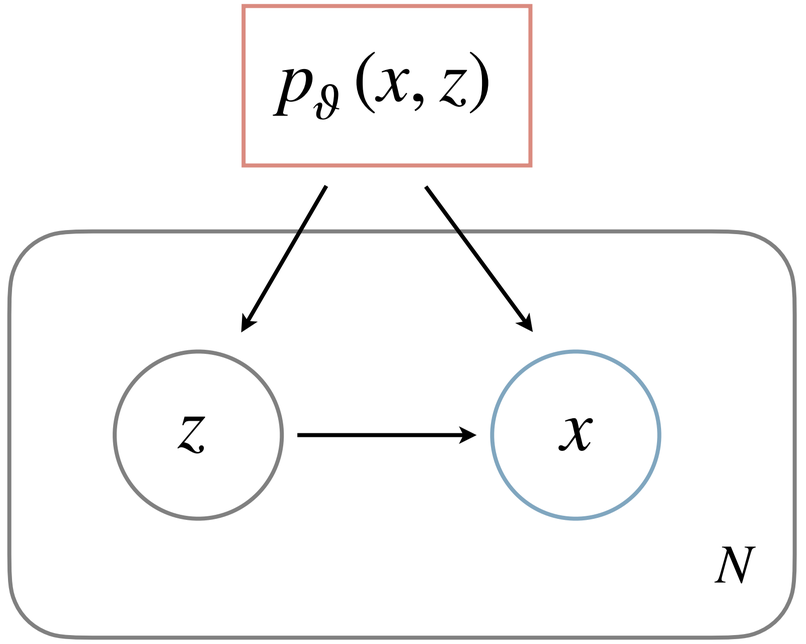
Learn lower-dimensional structure in the data distribution



Observed variables

Make the problem easier by making it "harder":

introduce *joint distribution* $p_\theta(x, z)$

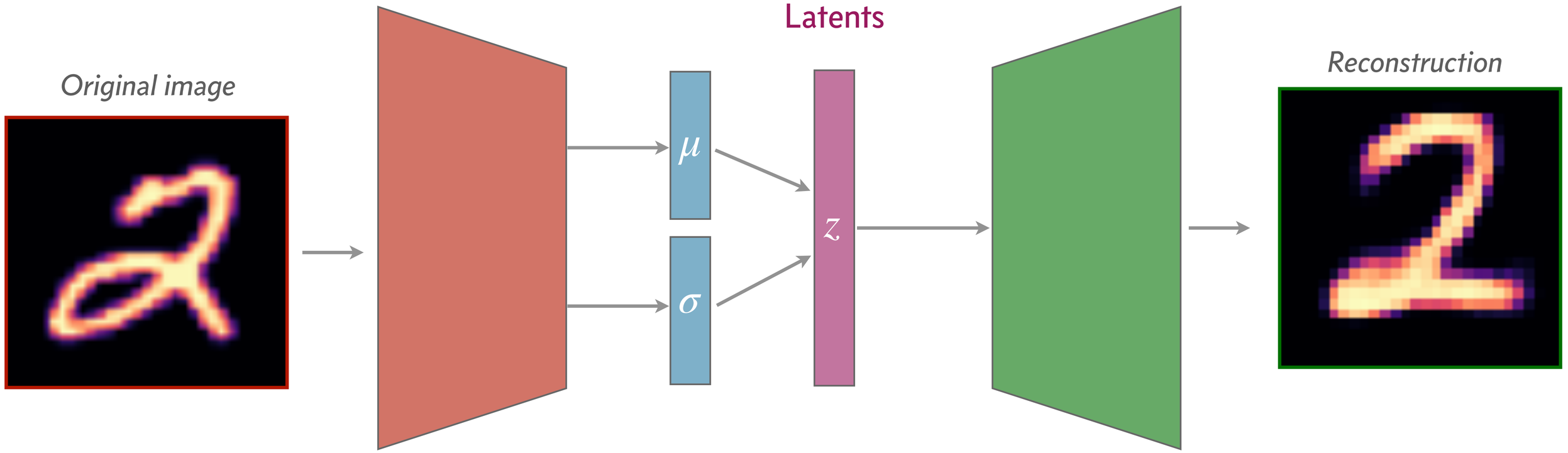


Latent variables *Observed variables*

Common factorization:

