7. Előadás: Diffúziós Modellek

Generatív AI és Inverz Módszerek a Képszintézisben

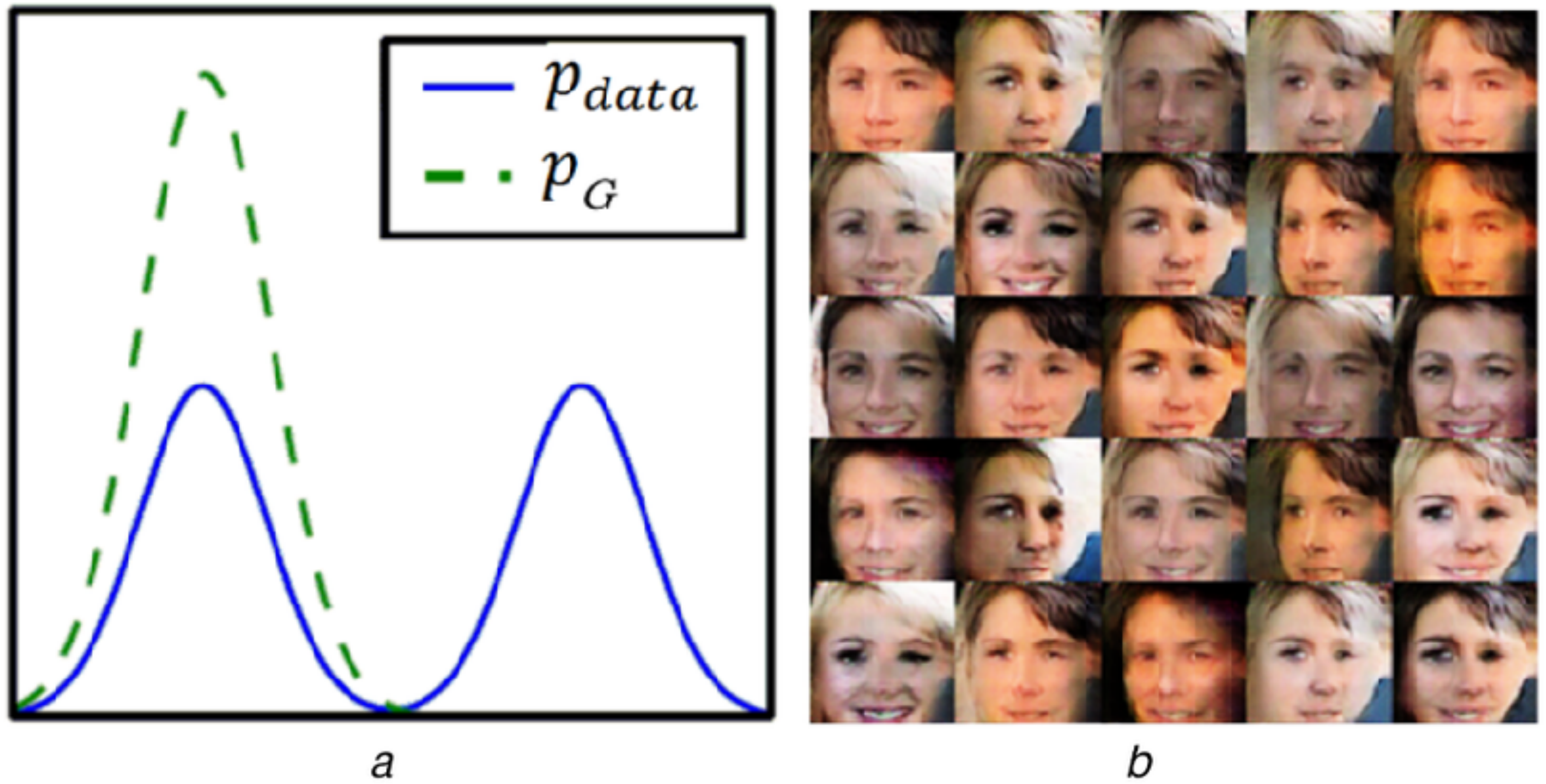
BME-VIK IIT, 2026

Dr. Vaitkus Márton

Generatív Modellezés

Korábbi módszerek

Autóenkóder:
Változatos, de homályos/“átlagos” képek



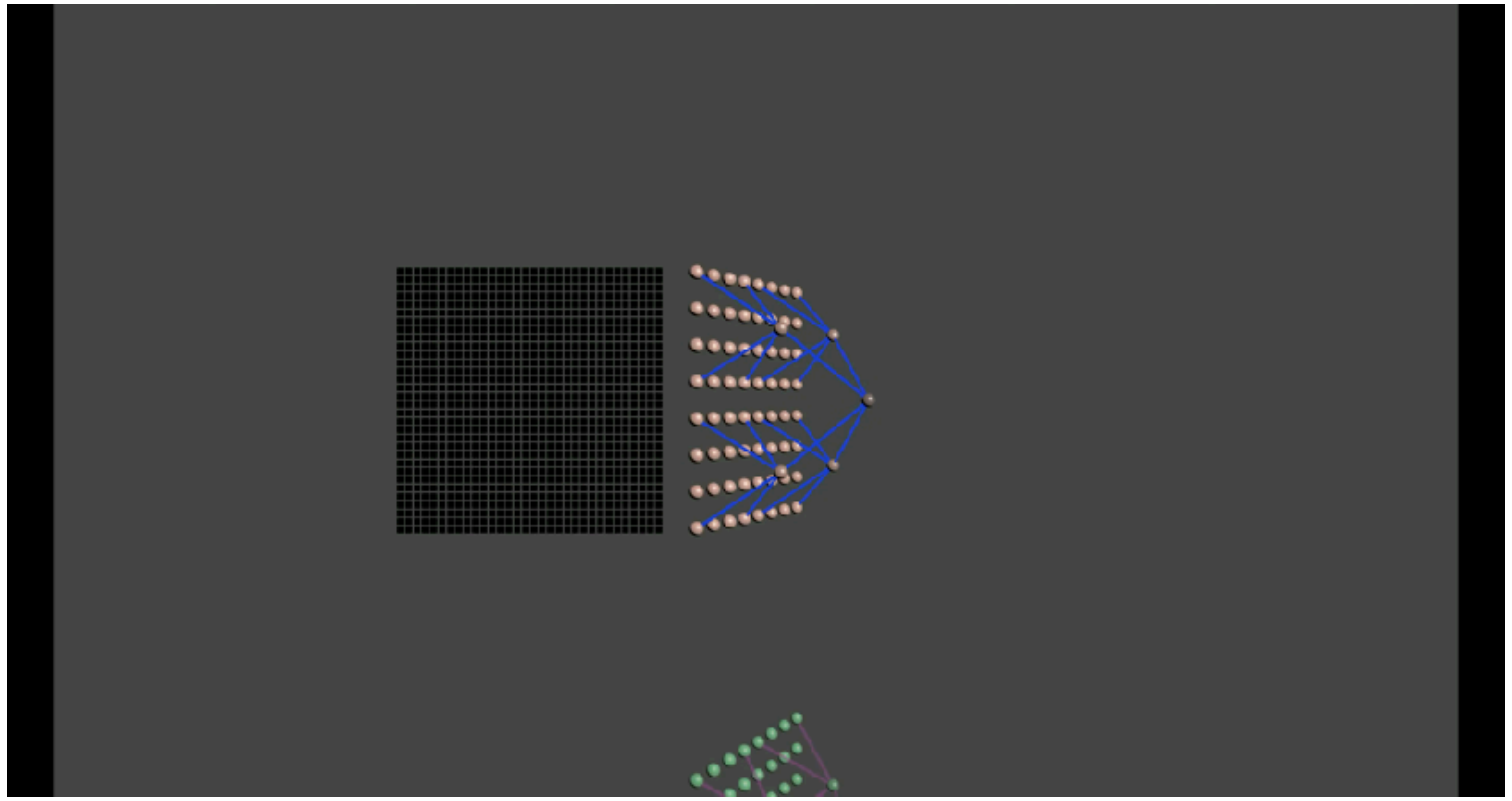
GAN:
Részletes, de egyhangú képek (“mode collapse”)
Nehéz tanítani

Nem várunk túl sokat egy neurális hálótól?

Diffúziós Modellek

Mi lenne, ha a képet nem egy lépésben generálnánk, hanem fokozatosan?

Diffúzió

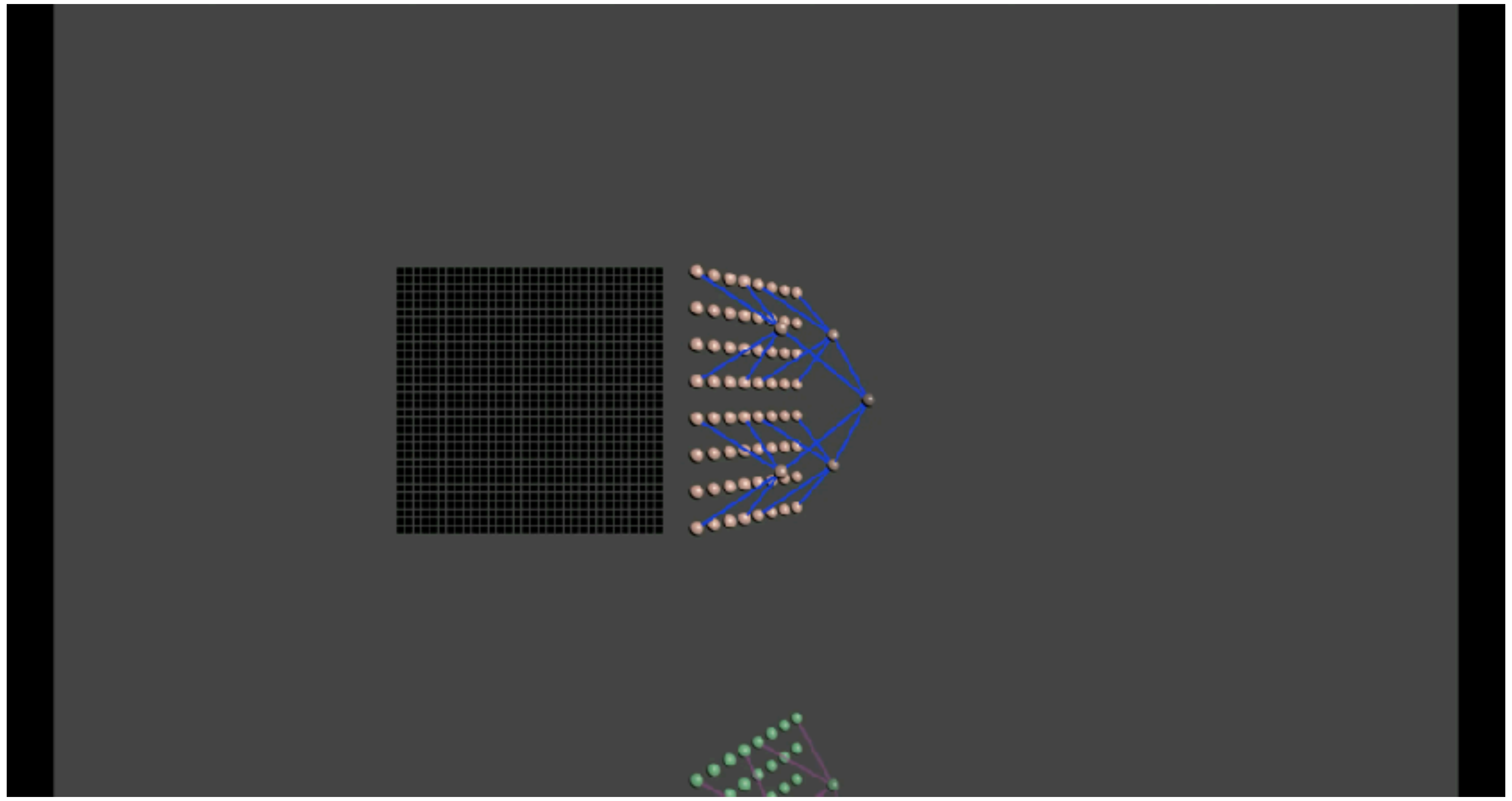


Autoregresszió

Diffúziós Modellek

Mi lenne, ha a képet nem egy lépésben generálnánk, hanem fokozatosan?

Diffúzió



Autoregresszió

Diffúziós Modellek

Denoizer hálók

- Tekintsünk egy $x \sim p(x)$ képet, adjunk hozzá Gaussi zajt σ_t szórással:

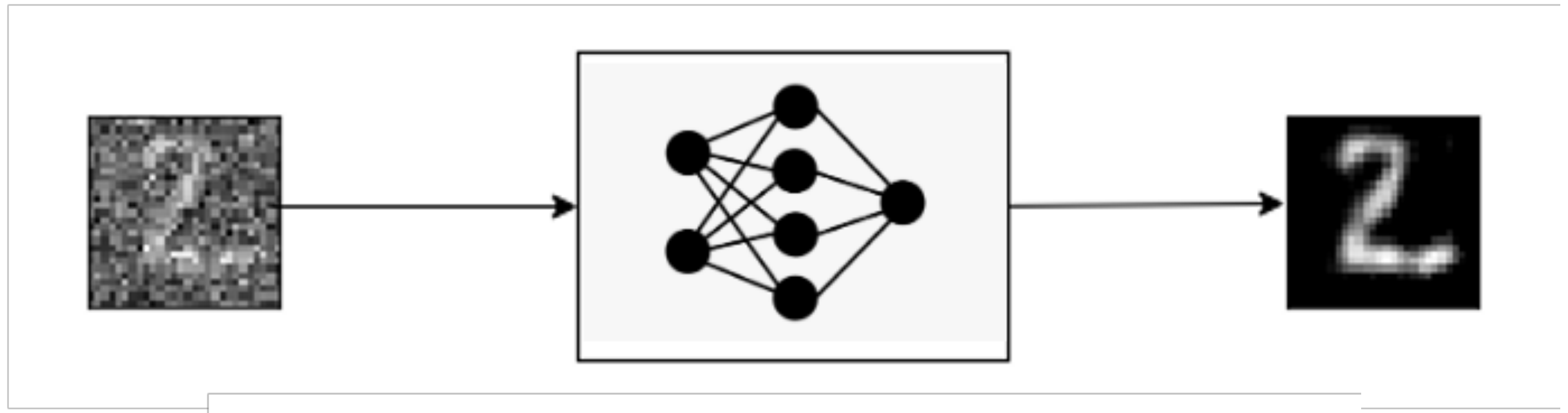
$$y = x + \epsilon, \quad \epsilon \sim p_\sigma(y \mid x) = \mathcal{N}(0,\sigma)$$

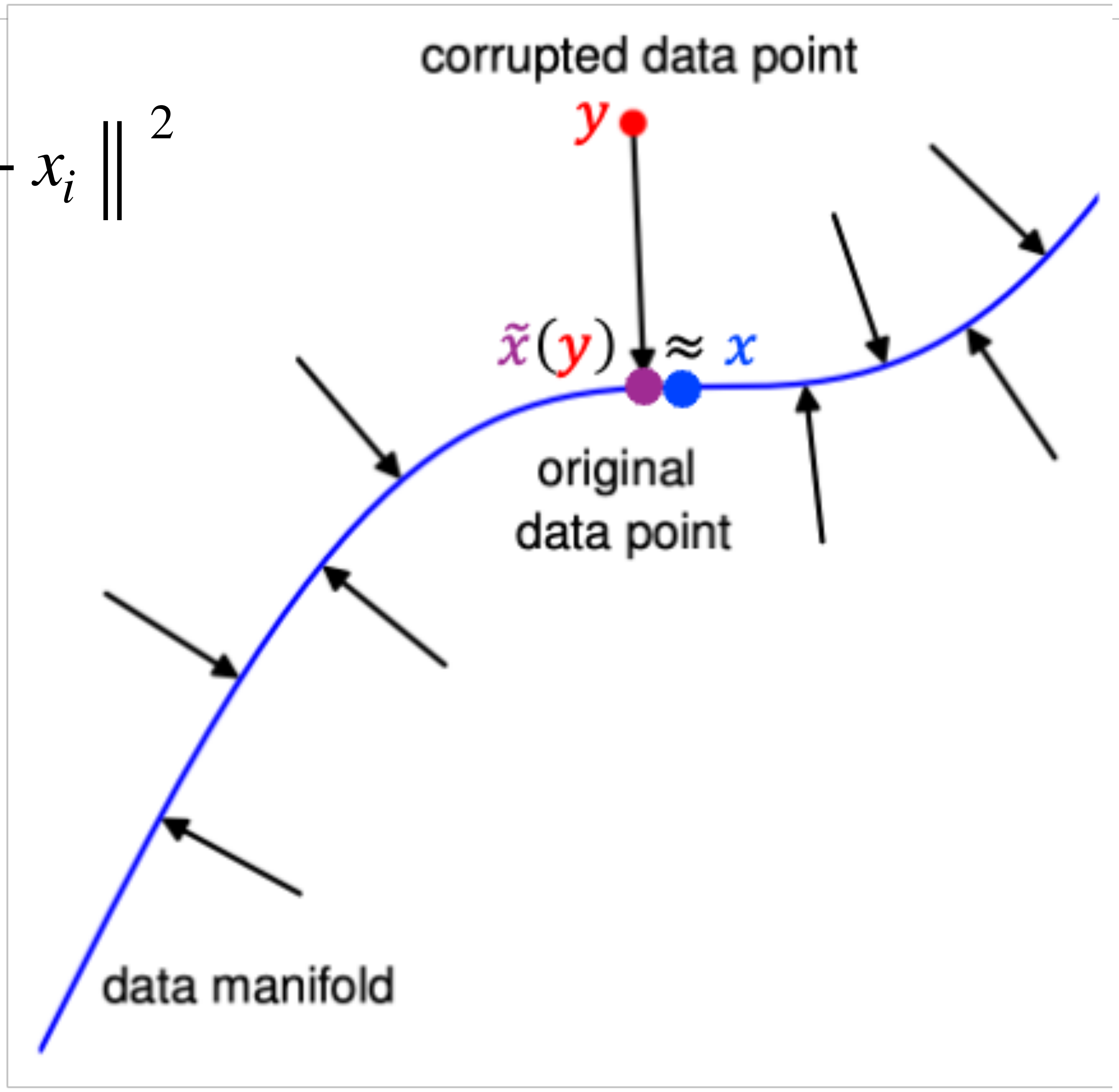
- Tanítsunk hálót a tiszta kép MSE rekonstrukciójára: $\min_{\tilde{x}} \sum_i \left\| \tilde{x}(y_i) - x_i \right\|^2$

- MSE szerint legjobb becslés a tiszta képre (**MMSE**):

$$\tilde{x}(y) = \int x \cdot \underbrace{p(x \mid y)}_{\frac{p(y \mid x) \cdot p(x)}{p(y)}} dx = \frac{1}{p(y)} \int x \cdot p(y \mid x) p(x) dx$$

- **Formulával**: $\tilde{x}(y) = \sum_i \frac{e^{\frac{(x_i - y)^2}{2\sigma^2}} \cdot x_i}{\sum_j e^{\frac{(x_j - y)^2}{2\sigma^2}}}$ Sok adatpontra: A hálóval csak közelítjük…





Diffúziós Modellek

Denoizer hálók

Zajos adatpontok eloszlásfüggvénye és gradiense:

"Gaussian blur" (az eloszlásra!!)

$$p_\sigma(y) = \int p_\sigma(y \mid x)p(x)dx \propto \int e^{-\frac{-\|y-x\|^2}{2\sigma^2}} p(x)dx$$

$$\nabla p_\sigma(y) = \frac{1}{\sigma^2}\int (x-y)p_\sigma(y \mid x)p(x)$$

<u>DEMÓ!</u>

- Tekintsünk egy zajos és eredeti pontok közötti *különbségvektort* (becsült zajt):

$$\tilde{x}(y,\sigma) - y = \int x \cdot p_\sigma(x \mid y)dx - y \cdot \underbrace{\int p_\sigma(x \mid y)dx}_{1} = \int (x-y) \cdot p_\sigma(x \mid y)dx$$

$$= \frac{1}{p_\sigma(y)}\int (x-y)p_\sigma(y \mid x)p(x)dx = \sigma^2 \frac{\nabla p_\sigma(y)}{p_\sigma(y)} = \sigma^2 \boxed{\nabla \log p_\sigma(y)}$$

<u>Score függvény!</u>

Tweedie formula

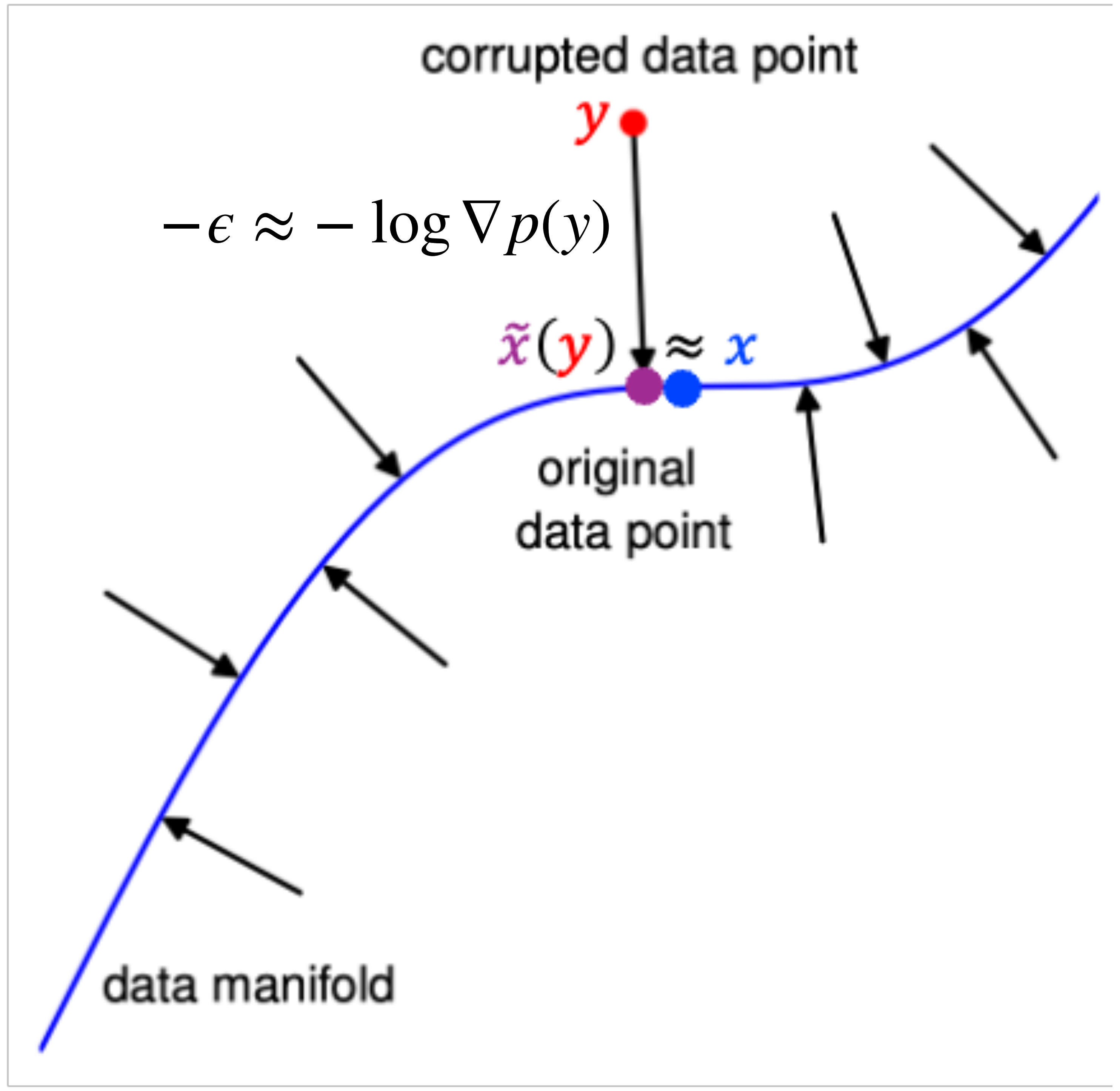
M. Tweedie (1919–1996)

- **Ideális denoizer = zajos eloszlás score-függvénye!**

Diffúziós Modellek

Denoizer hálók

- Több lehetséges paraméterezés a denoizer hálókra
  - Tiszta kép becslése: $\tilde{x}_\theta(y, \sigma) \approx x$
  - Zaj becslése: $\tilde{\epsilon}_\theta(y, \sigma) \approx y - x$ (a leggyakoribb!)
  - Score matching: $\tilde{s}_\theta(y, \sigma) \approx \nabla \log p_\sigma(y)$



Elméletileg ekvivalensek, de mindig ellenőrizzük az implementáció melyiket használja!