$$p_\vartheta(x, z) = p(z) \cdot p_\vartheta\left(x \mid z\right)$$

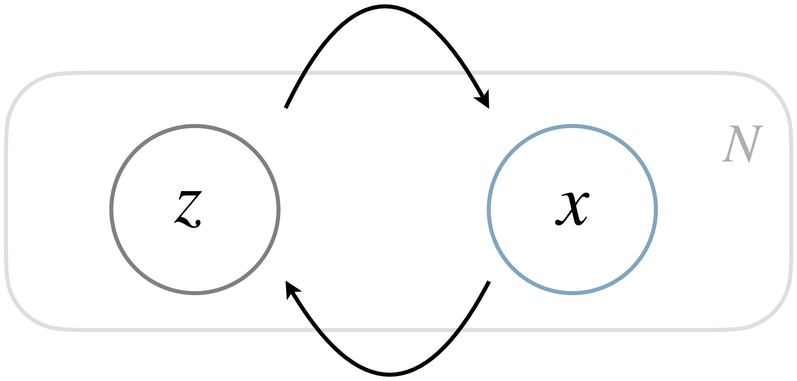
A bird-eye view



A Bayesian latent-variable model optimized with variational inference

Reverse process

$$p_\vartheta\left(x \mid z\right) \cdot p(z)$$



$$q_\varphi(z \mid x) \cdot p(x)$$

Forward process

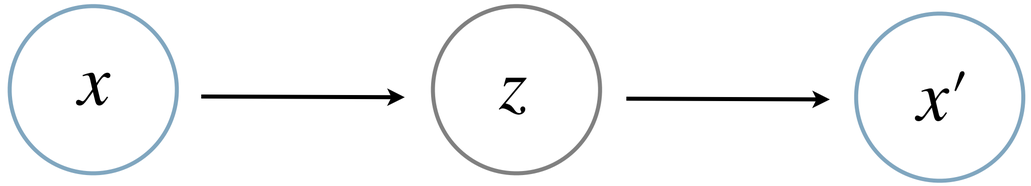
Maximizing ELBO

≡ Minimizing *reverse* KL

≡ "Aligning the forward and reverse processes"

Minimize $\left\langle \log \frac{q(x,z)}{p(x,z)} \right\rangle$

Forward process Reverse process

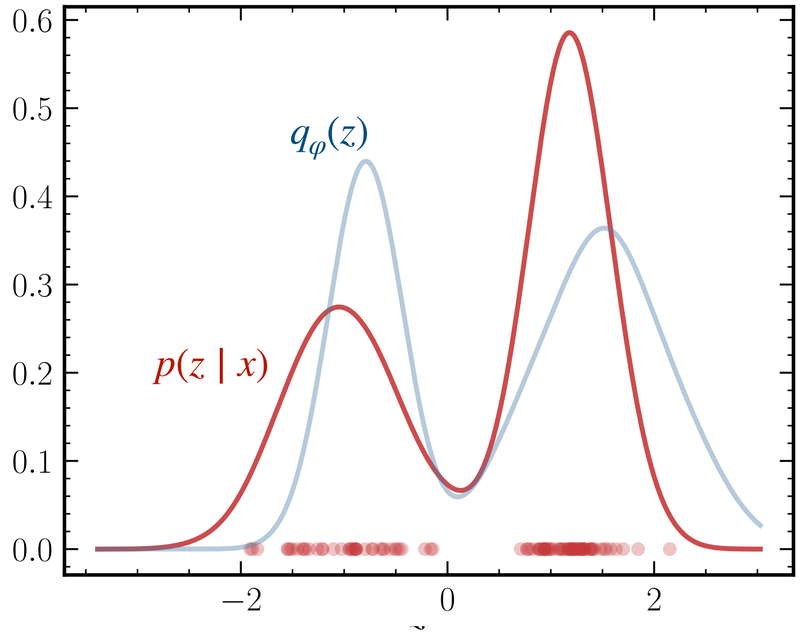


Latents

Variational inference

A general-purpose technique for posterior estimation

$$\underbrace{D_{\mathrm{KL}}\left(q_\varphi(z) \| p(z \mid x)\right)}_{\geq 0} = \underbrace{\log p(x)}_{\text{Evidence}} - \underbrace{\left\langle \log p_\vartheta(x, z) - \log q_\varphi(z) \right\rangle_{q_\varphi(z)}}_{\text{Evidence Lower BOund (ELBO)}}$$