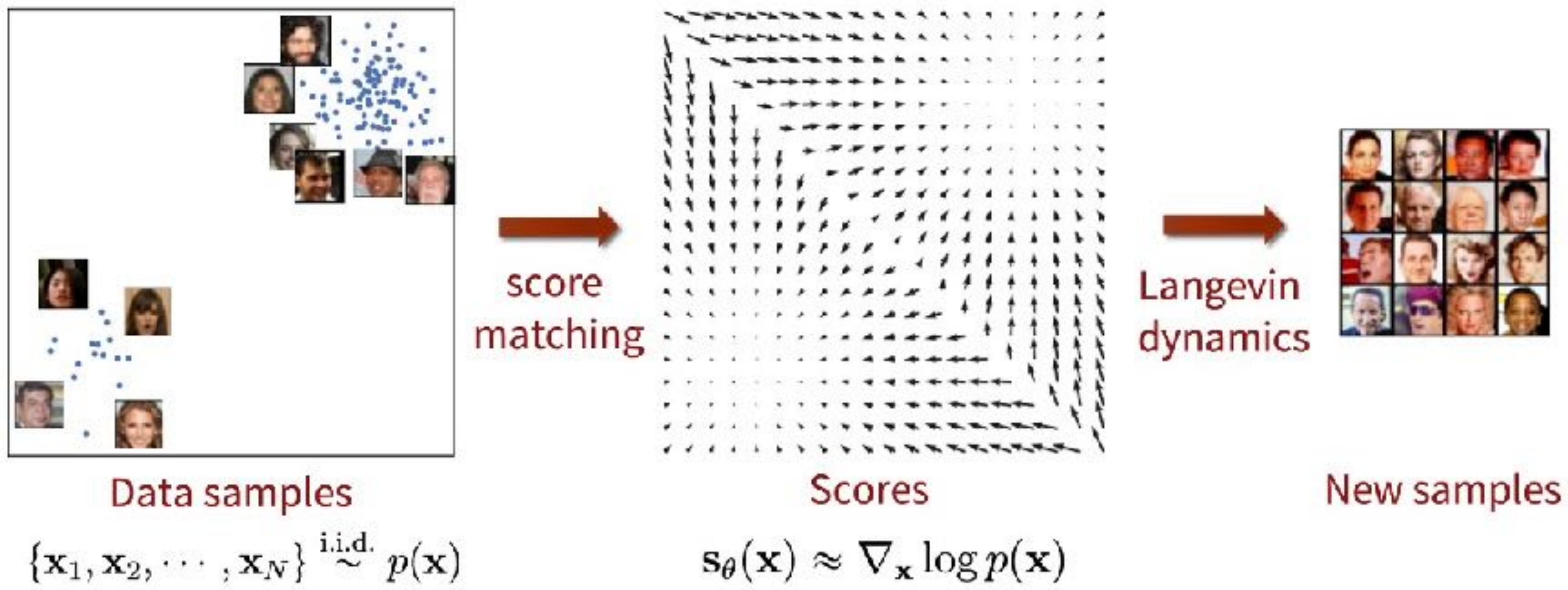
Diffúziós Modellek

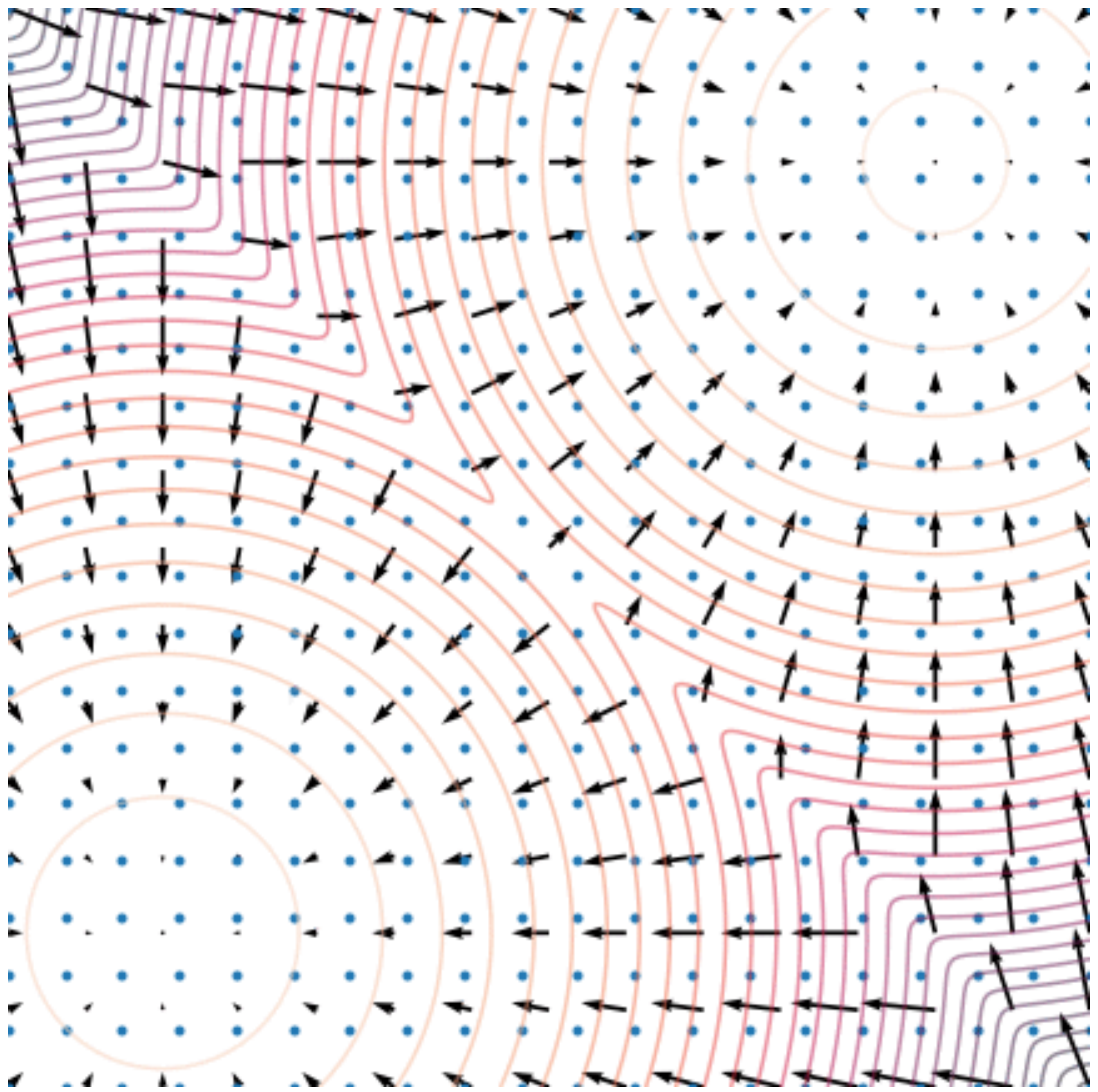
Score matching — Langevin mintavételezés

- Hogyan kapunk denoizer modellből generatív modellt?
- A denoizer becslést ad a score függvényről (**score matching**) — alkalmazzunk **Langevin mintavételezést**!

$$x_{k+1} = x_k + \sigma^2 \cdot \nabla \log p_\sigma(x_k) + \sqrt{2}\sigma \cdot \epsilon_k$$

$$\epsilon_k \in \mathcal{N}(0,\mathbf{I})$$





Diffúziós Modellek

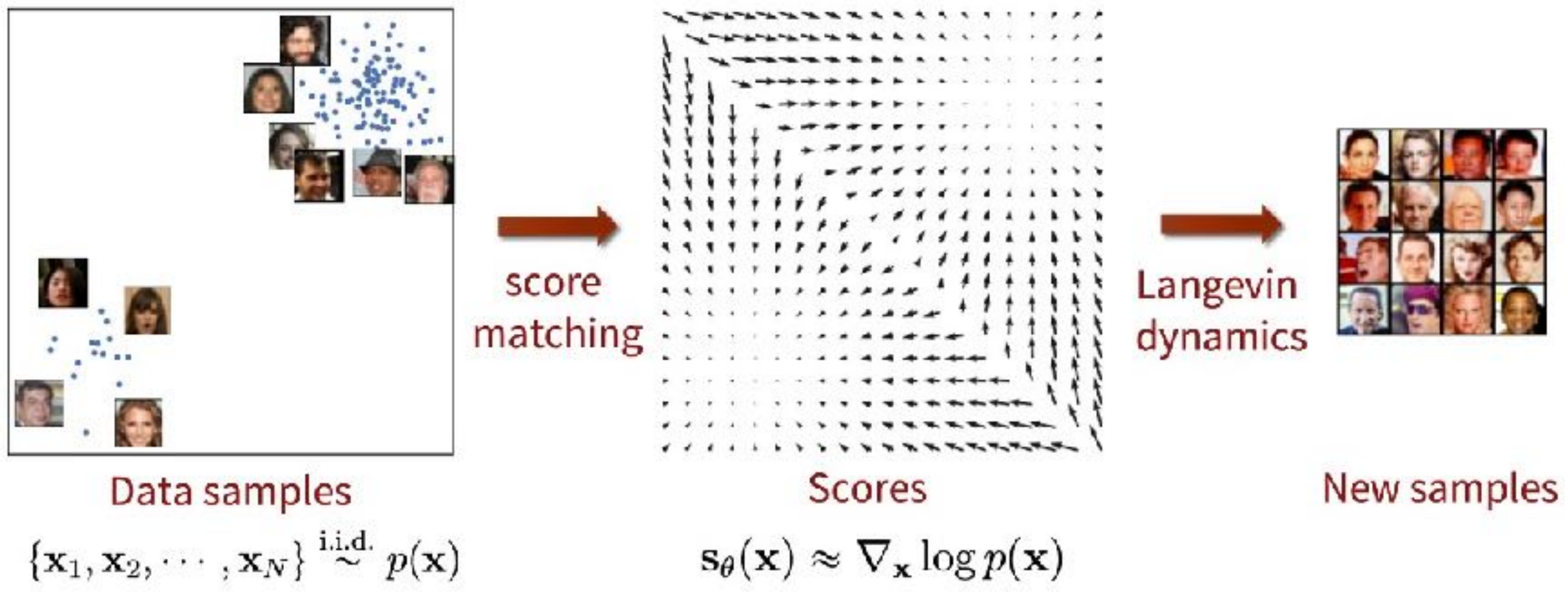
Score matching — Langevin mintavételezés

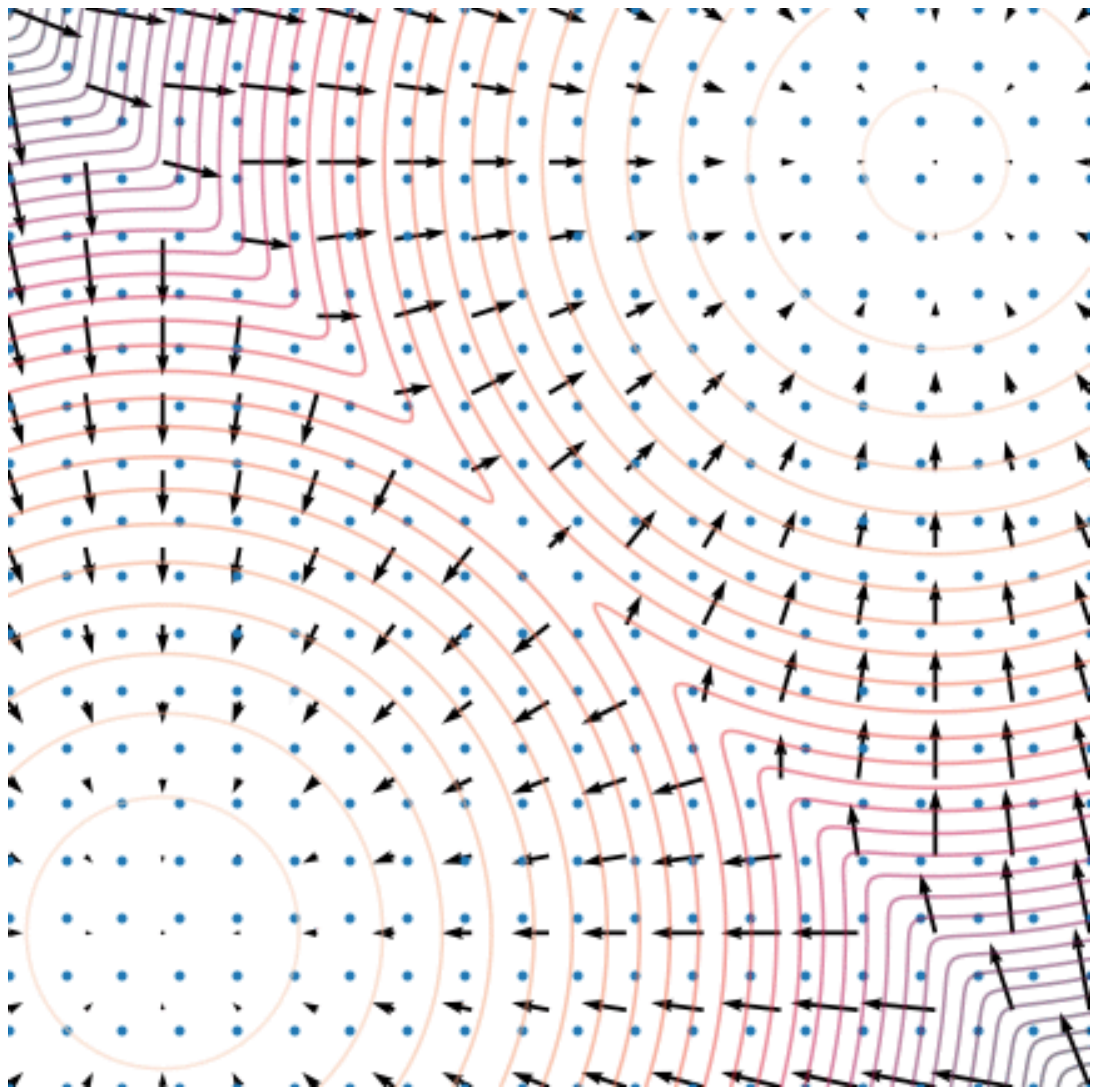
- Hogyan kapunk denoizer modellből generatív modellt?

- A denoizer becslést ad a score függvényről (**score matching**) — alkalmazzunk **Langevin mintavételezést**!

$$x_{k+1} = x_k + \sigma^2 \cdot \nabla \log p_\sigma(x_k) + \sqrt{2}\sigma \cdot \epsilon_k$$

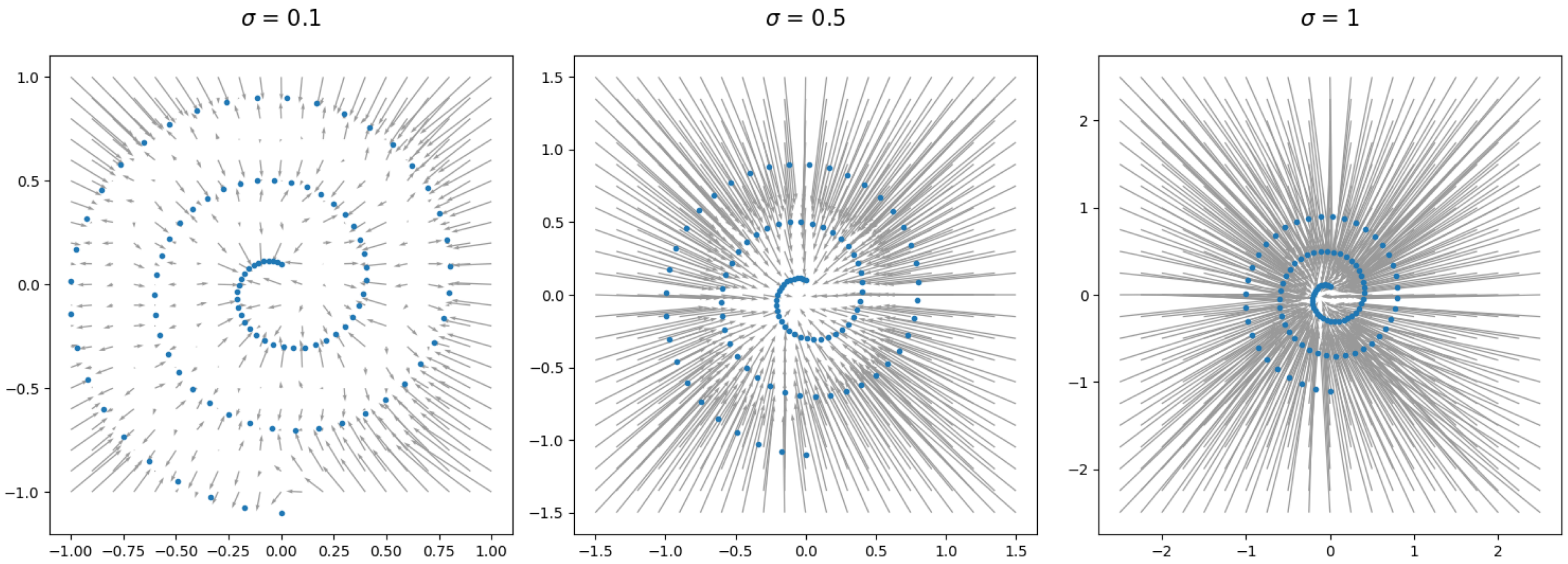
$$\epsilon_k \in \mathcal{N}(0,\mathbf{I})$$





Diffúziós Modellek

Score matching — Zajszint megválasztása

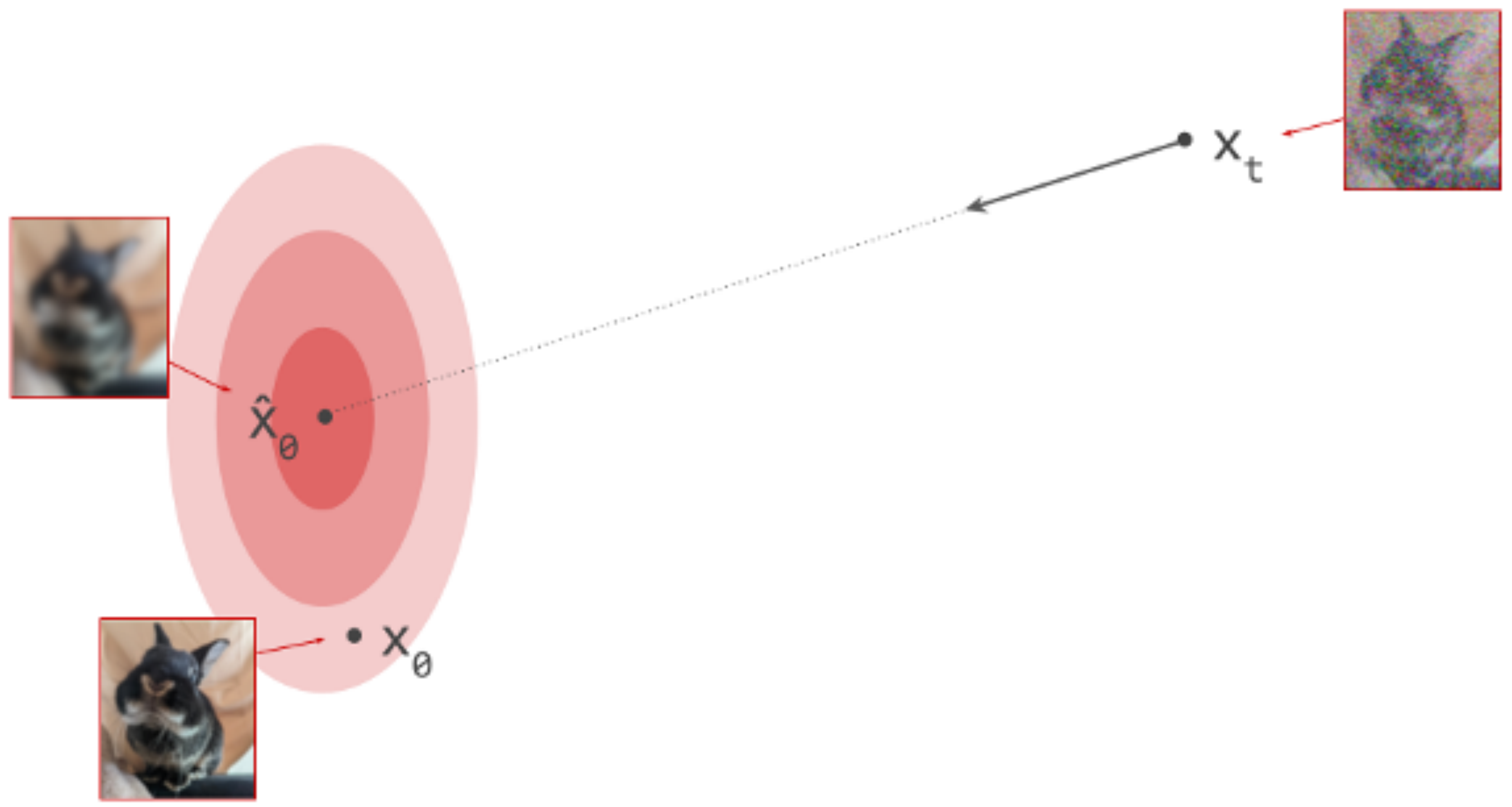


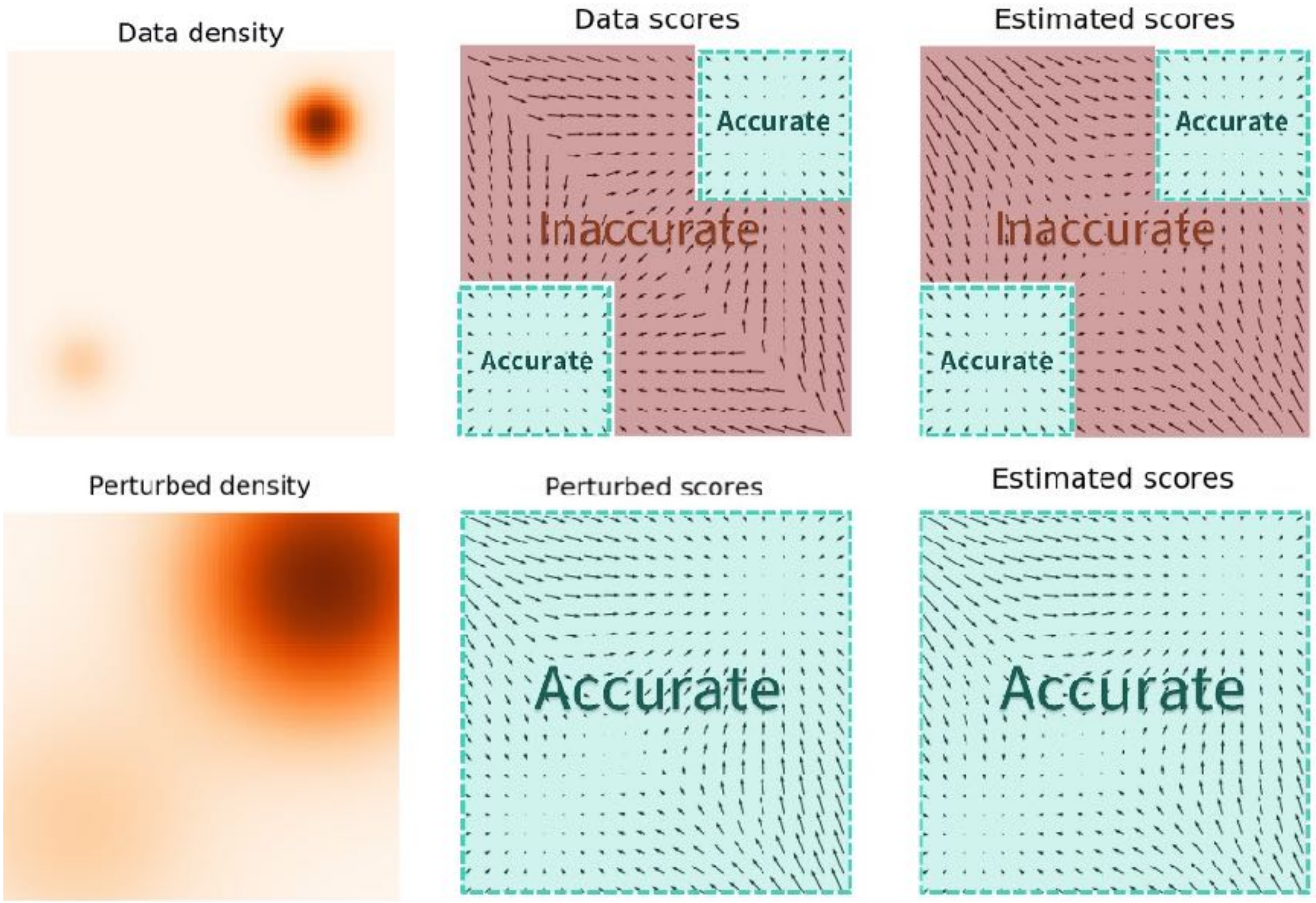
Score függvény különböző zajszintekre — DEMO