$$\begin{aligned}
\mathrm{ELBO} &= \left\langle \log p_\vartheta(x, z) - \log q_\varphi(z \mid x) \right\rangle_{q_\varphi} \\
&= \left\langle \log p_\vartheta(x \mid z) + \log p(z) - \log q_\varphi(z \mid x) \right\rangle_{q_\varphi} \\
&= \underbrace{\left\langle \log p_\vartheta(x \mid z) \right\rangle_{q_\varphi}}_{\text{“Reconstruction”}} - \underbrace{D_{\mathrm{KL}}\left(q_\varphi(z \mid x) \,\|\, p(z)\right)}_{\text{“Regularization”}}
\end{aligned}$$

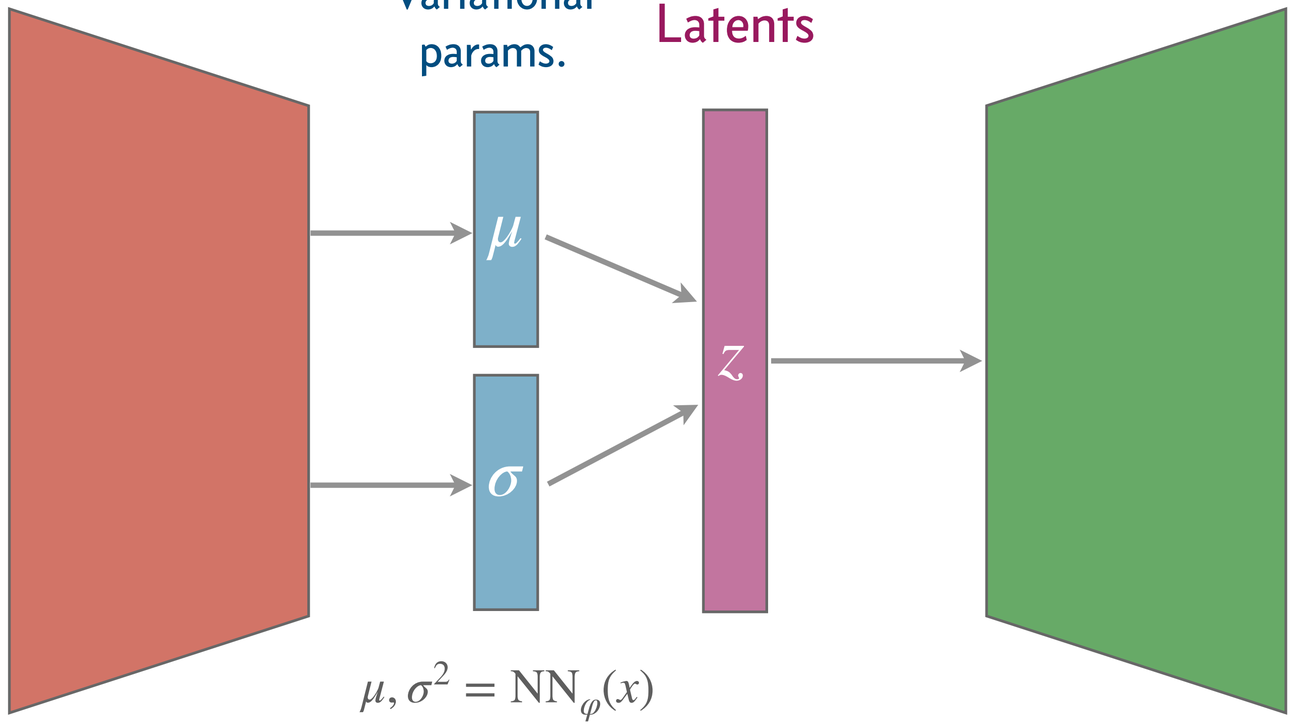
VAEs in practice

$$p(z) = \mathcal{N}(z; 0, I)$$

Variational params.

Latents

Original image



Reconstruction

$$\mu, \sigma^2 = \mathrm{NN}_\varphi(x)$$

$$x \sim p(x)$$

$$z \sim q_\varphi(z \mid x)$$

$$x' \sim p_\vartheta(x \mid z)$$

(Probabilistic) Encoder

$$q_\varphi(z \mid x) = \mathcal{N}(z; \mu, \sigma^2 \mathbb{I})$$

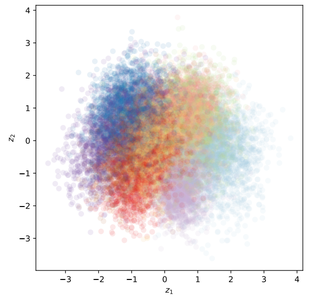
(Probabilistic) Decoder

Noise model / data likelihood

VAEs in practice

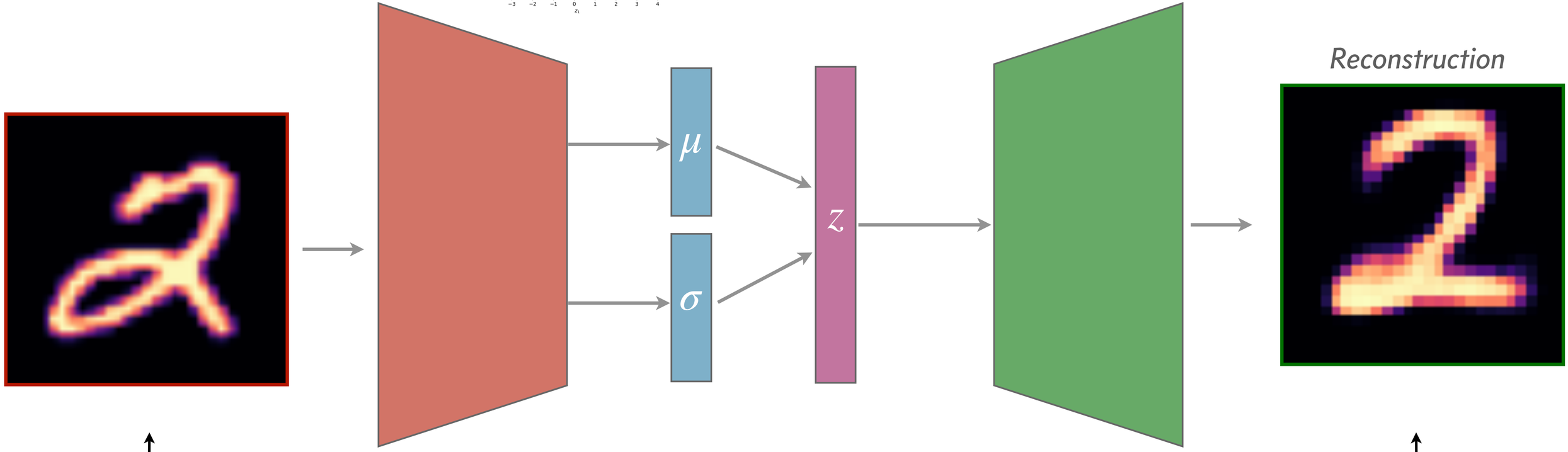
$$\text{ELBO} = \left\langle \log p_\vartheta(x \mid z) \right\rangle_{q_\varphi} - D_{\text{KL}}\left(q_\varphi(z \mid x) \,\|\, p(z)\right)$$