Diffúziós Modellek

Score matching — Zajszint megválasztása

- Mennyi zajt adjunk az adatokhoz?
- Túl kevés zaj: pontatlanul becsüljük az eloszlást
- Túl sok zaj: a denoizer homályos/ “átlagos” mintákat generál





Diffúziós Modellek

Score matching

Generative Modeling by Estimating Gradients of the Data Distribution

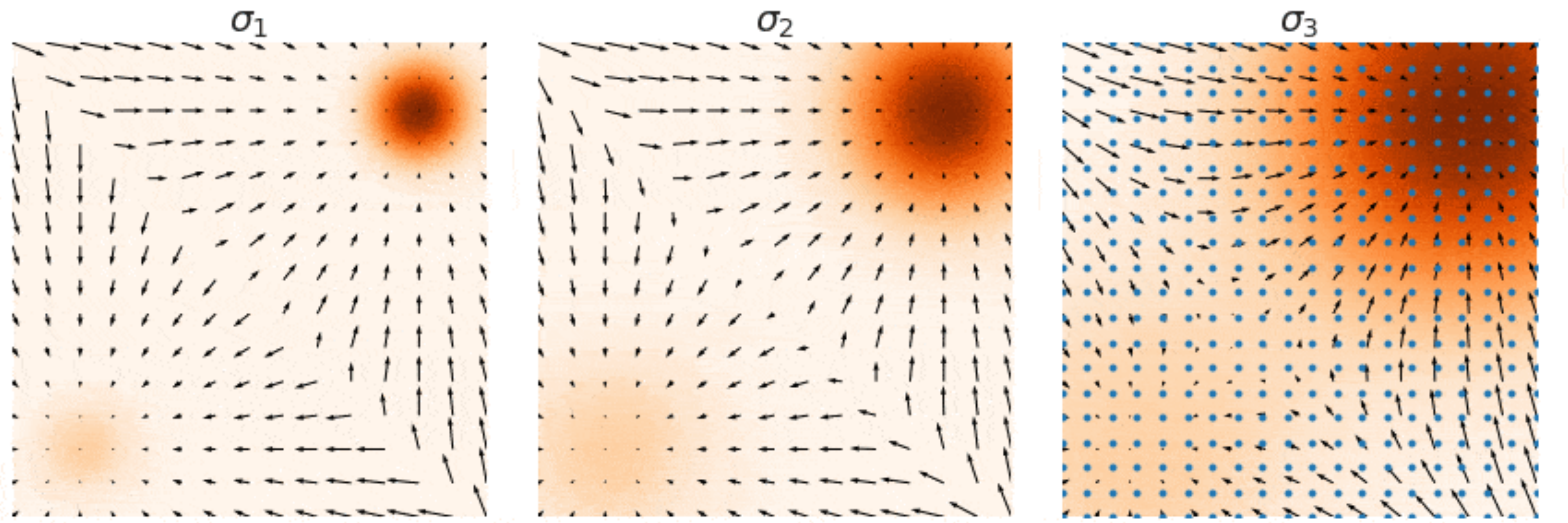
Yang Song
Stanford University

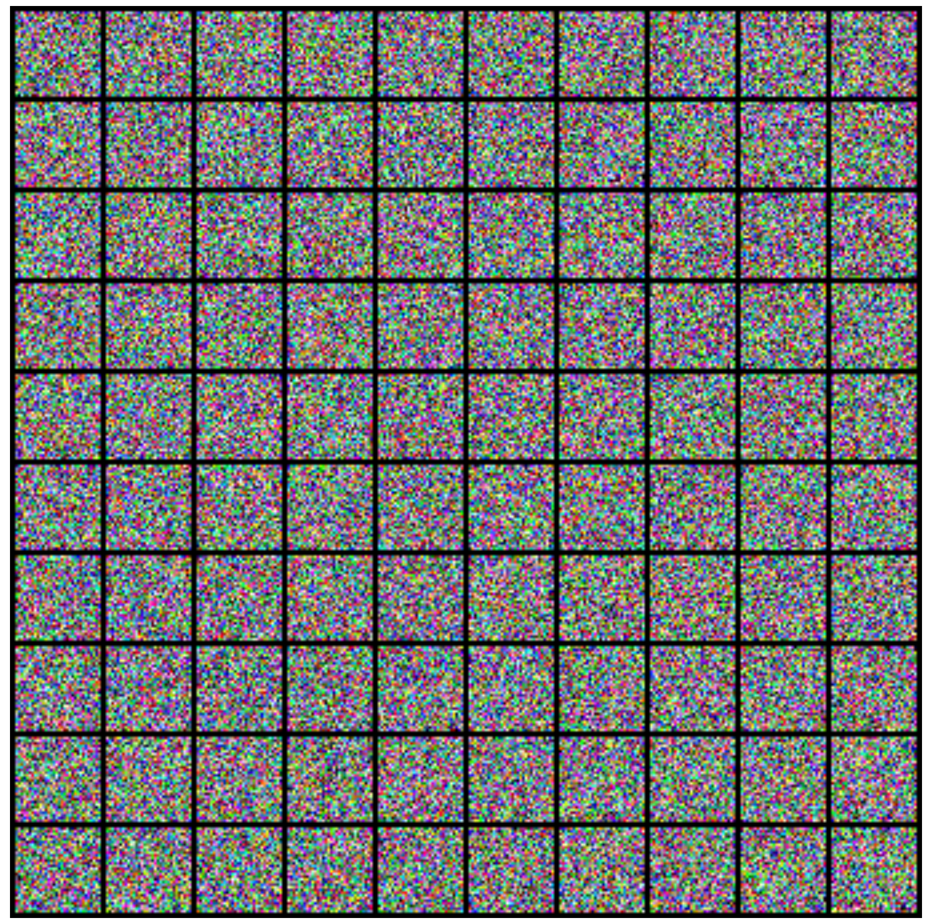
Stefano Ermon
Stanford University

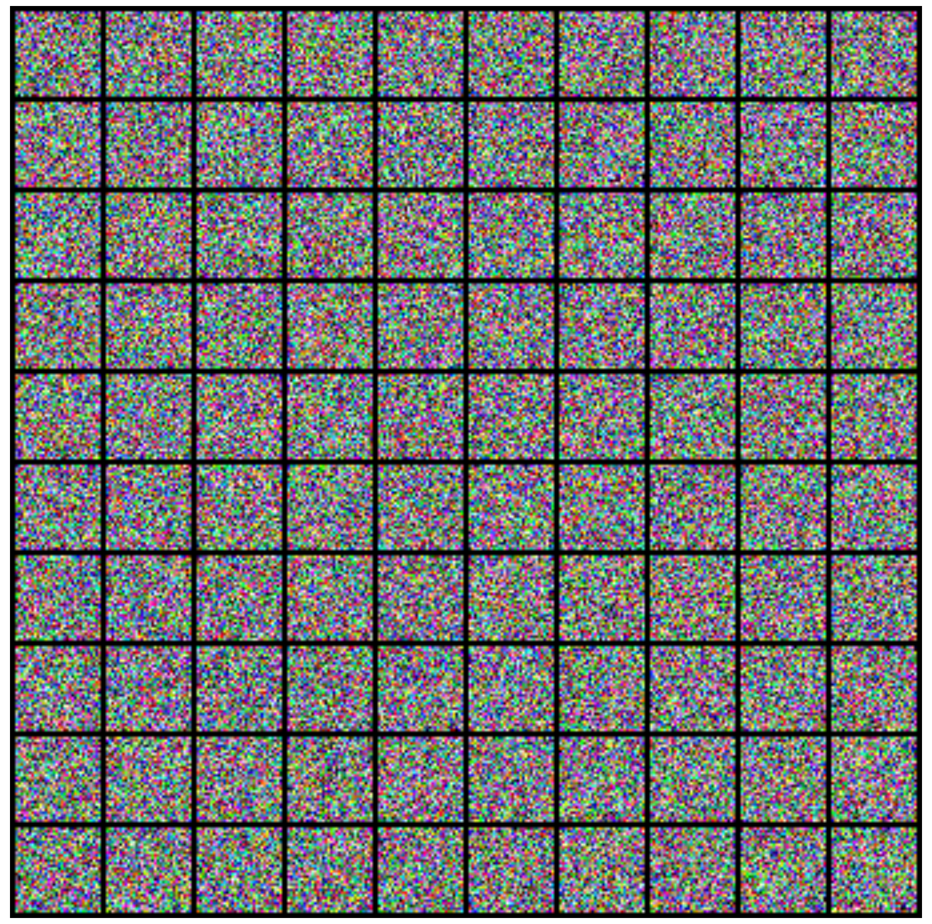
- <u>Ötlet</u>: fokozatosan növeljük a zaj szintjét valamilyen $\sigma(t),\ t \in [0,1]$ ütemezés szerint és fokozatosan távolítsuk el!
- Tanítsunk egy <u>közös</u> denoizer hálót a különböző zajszintekre!

$$\min_{\theta} \mathbb{E}_{\sigma}\mathbb{E}_{x,\epsilon\sim\mathcal{N}(0,I)}\left[\ \left\| \epsilon_\theta(x+\sigma\cdot\epsilon,\sigma)-\epsilon \right\|^2\right]$$

- Mintavételezés: <u>**“lágyított” Langevin**</u> módszerrel
    - Induláskor $\sigma(t)\approx 1$, aztán minden lépésben fokozatosan kisebb zajszint szerint lépünk
- <u>**Diffúziós modellek**</u> egy példája







Diffúziós Modellek

Score matching

- <u>Ötlet</u>: fokozatosan növeljük a zaj szintjét valamilyen $\sigma(t),\ t \in [0,1]$ ütemezés szerint és fokozatosan távolítsuk el!

- Tanítsunk egy <u>közös</u> denoizer hálót a különböző zajszintekre!

$$\min_{\theta} \mathbb{E}_{\sigma} \mathbb{E}_{x,\epsilon \sim \mathcal{N}(0,I)} \left[\| \epsilon_{\theta}(x + \sigma \cdot \epsilon, \sigma) - \epsilon \|^2 \right]$$

- Mintavételezés: **<u>“lágyított” Langevin</u>** módszerrel

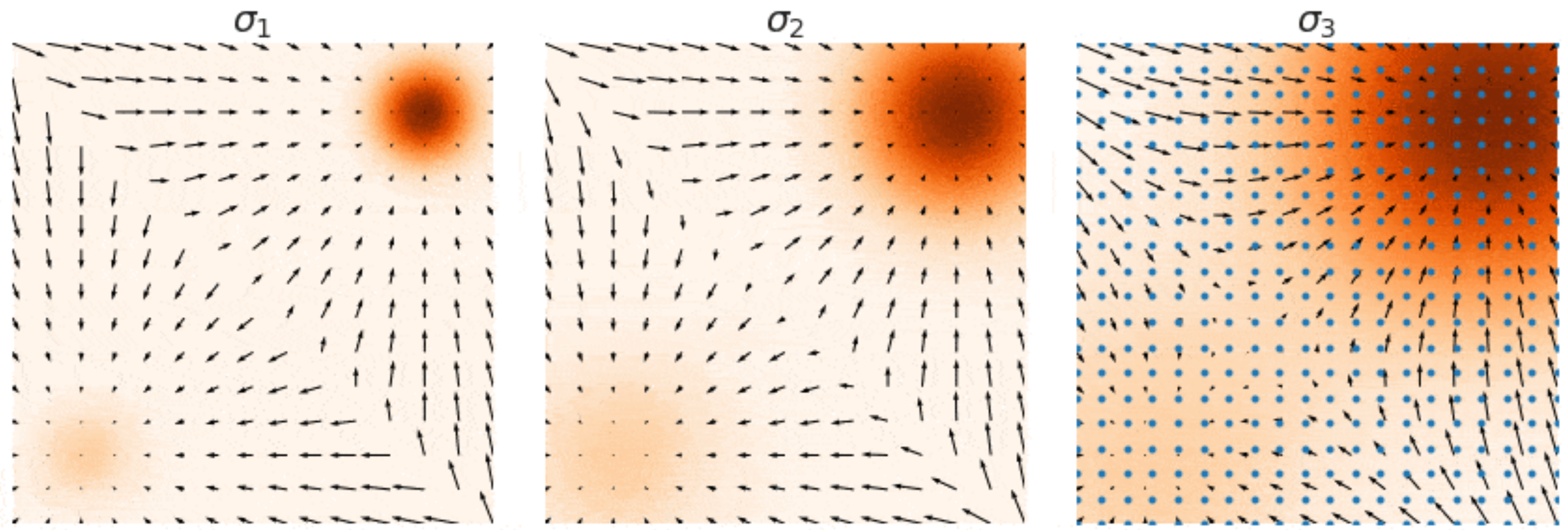
  - Induláskor $\sigma(t) \approx 1$, aztán minden lépésben fokozatosan kisebb zajszint szerint lépünk

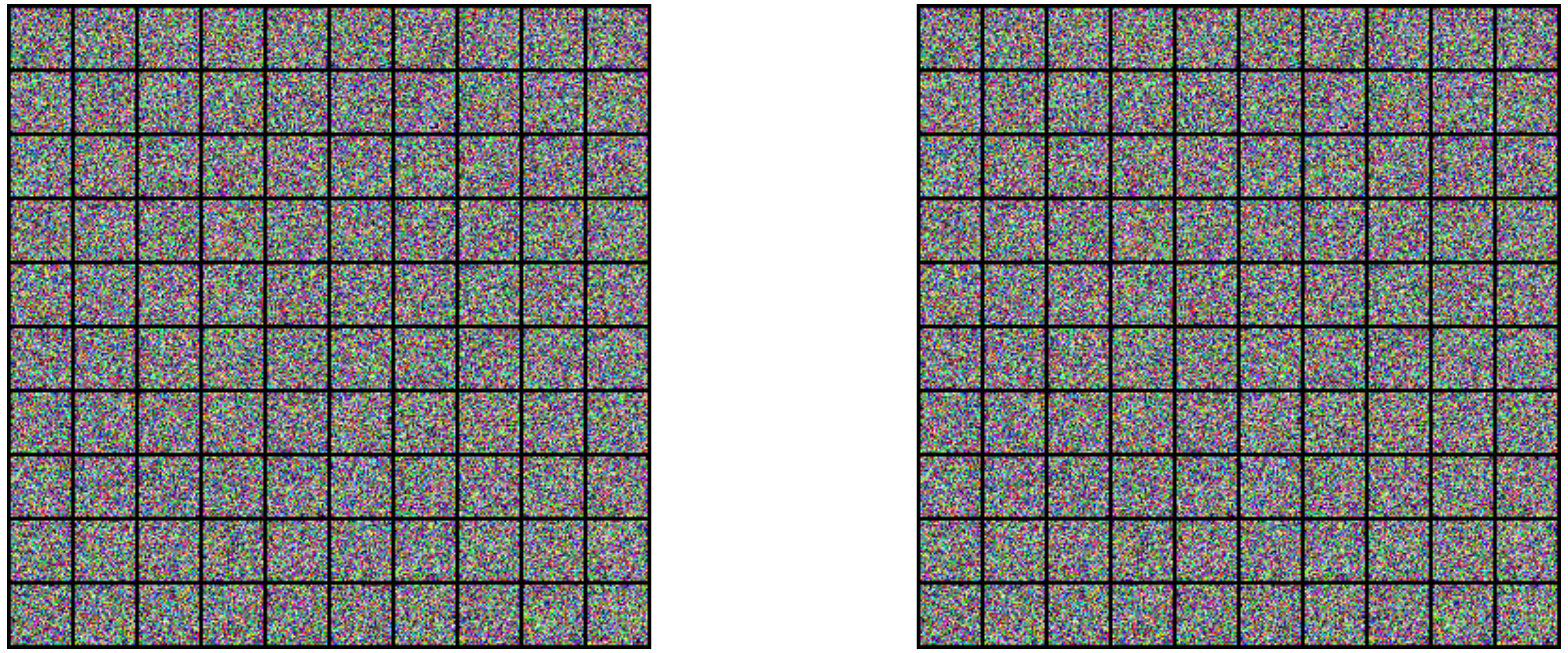
- **<u>Diffúziós modellek</u>** egy példája

Generative Modeling by Estimating Gradients of the Data Distribution

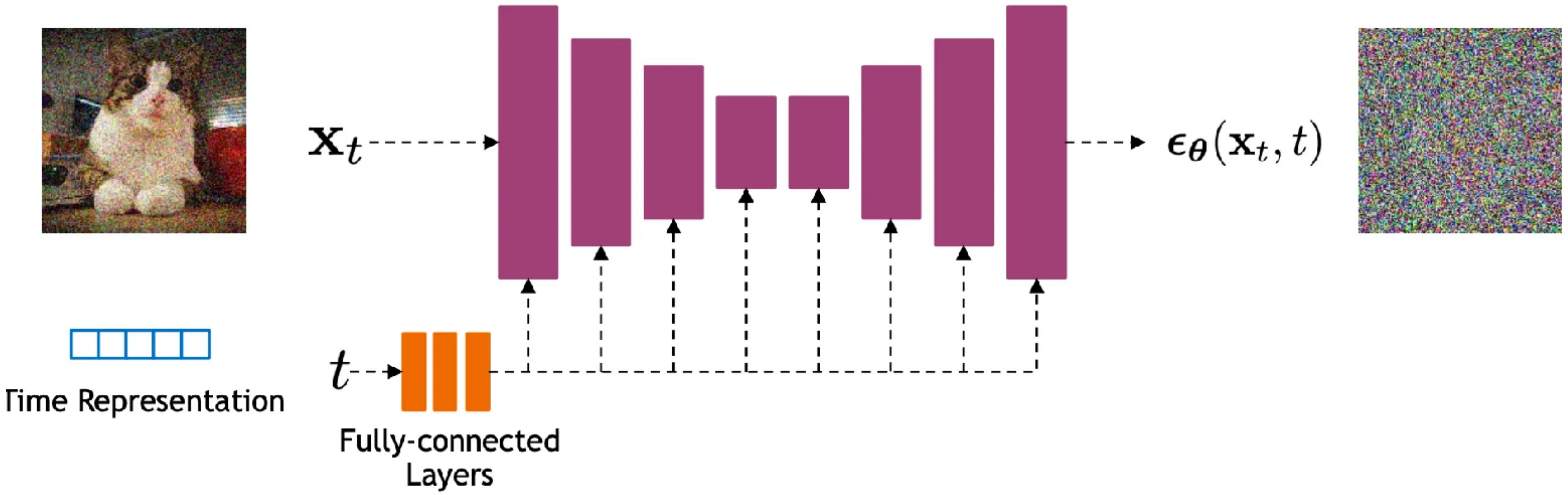
Yang Song
Stanford University

Stefano Ermon
Stanford University



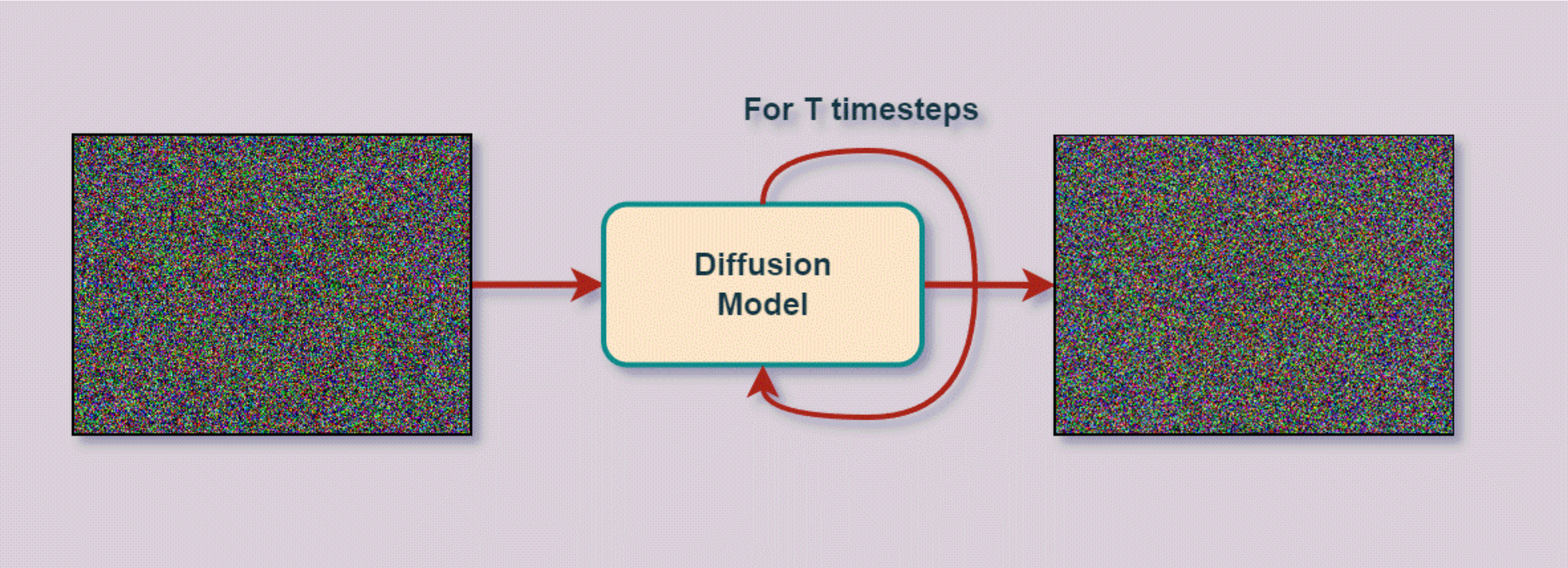


Diffúziós Modellek
Neurális architektúra



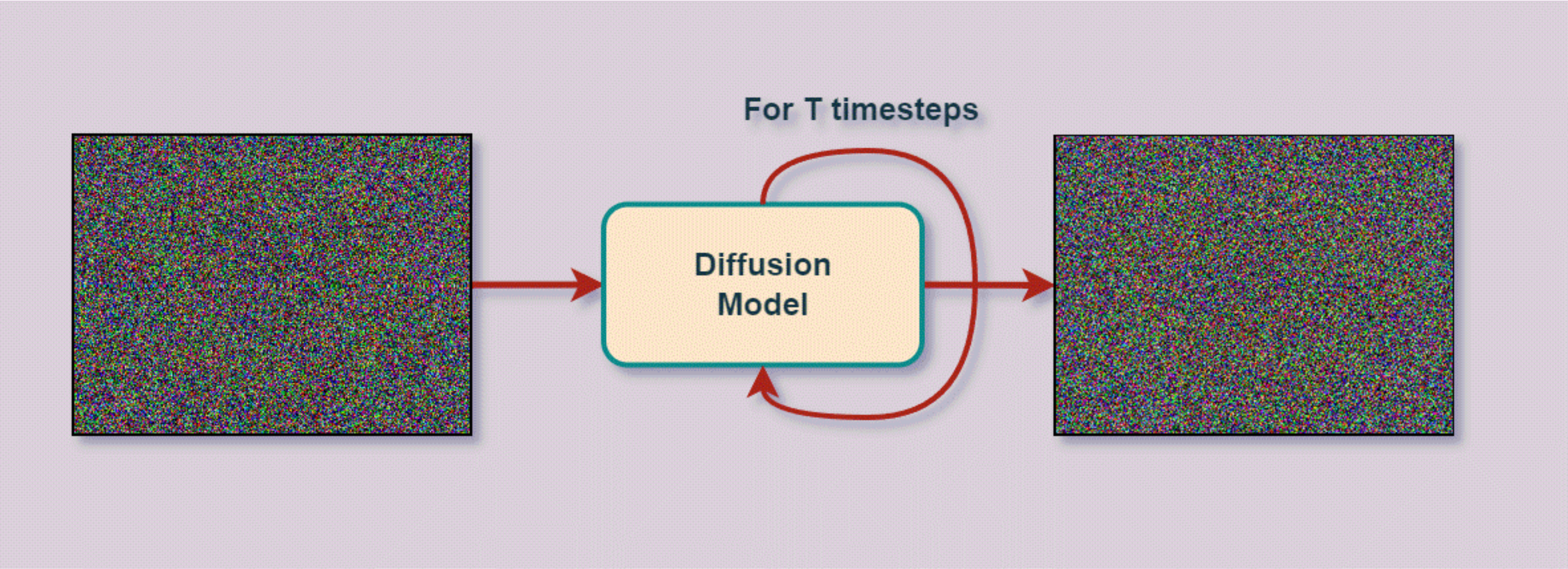
A diffúziós modellek architektúráira (U-Net, Transformer) még visszatérünk…

Diffúziós Modellek



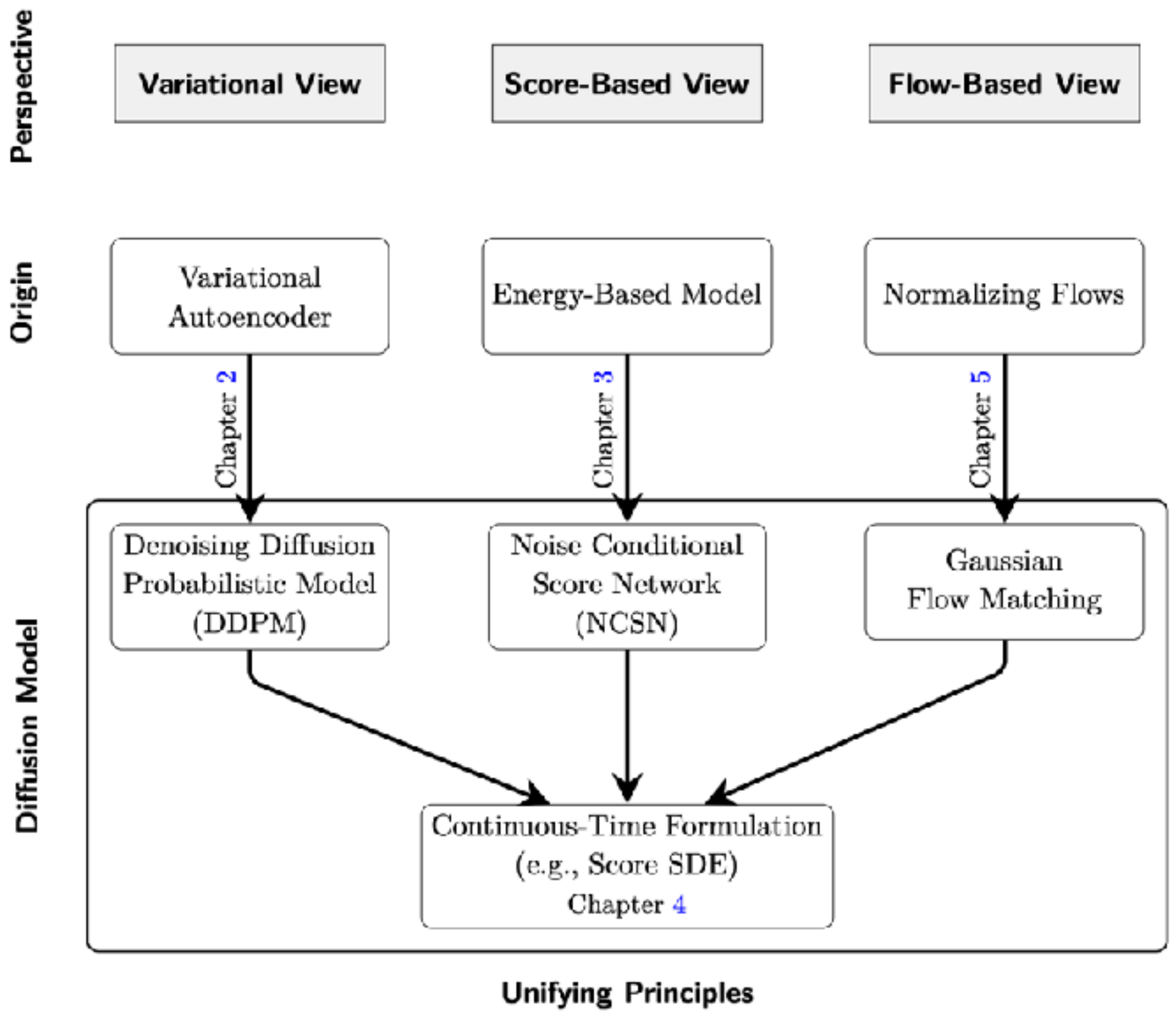
<u>Diffúziós modellek</u>: iteratív generatív modellek egy általános családja

Diffúziós Modellek



<u>Diffúziós modellek</u>: iteratív generatív modellek egy általános családja

Diffúziós Modellek



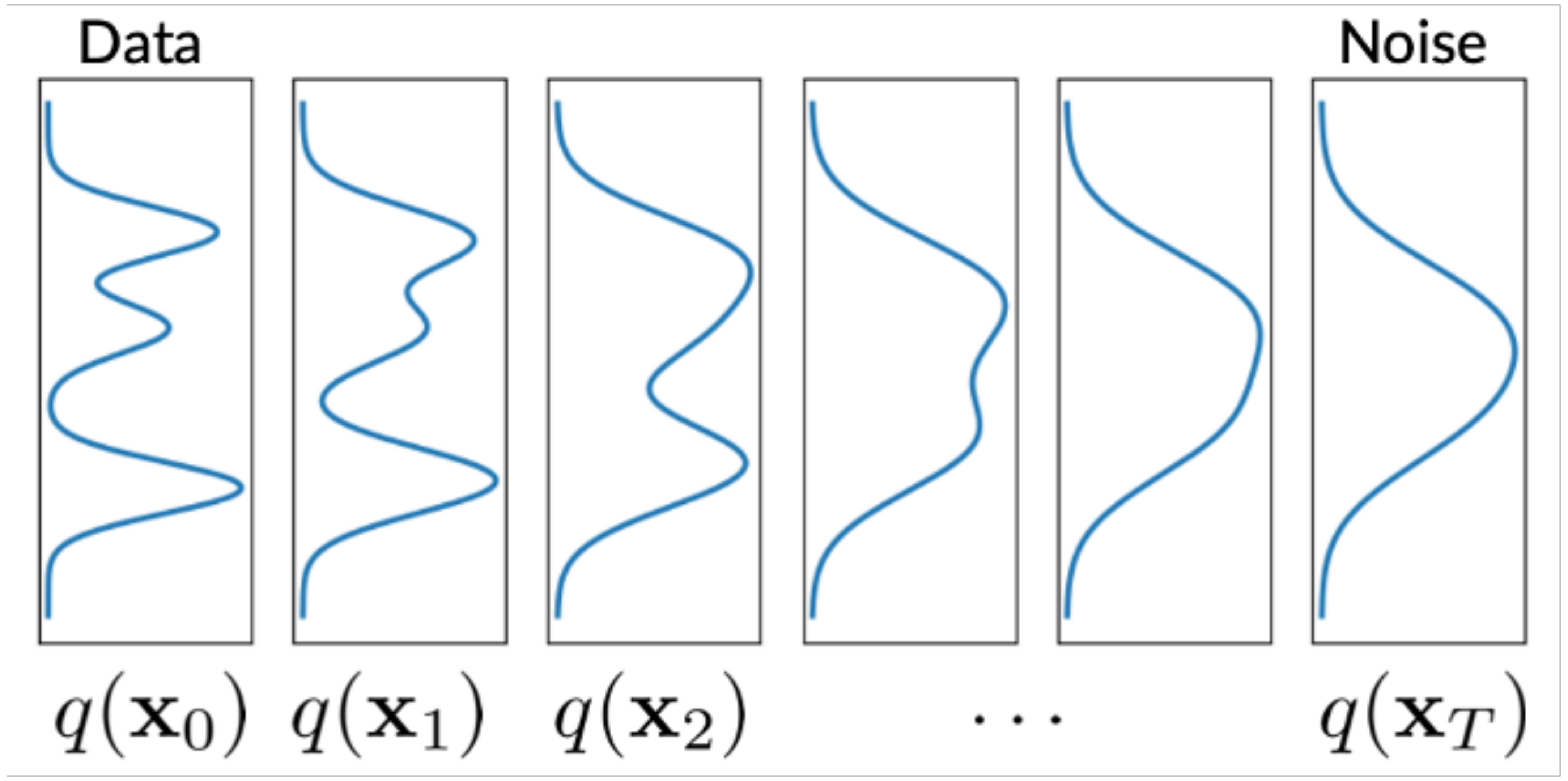
A score matching a diffúziós modellek egy lehetséges felfogása…
De nem az egyetlen, számos különböző perspektíva létezik!

Diffúziós Modellek

Forward folyamat

Emlékezzünk:

$$\mathrm{Var}(A \cdot x + B \cdot y) = A^2 \cdot \mathrm{Var}(x) + B^2 \cdot \mathrm{Var}(y) + 2AB \cdot \mathrm{Cov}(A, B)$$

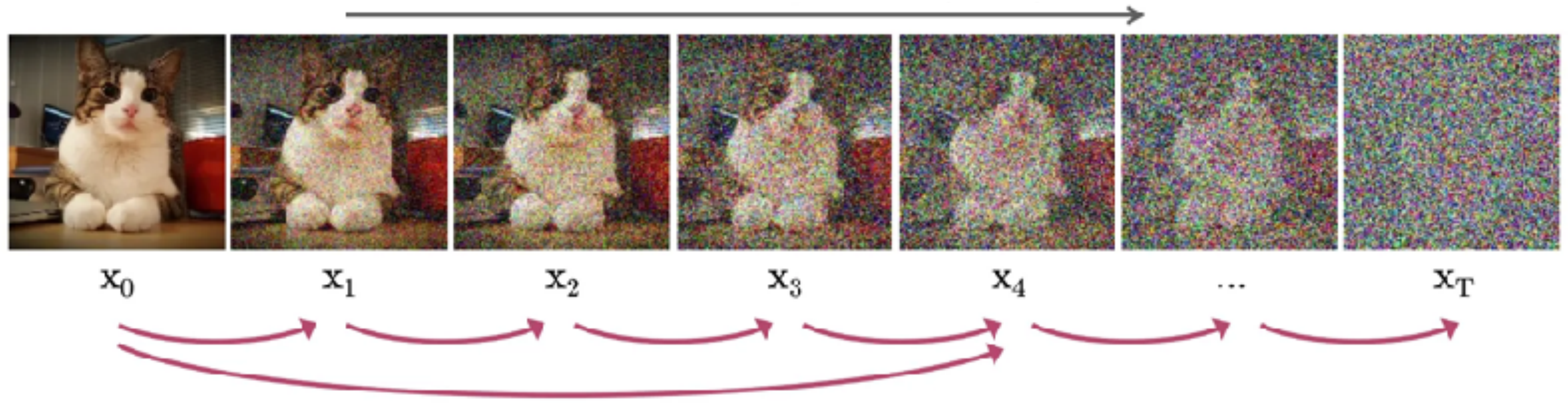


- **<u>Forward diffúziós folyamat</u>** (fix!):

$$q(x_t \mid x_0) \sim \mathcal{N}(\sqrt{\bar{\alpha}_t} \cdot x_0, (1 - \bar{\alpha}_t) \cdot I)$$

$$\bar{\alpha}_t = \prod_{s=1}^{t} (1 - \beta_s)$$

$\beta_t \in (0,1),\ t \in [1, T-1]$:
<u>zaj ütemezés</u>
(interpoláció az adat ($T = 0$) és $\mathcal{N}(0, I)$ zaj ($T = 1$) között)



- Mintavétel:

$$q(x_t \mid x_0) = \sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon \qquad \epsilon \sim \mathcal{N}(0, I)$$

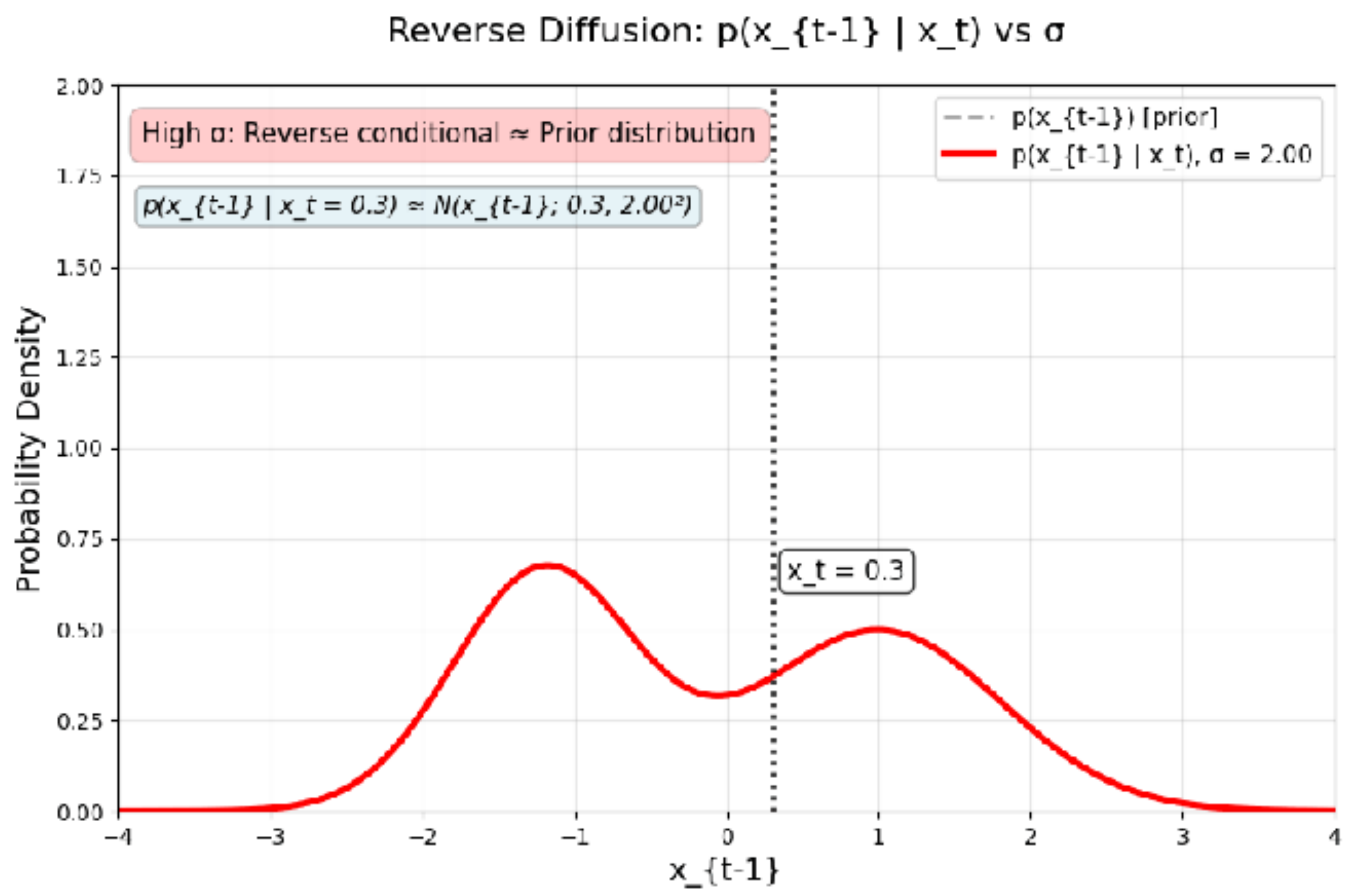
Diffúziós Modellek

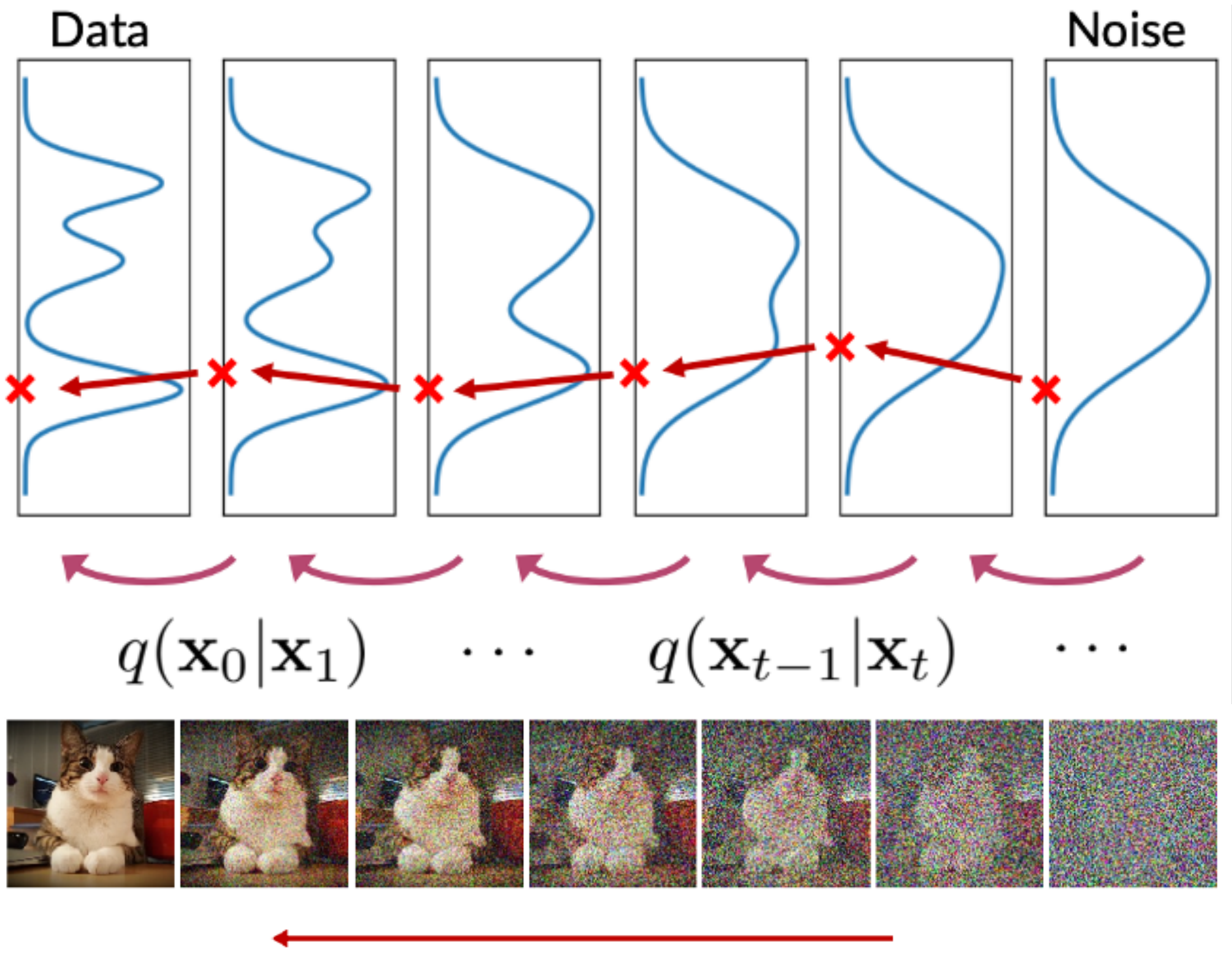
Reverse folyamat

- **<u>Reverse diffúziós folyamat:</u>**

$$p(x_{t-1} \mid x_t) \underset{Bayes}{\propto} p(x_t \mid x_{t-1}) \cdot p(x_{t-1})$$ ☠

- Kis lépésre közelíthető <u>normális eloszlással</u>!