$q_\varphi(z \mid x)$



$$D_{\text{KL}}\left(q_\varphi(z \mid x) \,\|\, p(z)\right)$$

Regularization



$$\left\langle \log p_\vartheta(x \mid z) \right\rangle_{q_\varphi}$$

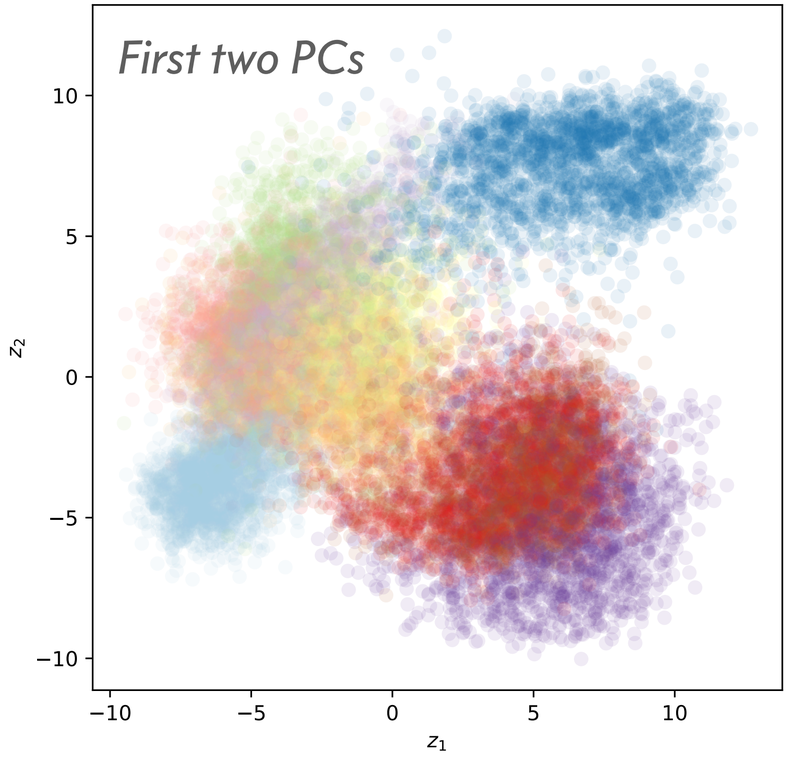
Reconstruction (e.g., MSE, …)