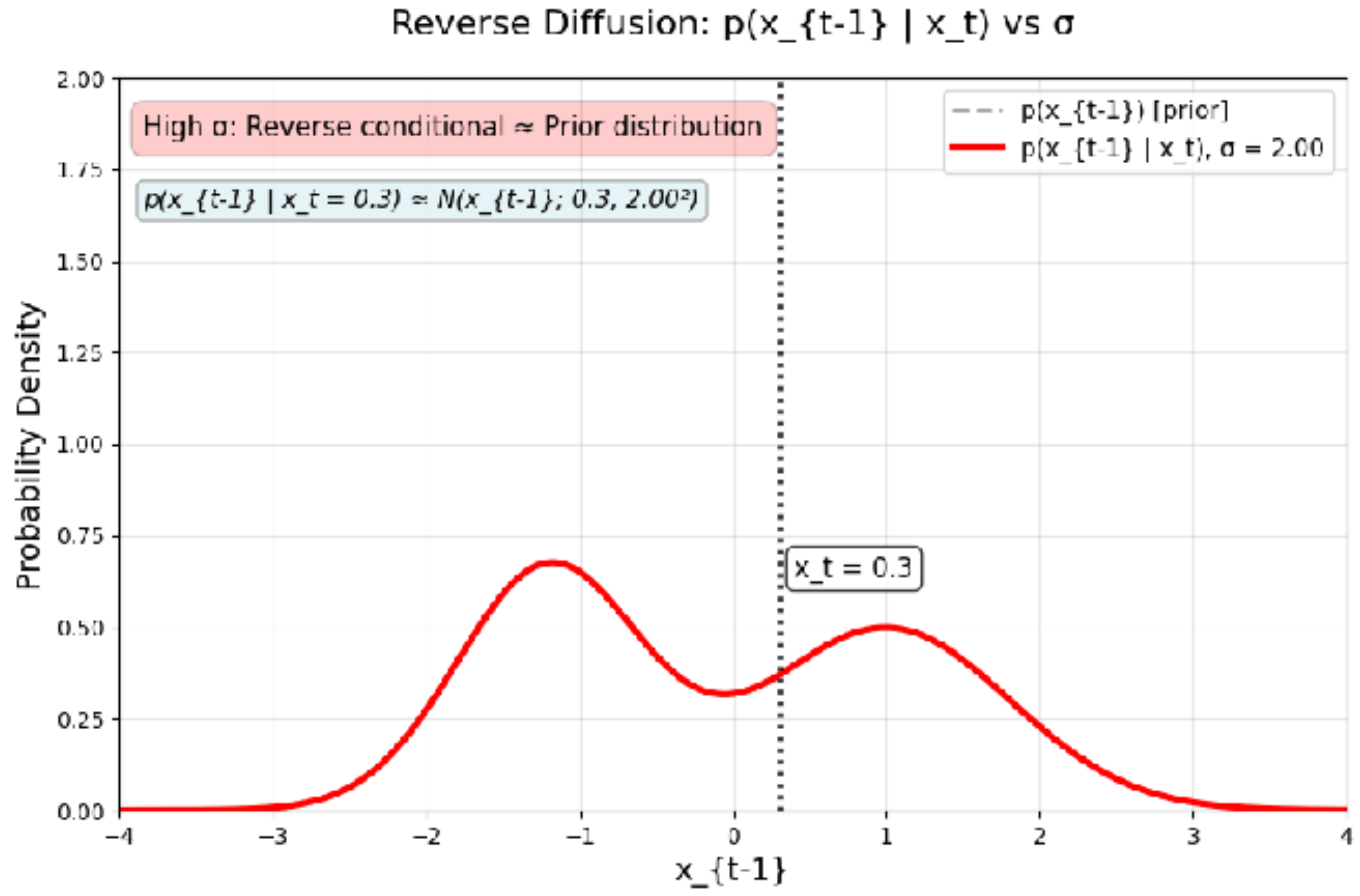
Diffúziós Modellek

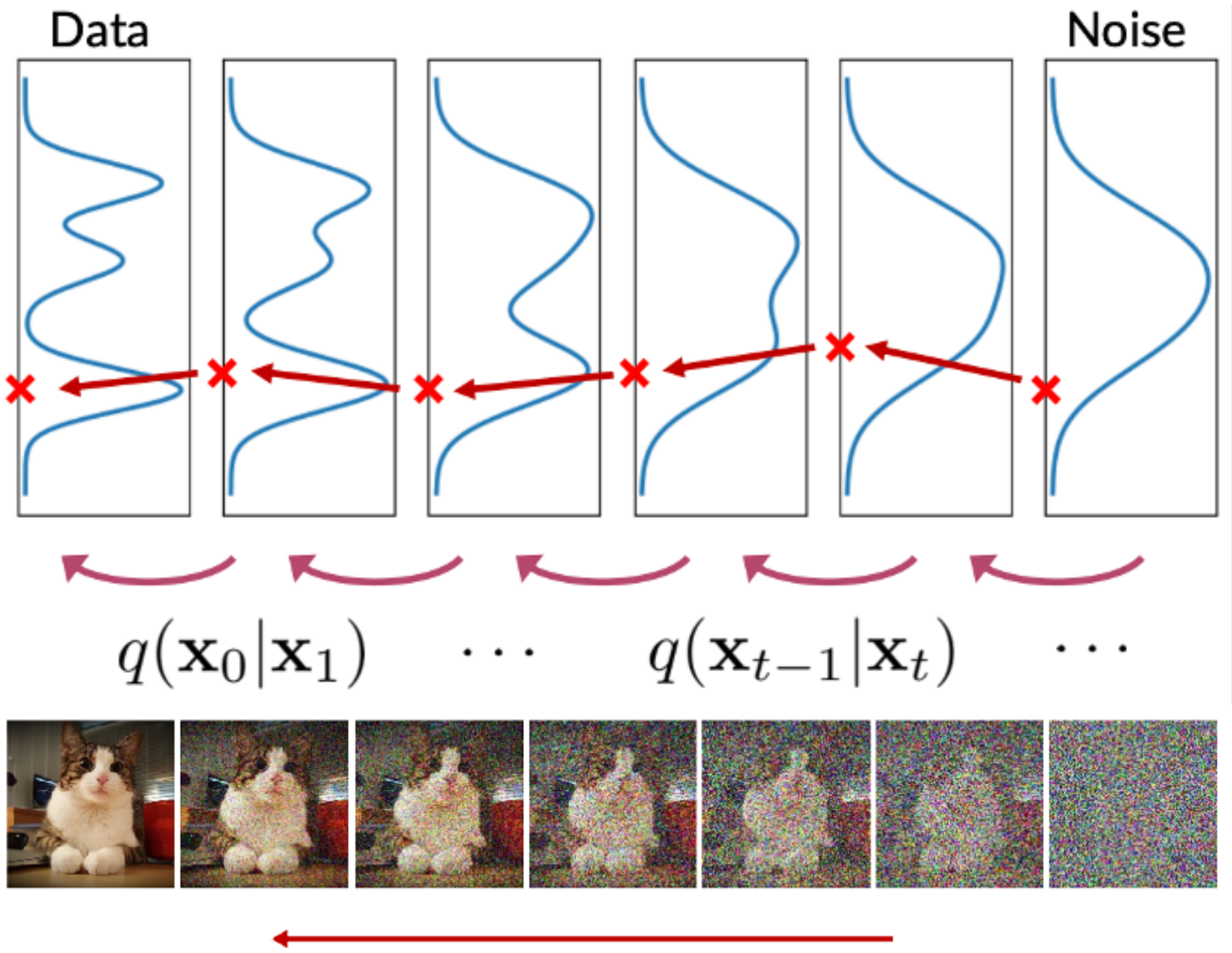
Reverse folyamat

- <u>**Reverse diffúziós folyamat:**</u>

$$p(x_{t-1} \mid x_t) \underset{Bayes}{\propto} p(x_t \mid x_{t-1}) \cdot p(x_{t-1}) ☠$$

- Kis lépésre közelíthető <u>normális eloszlással</u>!





Diffúziós Modellek
Reverse folyamat

- Reverse posterior eloszlása kezelhetőbb, ha ismerjük a zajmentes képet:

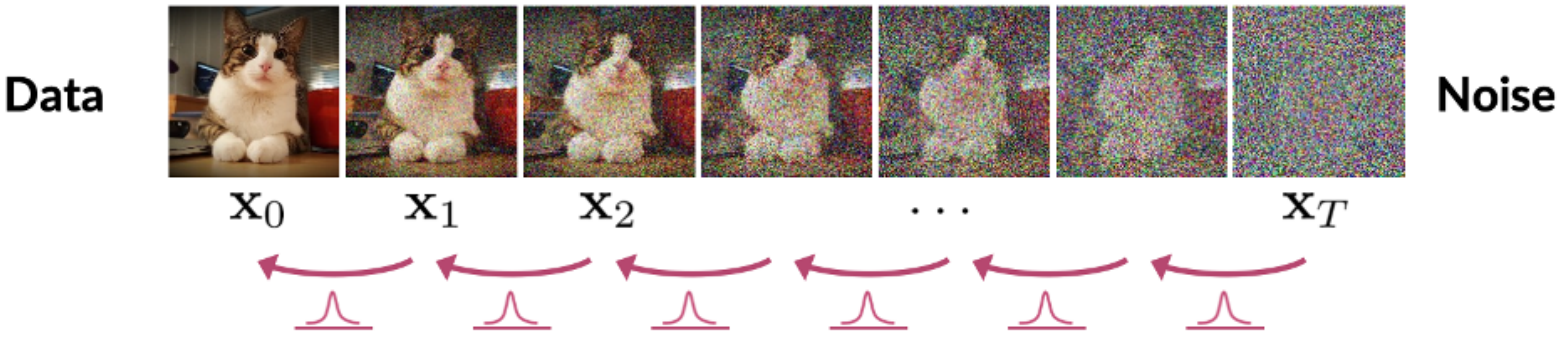
$$q(x_{t-1} \mid x_t, x_0) \propto \underbrace{q(x_t \mid x_{t-1}, x_0)}_{q(x_t|x_{t-1})} \cdot q(x_{t-1} \mid x_0)$$

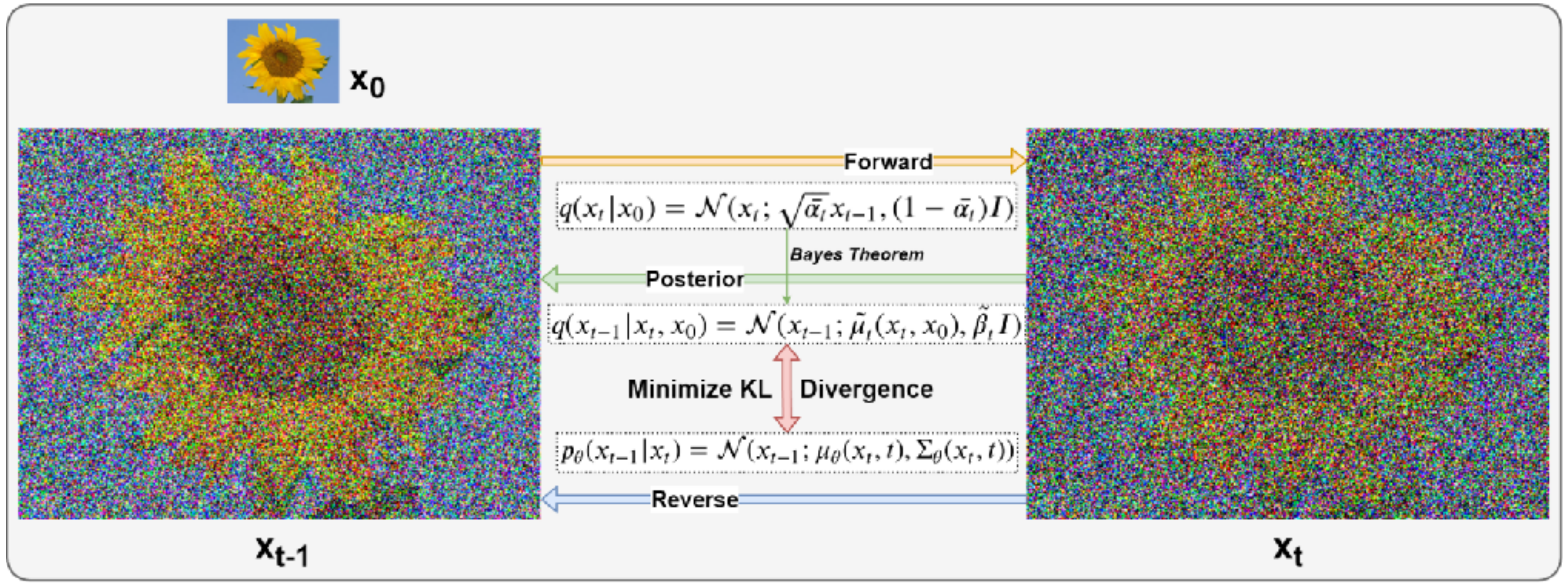
Minden normális eloszlású!

- Közelítsük neurális hálóval:

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(\mu_\theta(x_t, \boxed{t}), \beta_t \cdot I)$$

Azonos háló minden lépésre!





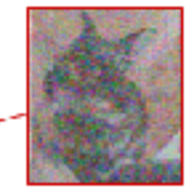
Diffúziós Modellek

Denoising Diffusion Probabilistic Model (DDPM)

- Tanítás log-likelihood alapján — közelítő ELBO (a la VAE)
  $\Downarrow$
  Gauss eloszlásokra szimpla négyzetes hiba!

- **<u>DDPM</u>** $\approx$ score matching!

- A legtöbb diffúziós modell hasonló loss-t minimalizál, csak a generálás (mintavételezés) módszerében különbözik!

x_t 

 x_0

Algorithm 1 Training

1: **repeat**
2: $\quad \mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $\quad t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: $\quad$ Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boxed{\sqrt{\bar\alpha_t}\mathbf{x}_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon}}, t) \right\|^2$$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
4: $\quad \mathbf{x}_{t-1} = \boxed{\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar\alpha_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right)} + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

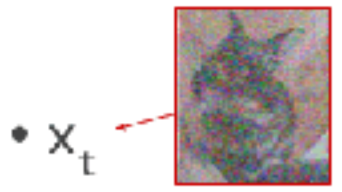
Diffúziós Modellek

Denoising Diffusion Probabilistic Model (DDPM)

- Tanítás log-likelihood alapján — közelítő ELBO (a la VAE)
  $\Downarrow$
  Gauss eloszlásokra szimpla négyzetes hiba!

- **DDPM** $\approx$ score matching!

- A legtöbb diffúziós modell hasonló loss-t minimalizál, csak a generálás (mintavételezés) módszerében különbözik!



Algorithm 1 Training

1: **repeat**
2: $\quad \mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $\quad t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: $\quad$ Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\boxed{\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}}, t\right) \right\|^2$$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
4: $\quad \mathbf{x}_{t-1} = \boxed{\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right)} + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

Diffúziós Modellek

Zaj ütemezési stratégiák

Ami valójában számít: $SNR_t = \frac{\alpha_t^2}{\sigma_t^2}$

- A zajos kép paraméterezése:

$$x_t = \alpha_t \cdot x_0 + \sigma_t \cdot \epsilon$$

Tiszta kép Gaussi zaj $\sim \mathcal{N}(0,I)$

- **Variance-Preserving (VP)**:

$$\alpha(t) = \sqrt{1 - \sigma^2(t)}$$

- **Variance-Exploding**:

$$\alpha(t) = 1$$

$$\sigma(t) = t \text{ ("EDM")}$$

- **Sub-VP** (folyamillesztés):

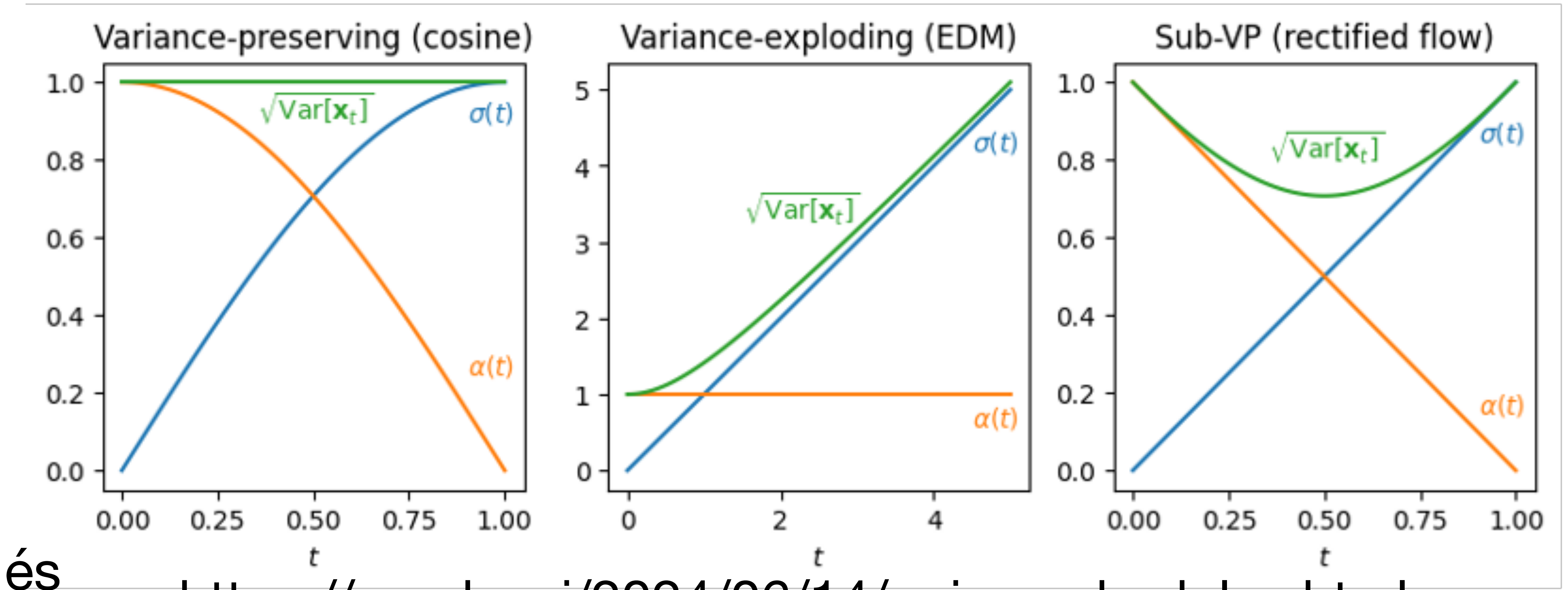
$$\alpha(t) = t$$

$$\sigma(t) = 1 - t$$

DEMO!

DDPM: $\sigma(t) = \sqrt{1 - \prod (1 - \beta_i)}$ — ma már kevésbé használatos

Cosine scheduling: $\sigma(t) = \sin\left(\frac{t/T + s}{s} \cdot \frac{\pi}{2}\right)$ — zajos képekre gyorsabb, tisztábbakra lassabb, gyakran használt!



Akár különbözhet tanítás és mintavételezés közben!

https://sander.ai/2024/06/14/noise-schedules.html

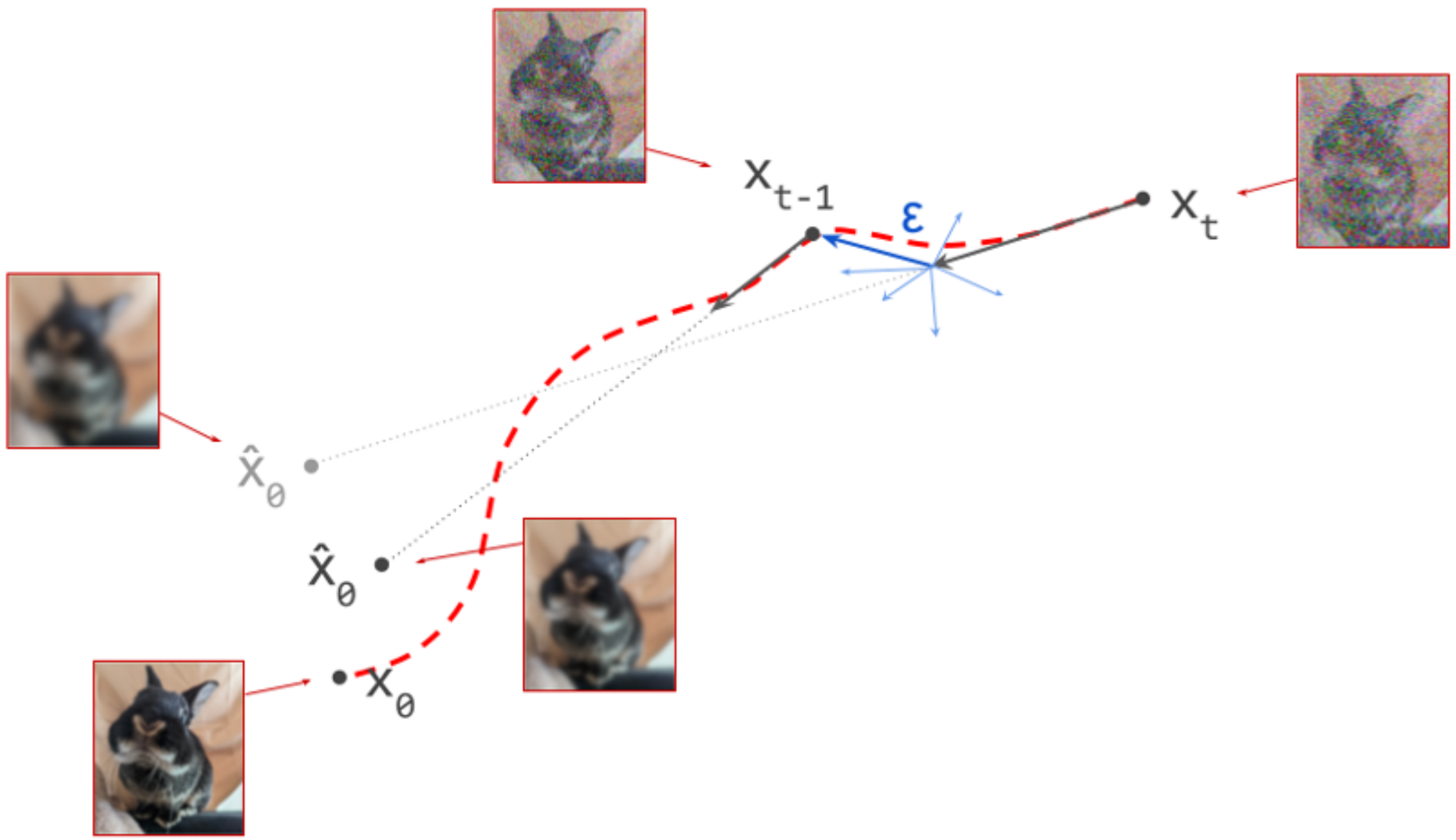
Diffúziós Modellek

!FIGYELEM!