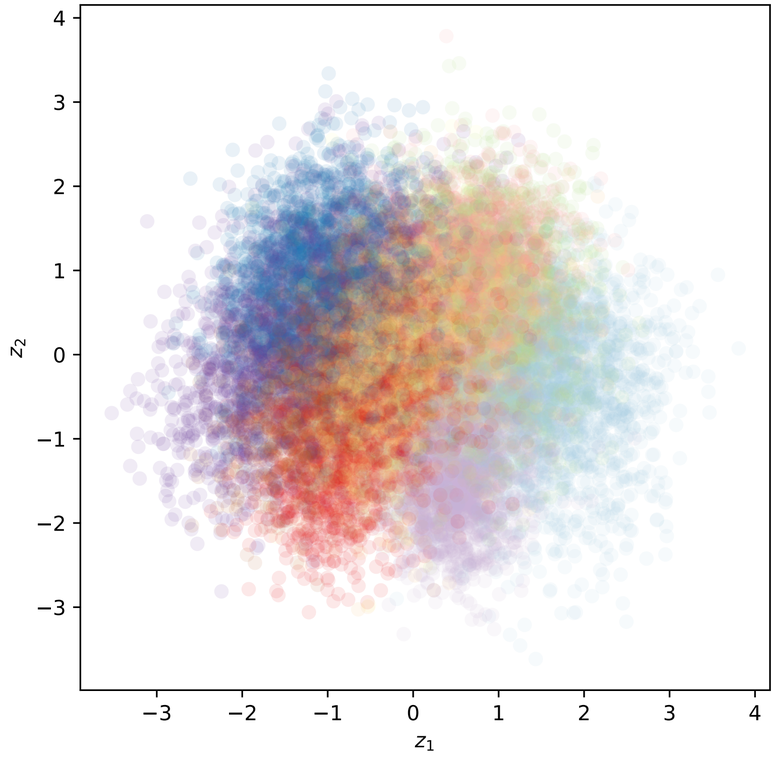
$$\|x - x'\|_2^2$$

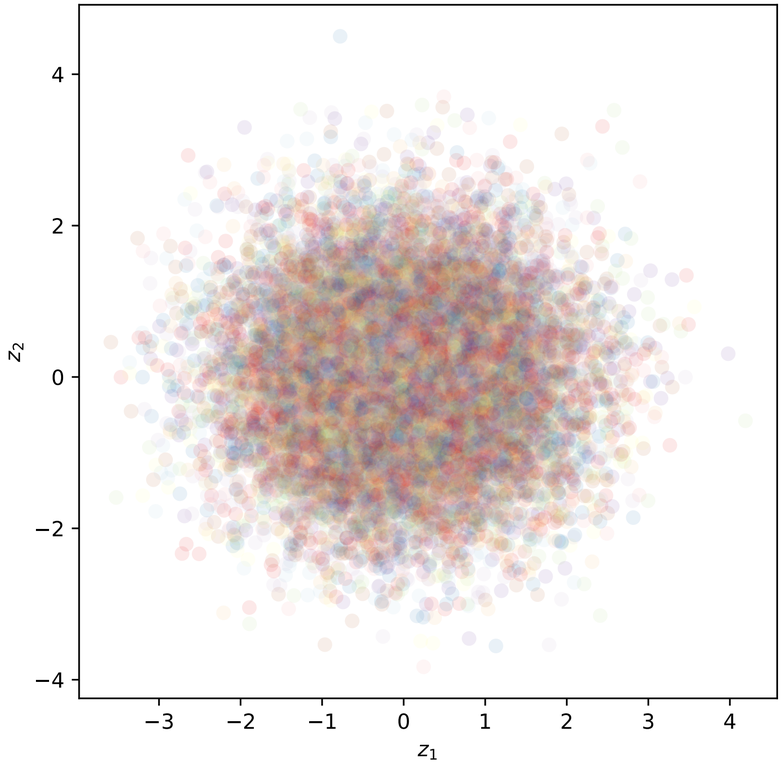
A semantically meaningful latent space

The KL-term enforces simplicity in the latent space, encouraging learned semantic structure and *disentanglement*

Pure reconstruction ⟷ *More latent regularization*







Neural compression: *Rate-distortion theory*

Autoencoding is a form of (neural) compression!

$$-\text{ELBO} = -\left\langle \log p_\vartheta(x \mid z) \right\rangle_{q_\varphi} + D_{\text{KL}}\left(q_\varphi(z \mid x) \,\|\, p(z) \right)$$

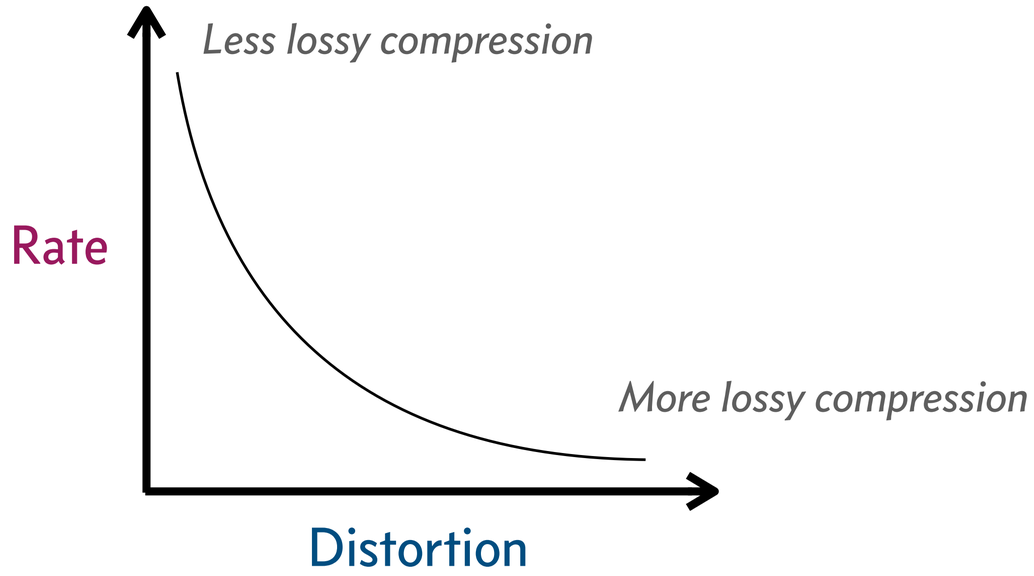
Distortion

"Reconstruction loss"

Rate

"Amount of compression"

Rate-distortion curve quantifies this tradeoff



Controlling compression and disentanglement: β-VAEs

$$-\text{ELBO} = -\left\langle \log p_\vartheta(x \mid z) \right\rangle_{q_\varphi} + \beta \cdot D_{\text{KL}}\left(q_\varphi(z \mid x) \,\|\, p(z) \right)$$

Distortion

Rate

- Larger σ: More of the data variation is attributed to the likelihood → larger "β", more compression
- Smaller σ: Latents z try to capture more of the variation in the data (e.g. small perceptual features)