A diffúziós szakirodalom (ezen fóliákat beleértve)
nem következetes a jelölések tekintetében.
Legyünk óvatosak!

Diffúziós Modellek

A DDPM-en túl?



A DDPM mintavételezés eléggé lassú (1000+ lépés is szükséges lehet)
Minden lépéssel a tiszta képre adott jelenlegi becslést “követjük”.
A sok kicsi lépés folytonos pályát ír le — Be tudjuk futni másként is?

Diffúziós Modellek

Denoising Diffusion Implicit Model (DDIM)

- Minden denoizer lépés becslést ad a tiszta képre:

$$x_{0,k} = \frac{x_k - \epsilon_\theta(x_k)}{\alpha_k}$$

- **<u>DDIM</u>** mintavételezés — általános lineáris kombináció:

$$x_{k-1} = A \cdot x_{0,k} + B \cdot x_k + C \cdot \epsilon$$

- Lehet $C = 0$ – **<u>determinisztikus mintavételezés</u>**!

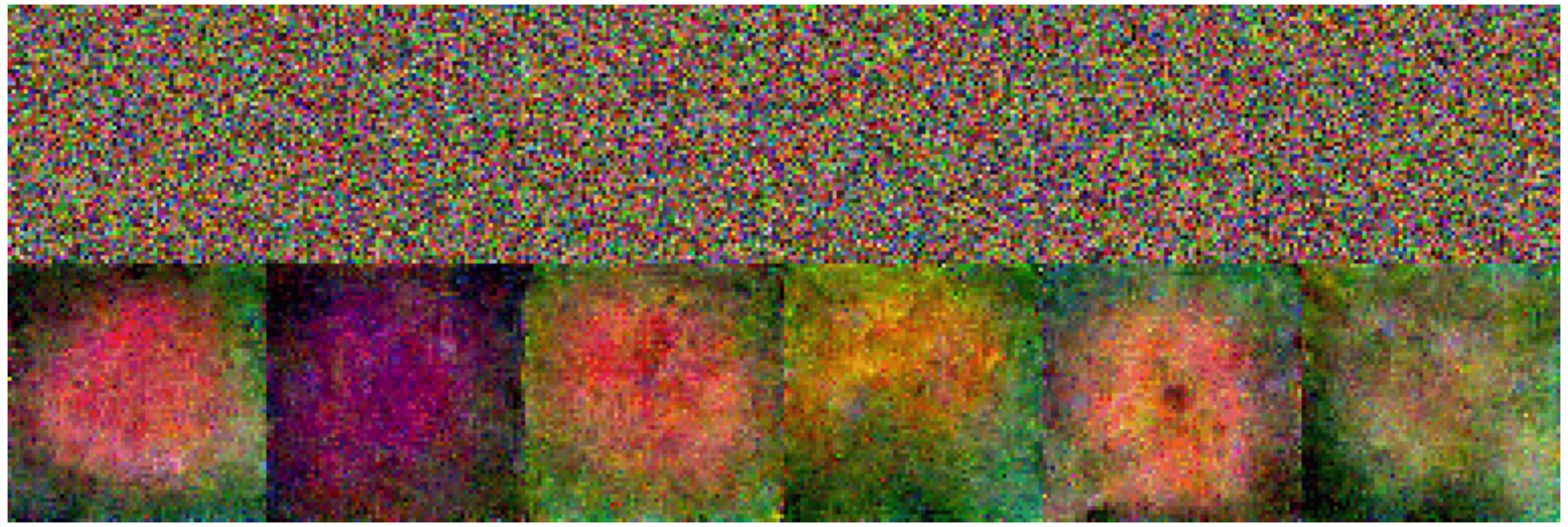
- DDPM-nél sokkal gyorsabb mintavétel: 20-100 lépés a több 1000 helyett!

- A modern ODE és folyam-alapú mintavételezés speciális esete


DENOISING DIFFUSION IMPLICIT MODELS

Jiaming Song, Chenlin Meng & Stefano Ermon
Stanford University

$A + B = 1$ (egyenes interpoláció zaj és kép között)
$A^2 \cdot \sigma_k + C^2 = \sigma_{k-1}$ (forward folyamat megőrzése)
<u>A, B, C közül egy szabadon választható!</u>



Diffúziós Modellek

Denoising Diffusion Implicit Model (DDIM)

- Minden denoizer lépés becslést ad a tiszta képre:

$$x_{0,k} = \frac{x_k - \epsilon_\theta(x_k)}{\alpha_k}$$

- **DDIM** mintavételezés — általános lineáris kombináció:

$$x_{k-1} = A \cdot x_{0,k} + B \cdot x_k + C \cdot \epsilon$$

- Lehet $C = 0$ — **determinisztikus mintavételezés**!

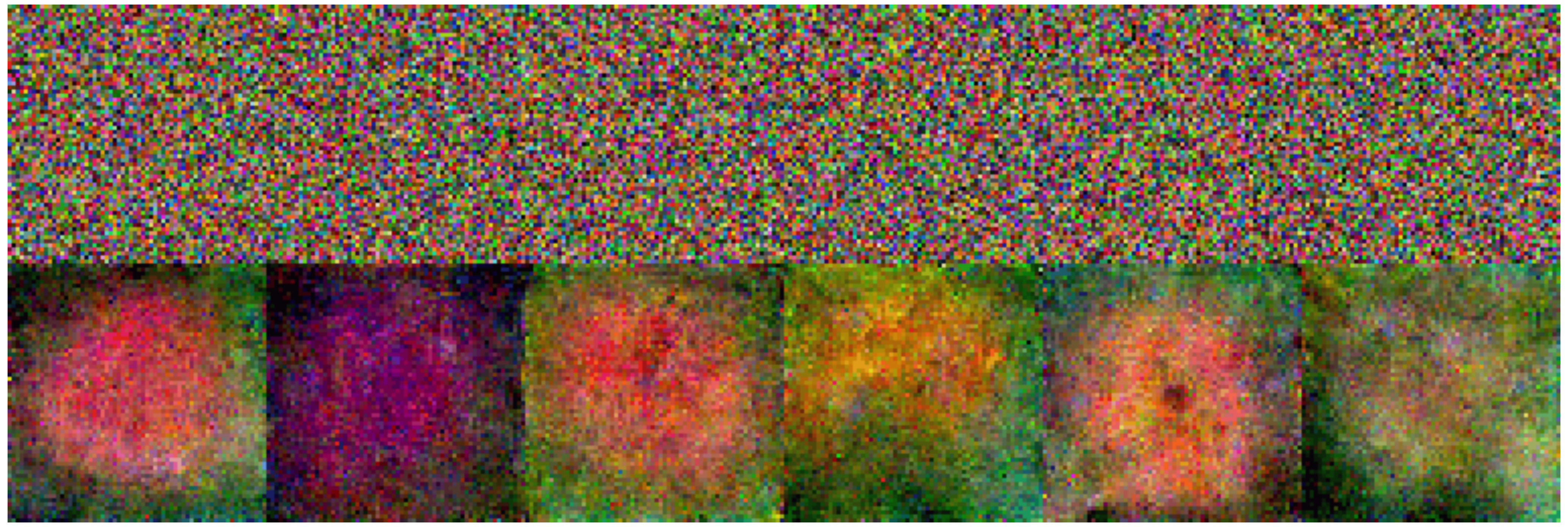
- DDPM-nél sokkal gyorsabb mintavétel: 20-100 lépés a több 1000 helyett!

- A modern ODE és folyam-alapú mintavételezés speciális esete


DENOISING DIFFUSION IMPLICIT MODELS

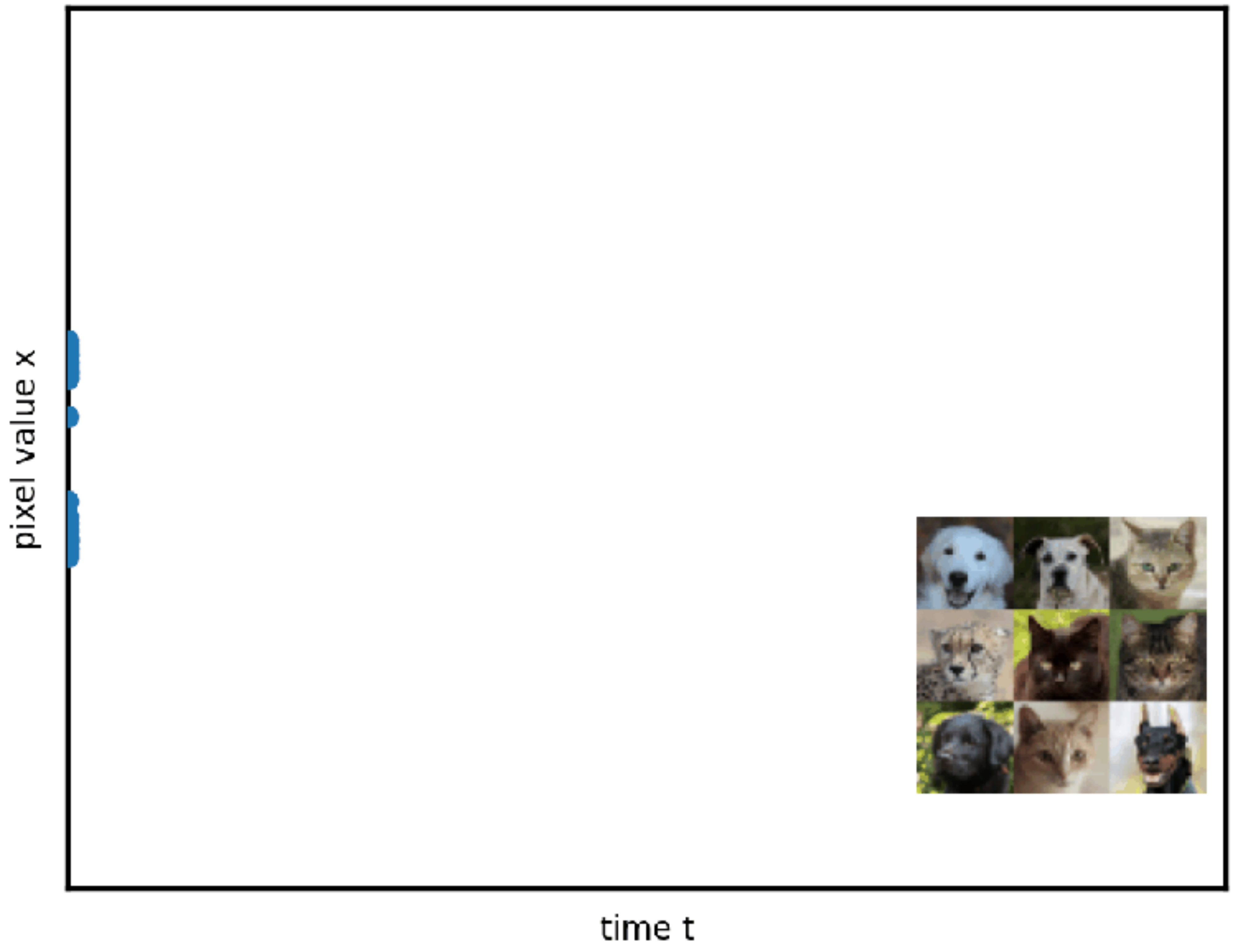
Jiaming Song, Chenlin Meng & Stefano Ermon
Stanford University

$A + B = 1$ (egyenes interpoláció zaj és kép között)
$A^2 \cdot \sigma_k + C^2 = \sigma_{k-1}$ (forward folyamat megőrzése)
A, B, C közül egy szabadon választható!



Diffúziós Modellek

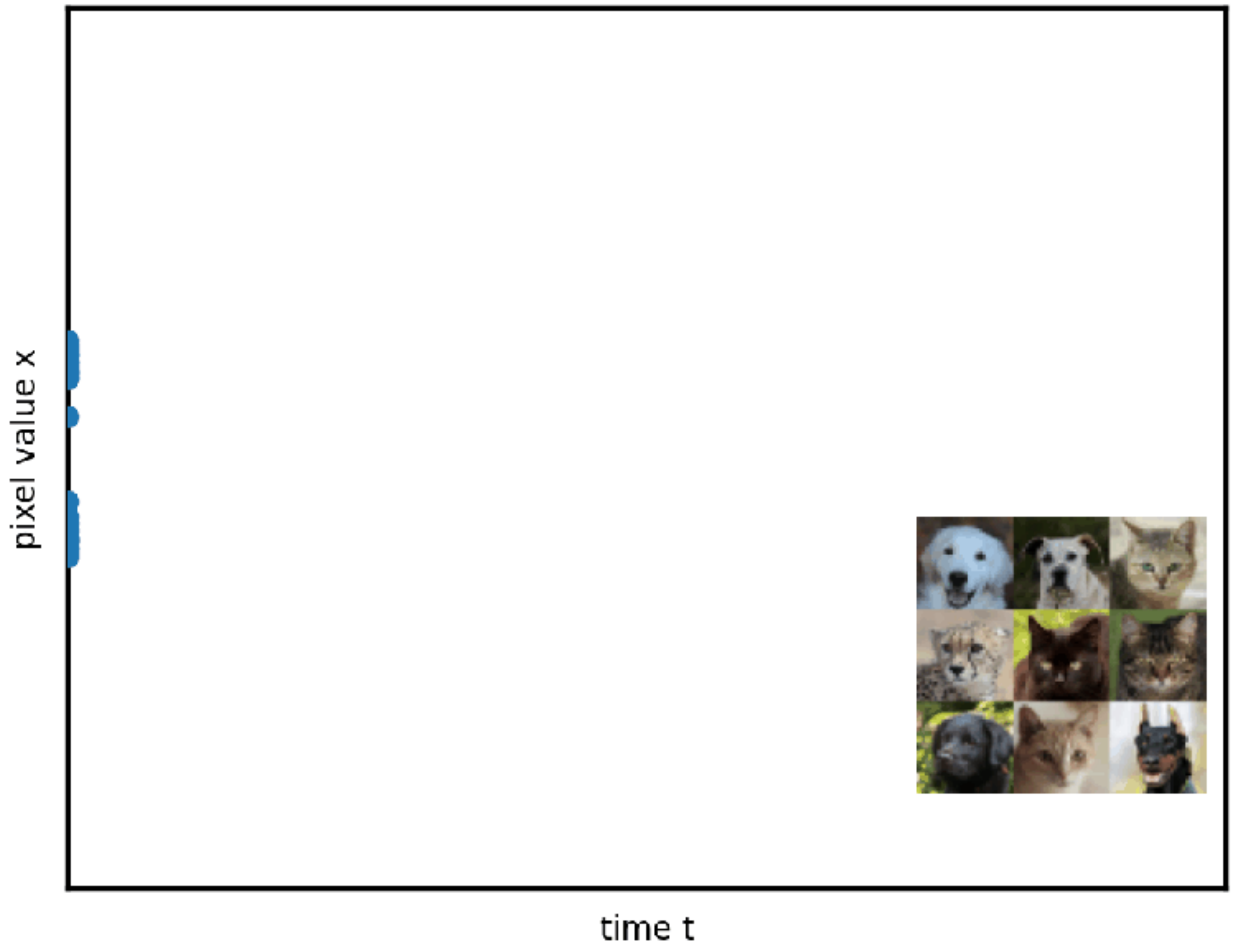
A folyam-alapú modellek felé…



Új nézőpont: egy diffúziós modell folytonos átmenetet képez eloszlások között

Diffúziós Modellek

A folyam-alapú modellek felé…



Új nézőpont: egy diffúziós modell folytonos átmenetet képez eloszlások között

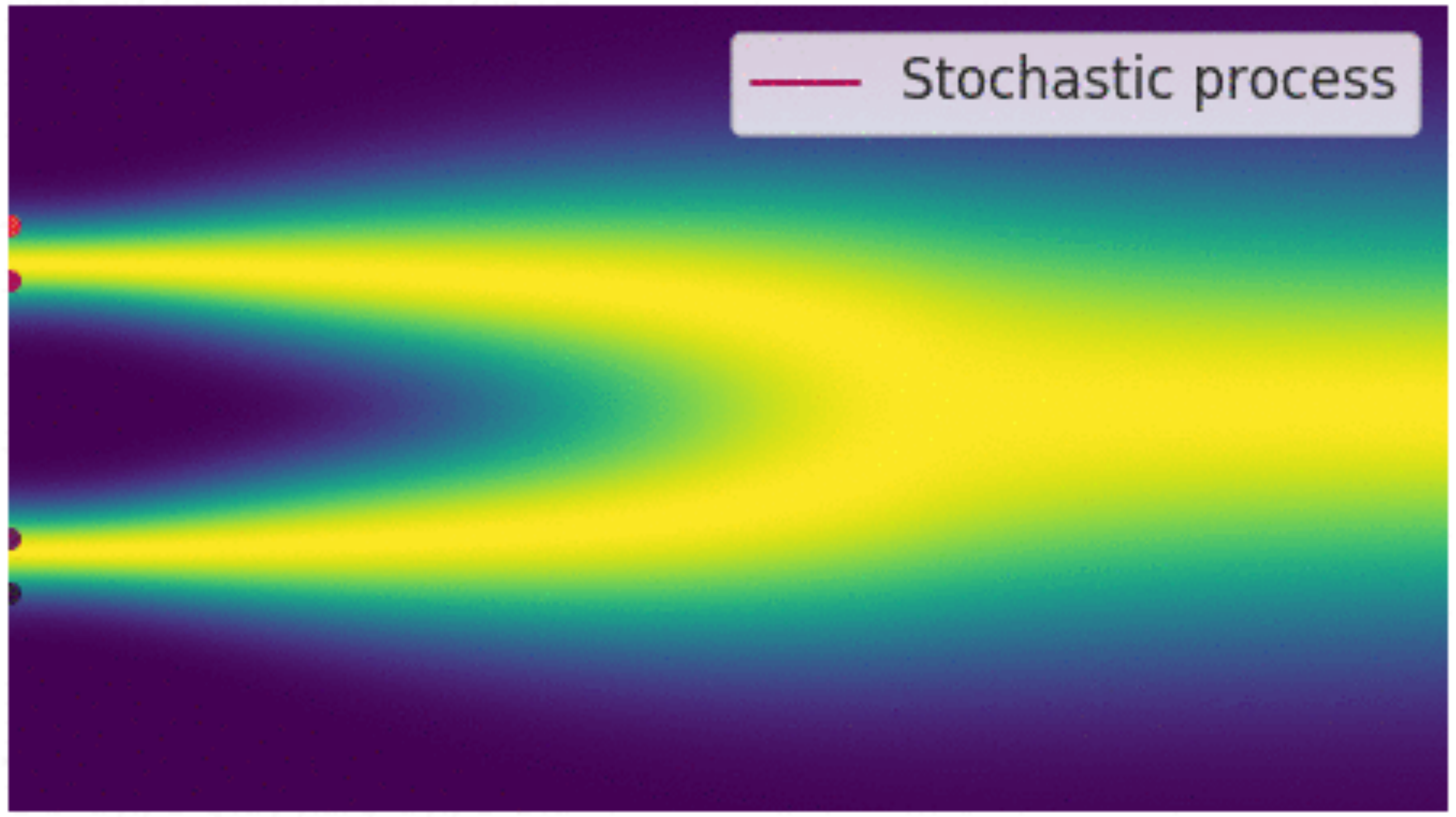
Diffúziós Modellek

SDE értelmezés

- Végtelen finom zajskálákra a forward folyamat megfelel egy **<u>sztochasztikus differenciálegyenletnek (SDE)</u>**:

$$dx = \sigma(t)dW_t$$

$\sigma(t)$: zaj ütemezés



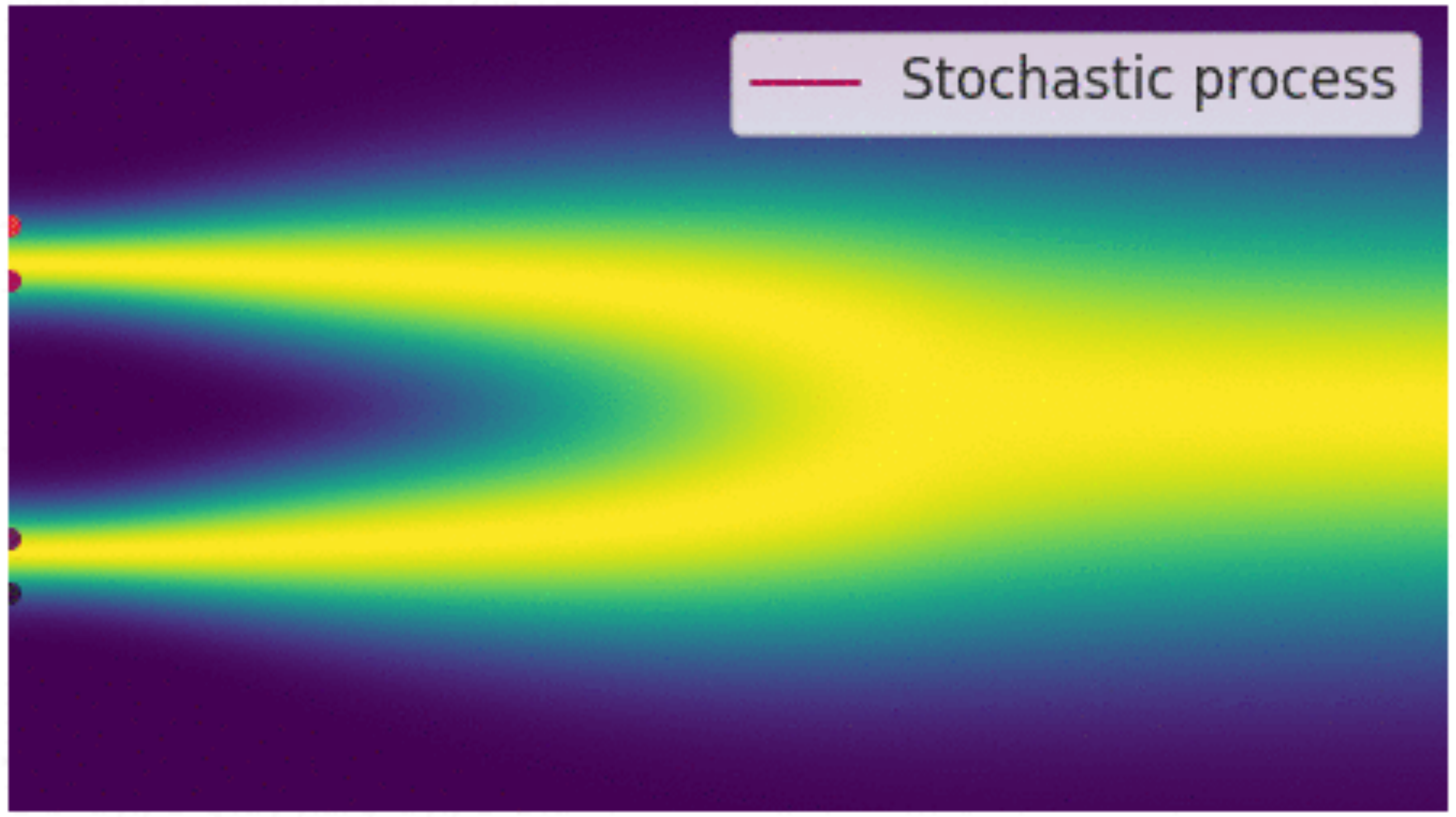
Diffúziós Modellek

SDE értelmezés

- Végtelen finom zajskálákra a forward folyamat megfelel egy **sztochasztikus differenciálegyenletnek (SDE)**:

$$dx = \sigma(t)dW_t$$

$\sigma(t)$: zaj ütemezés



Diffúziós Modellek

SDE értelmezés

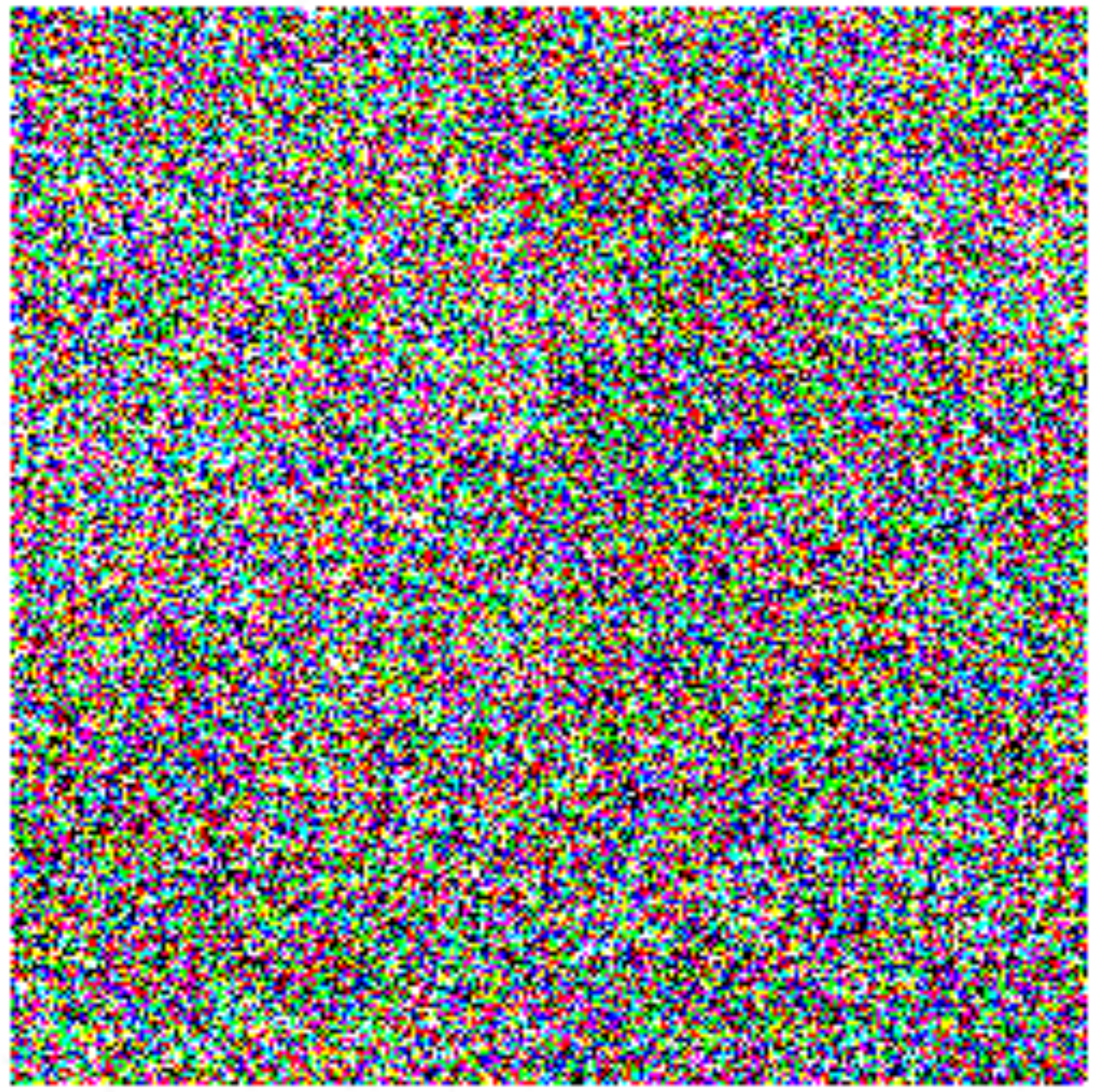
SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

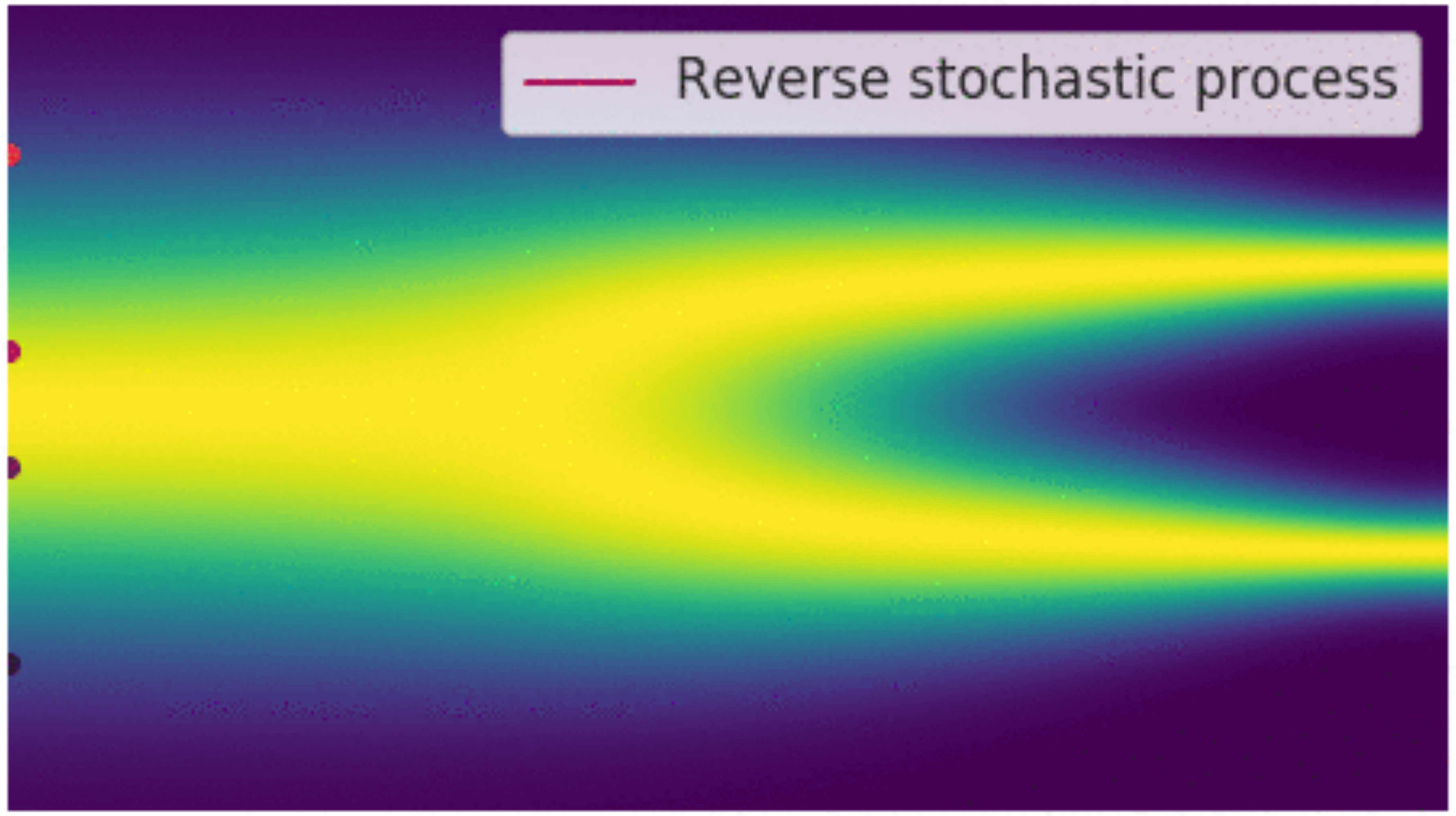
Yang Song* Stanford University yangsong@cs.stanford.edu
Jascha Sohl-Dickstein Google Brain jaschasd@google.com
Diederik P. Kingma Google Brain durk@google.com
Abhishek Kumar Google Brain abhishk@google.com
Stefano Ermon Stanford University ermon@cs.stanford.edu
Ben Poole Google Brain pooleb@google.com

- A reverse folyamat is megfelel egy SDE-nek:

$$dx = -\sigma(t)\dot{\sigma}(t)\boxed{\nabla \log p_{\sigma(t)}(x)}dt + \sigma(t)dW_t$$

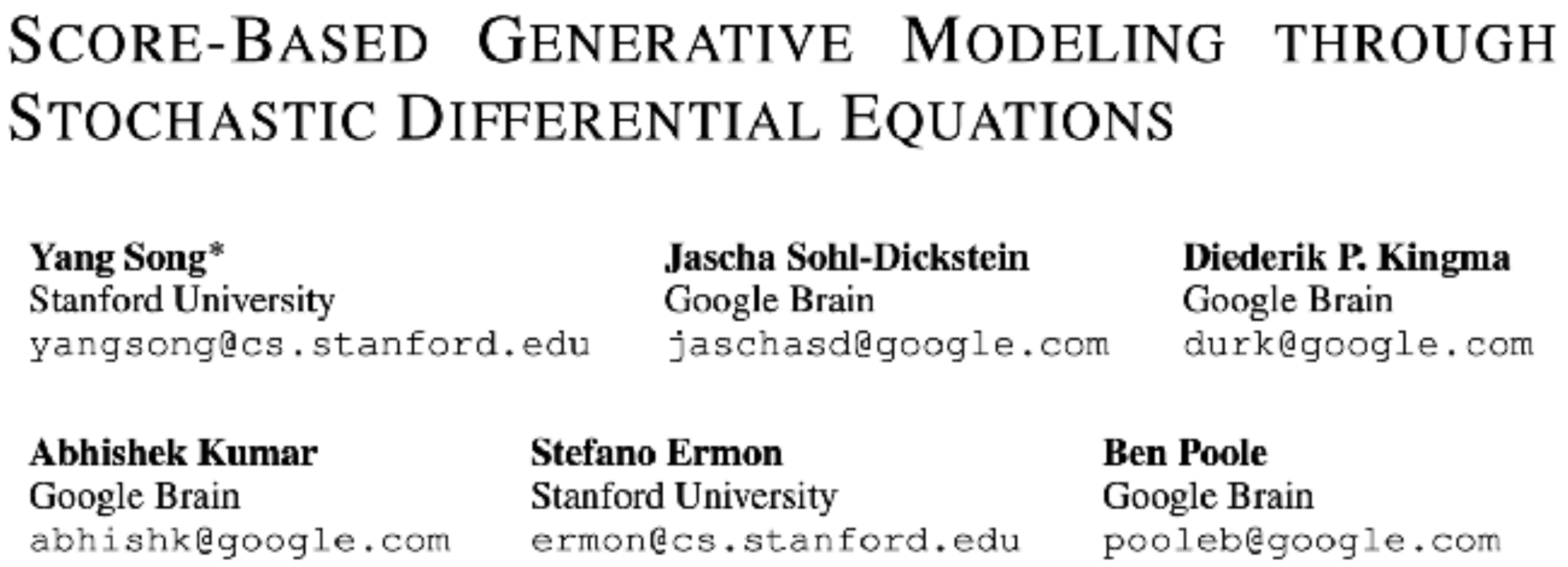
Score-függvény!





Diffúziós Modellek

SDE értelmezés


SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

Yang Song*
Stanford University
yangsong@cs.stanford.edu

Jascha Sohl-Dickstein
Google Brain
jaschasd@google.com

Diederik P. Kingma
Google Brain
durk@google.com

Abhishek Kumar
Google Brain
abhishk@google.com

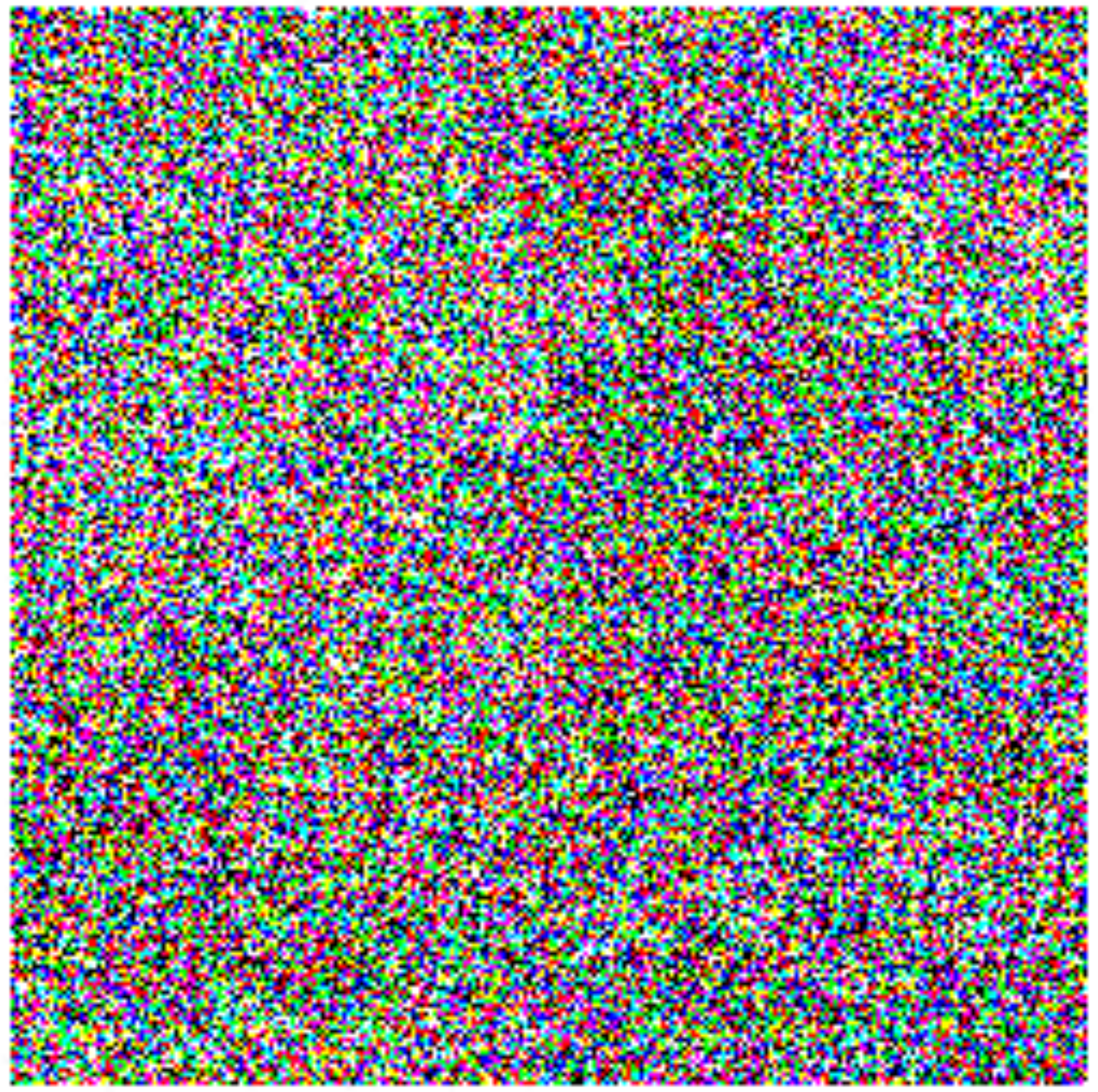
Stefano Ermon
Stanford University
ermon@cs.stanford.edu

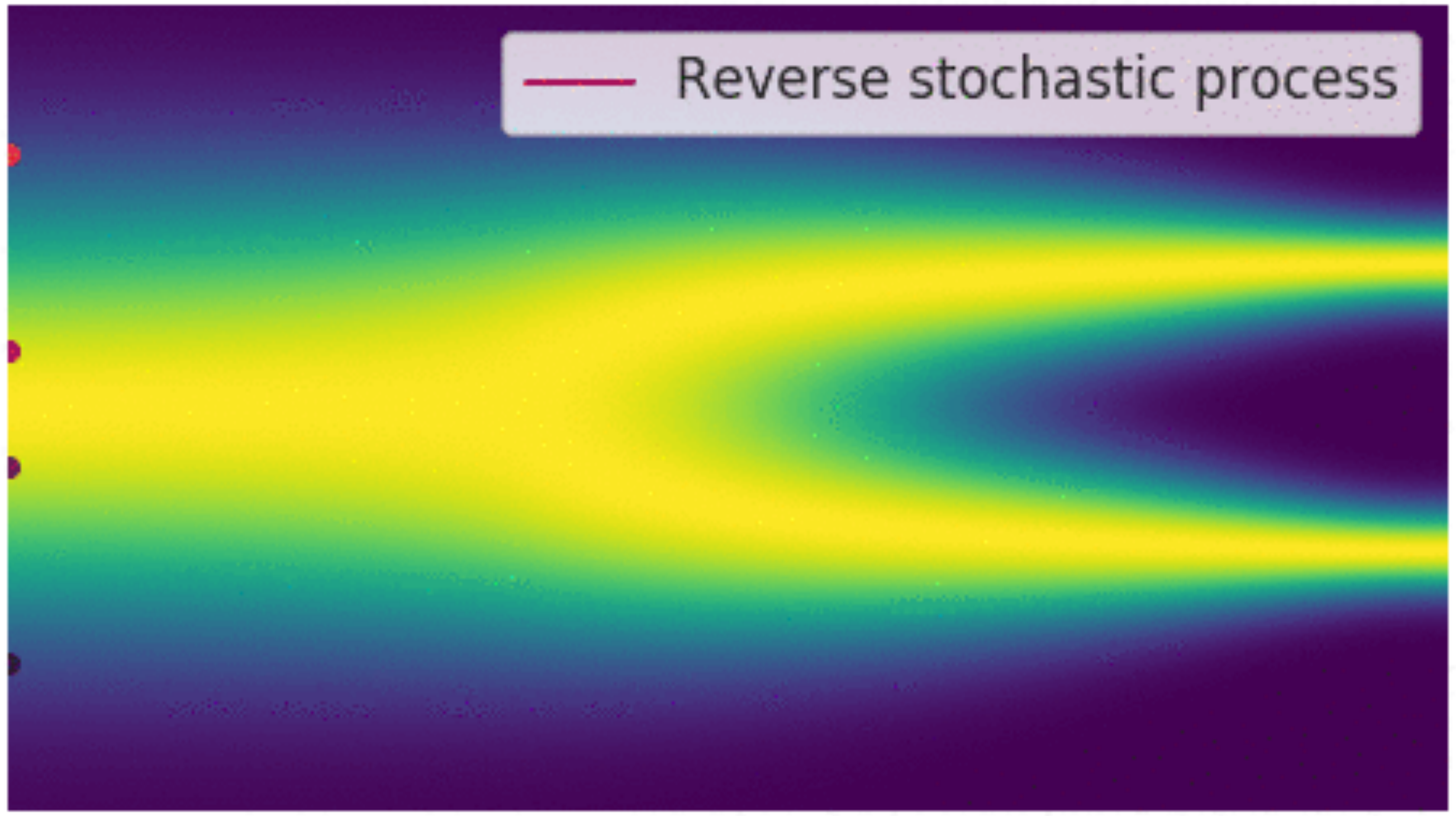
Ben Poole
Google Brain
pooleb@google.com

- A reverse folyamat is megfelel egy SDE-nek:

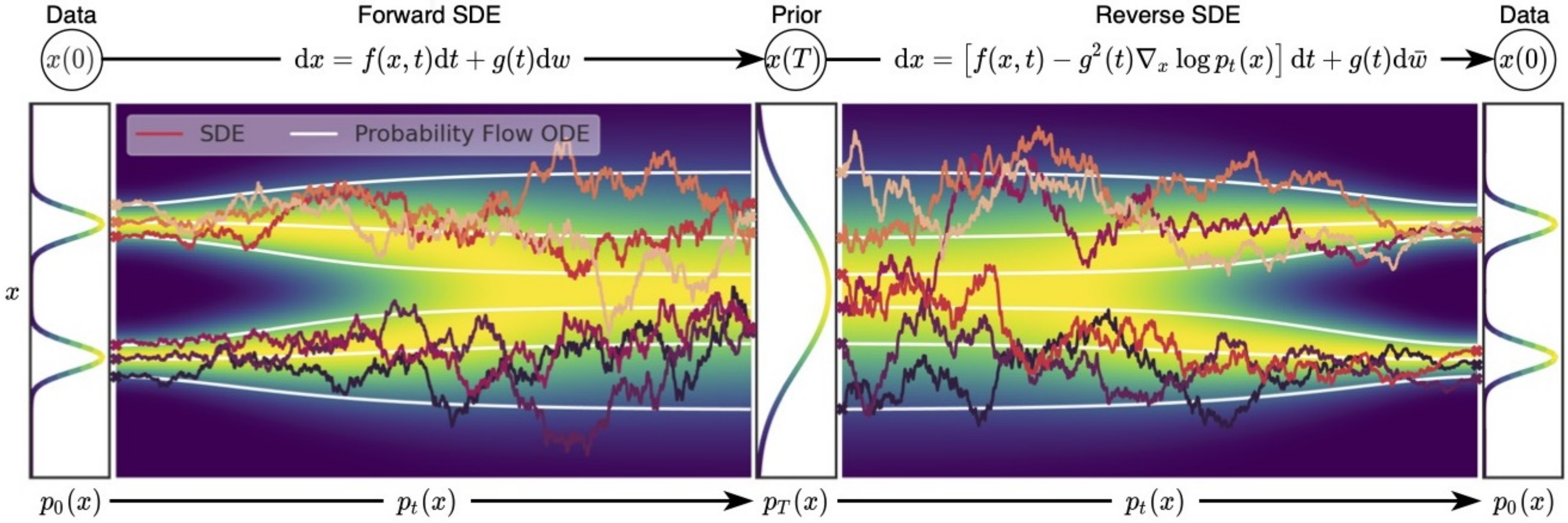
$$dx = -\sigma(t)\dot{\sigma}(t)\boxed{\nabla \log p_{\sigma(t)}(x)}dt + \sigma(t)dW_t$$

Score-függvény!





Diffúziós Modellek
SDE értelmezés



Diffúziós Modellek

ODE értelmezés

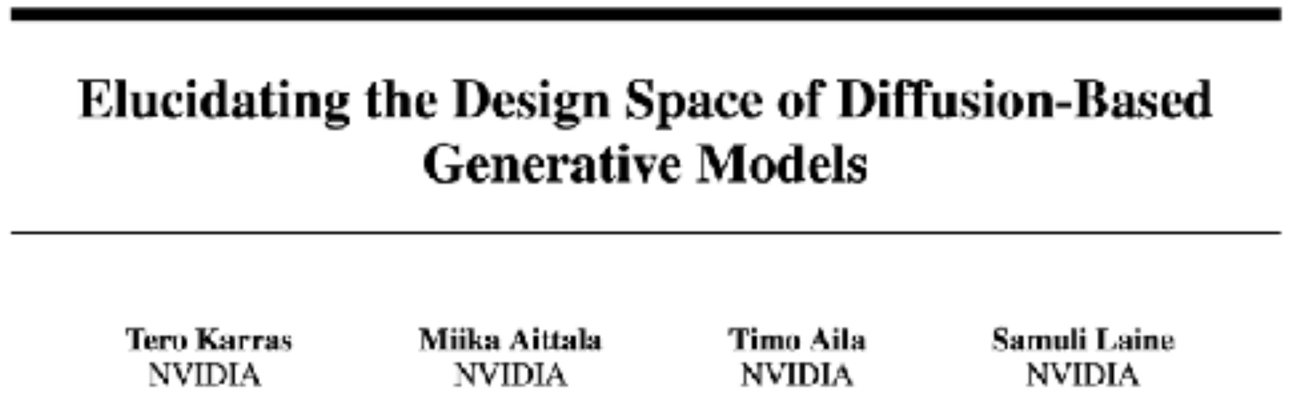
- Vehetjük az SDE-nek megfelelő (Neural) **ODE**-t:

$$\frac{dx}{dt} = -\sigma(t)\dot{\sigma}(t)\nabla \log p_{\sigma(t)}(x)$$

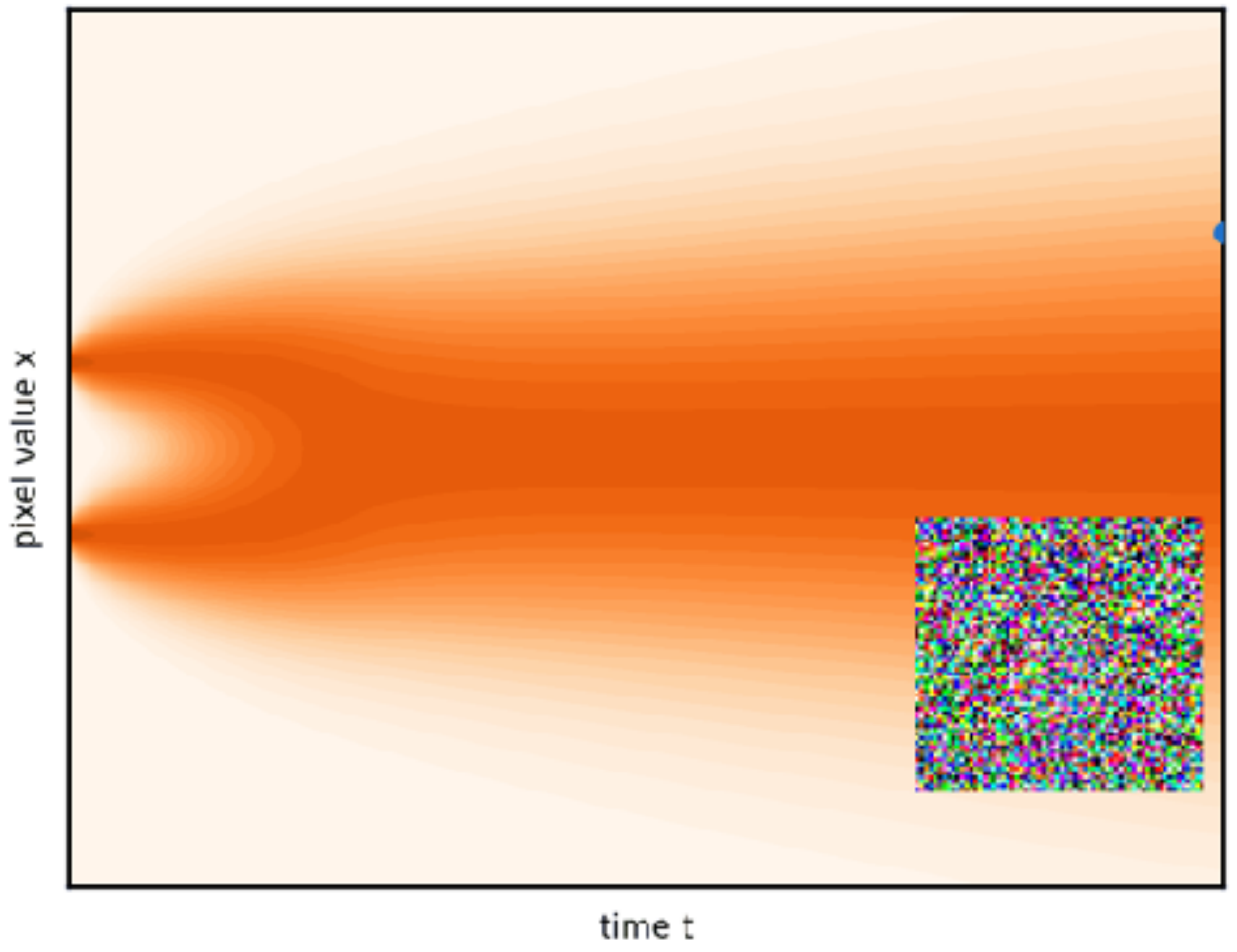
- Numerikus integrálás — **determinisztikus generálás**!

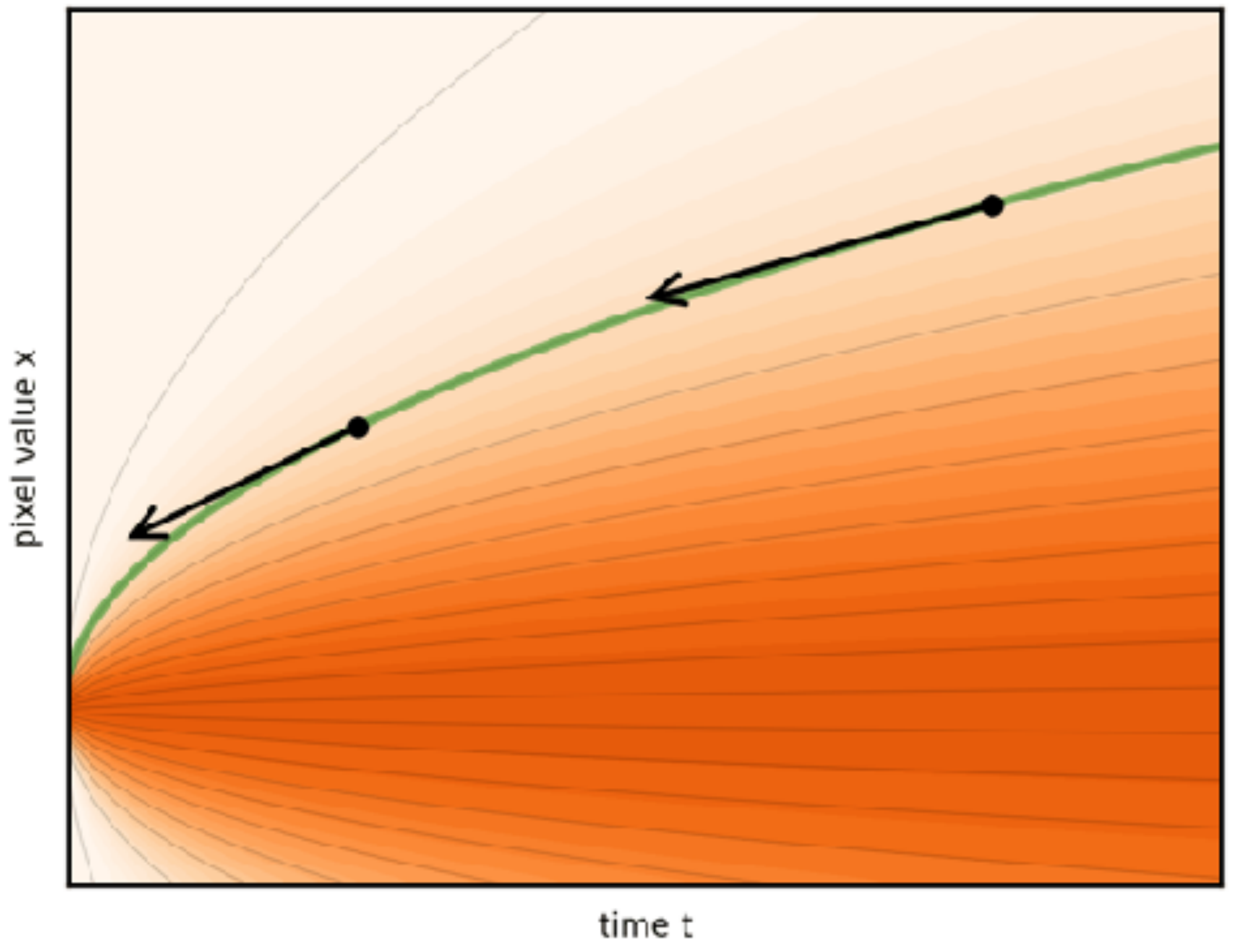
- Sok lehetséges ODE solver / $\sigma(t)$ kombináció:

		VP [49]	VE [49]	iDDPM [37] + DDIM [47]	Ours ("EDM")
Sampling (Section 3)					
ODE solver		Euler	Euler	Euler	2nd order Heun
Time steps	$t_{i<N}$	$1 + \frac{i}{N-1}(\epsilon_s - 1)$	$\sigma_{\max}^2 \left(\sigma_{\min}^2/\sigma_{\max}^2\right)^{\frac{i}{N-1}}$	$u_{\lfloor j_0 + \frac{M-1-j_0}{N-1} i + \frac{1}{2} \rfloor}$, where $u_M = 0$, $u_{j-1} = \sqrt{\frac{u_j^2+1}{\max(\bar{\alpha}_{j-1}/\bar{\alpha}_j, C_1)} - 1}$	$\left(\sigma_{\max}{}^{\frac{1}{\rho}} + \frac{i}{N-1}\left(\sigma_{\min}{}^{\frac{1}{\rho}} - \sigma_{\max}{}^{\frac{1}{\rho}}\right)\right)^{\rho}$
Schedule	$\sigma(t)$	$\sqrt{e^{\frac{1}{2}\beta_d t^2 + \beta_{\min} t} - 1}$	$\sqrt{t}$	t	t
Scaling	$s(t)$	$1/\sqrt{e^{\frac{1}{2}\beta_d t^2 + \beta_{\min} t}}$	1	1	1


Elucidating the Design Space of Diffusion-Based Generative Models

Tero Karras NVIDIA
Miika Aittala NVIDIA
Timo Aila NVIDIA
Samuli Laine NVIDIA





Diffúziós Modellek

ODE értelmezés

- Vehetjük az SDE-nek megfelelő (Neural) **ODE**-t:

$$\frac{dx}{dt} = -\sigma(t)\dot{\sigma}(t)\,\nabla \log p_{\sigma(t)}(x)$$

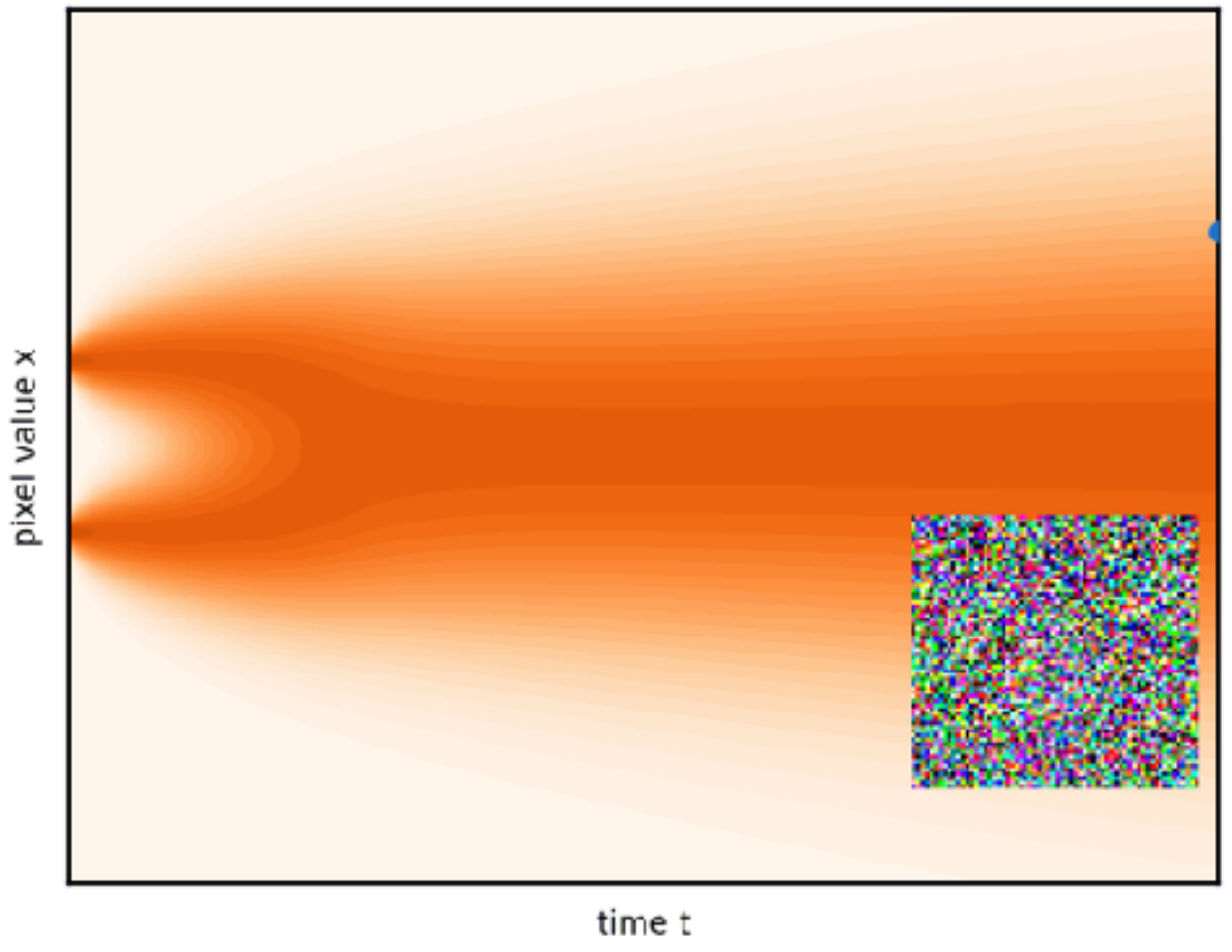
- Numerikus integrálás — **determinisztikus generálás**!

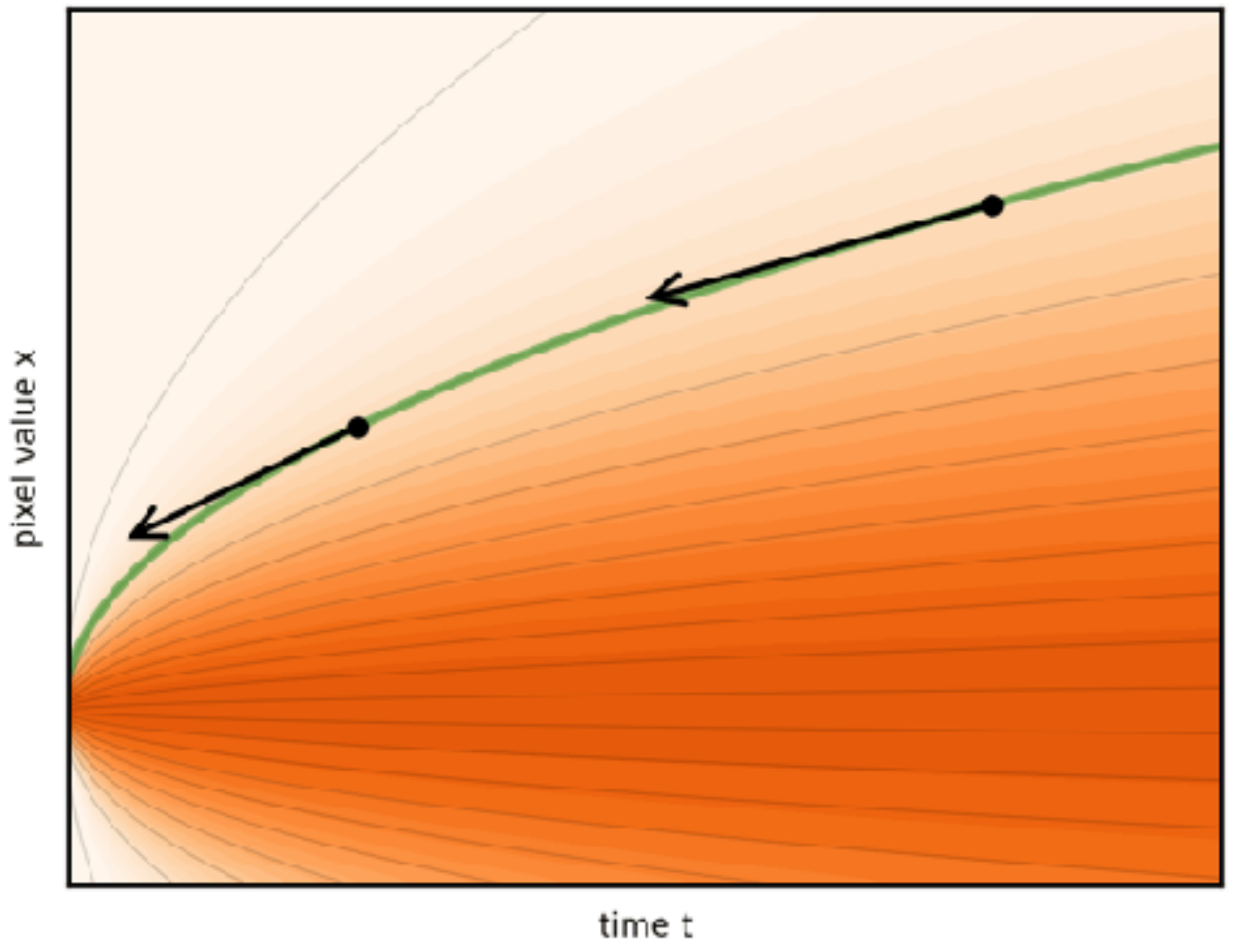
- Sok lehetséges ODE solver / $\sigma(t)$ kombináció:

		VP [49]	VE [49]	iDDPM [37] + DDIM [47]	Ours ("EDM")
Sampling (Section 3)					
ODE solver		Euler	Euler	Euler	2nd order Heun
Time steps	$t_{i<N}$	$1+\frac{i}{N-1}(\epsilon_s-1)$	$\sigma_{\max}^2\left(\sigma_{\min}^2/\sigma_{\max}^2\right)^{\frac{i}{N-1}}$	$u_{\lfloor j_0+\frac{M-1-j_0}{N-1}i+\frac{1}{2}\rfloor}$, where $u_M=0$, $u_{j-1}=\sqrt{\frac{u_j^2+1}{\max(\bar{\alpha}_{j-1}/\bar{\alpha}_j,C_1)}-1}$	$\left({\sigma_{\max}}^{\frac{1}{\rho}}+\frac{i}{N-1}({\sigma_{\min}}^{\frac{1}{\rho}}-{\sigma_{\max}}^{\frac{1}{\rho}})\right)^{\rho}$
Schedule	$\sigma(t)$	$\sqrt{e^{\frac{1}{2}\beta_d t^2+\beta_{\min}t}-1}$	$\sqrt{t}$	t	t
Scaling	$s(t)$	$1/\sqrt{e^{\frac{1}{2}\beta_d t^2+\beta_{\min}t}}$	1	1	1

Elucidating the Design Space of Diffusion-Based Generative Models

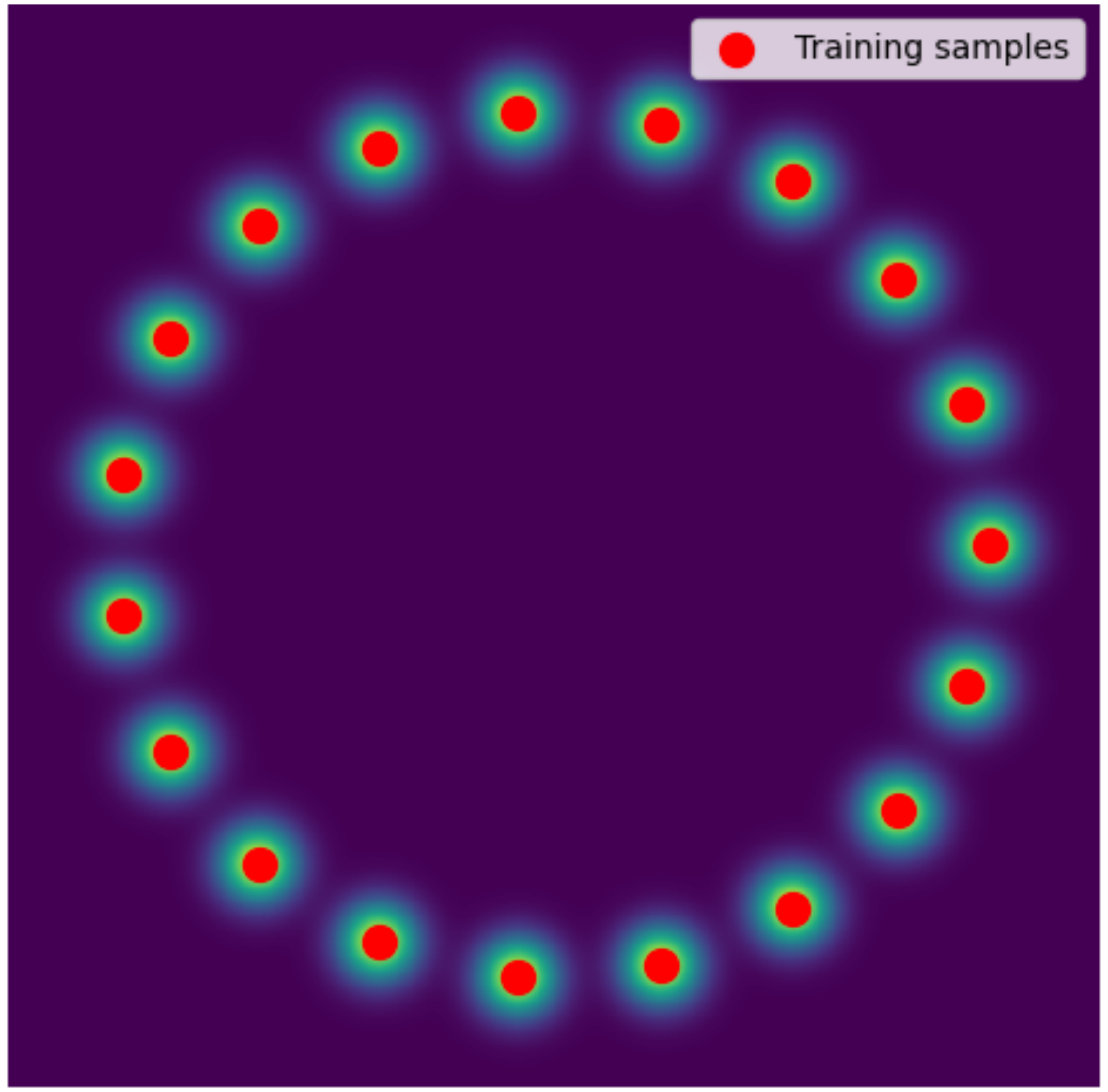
Tero Karras NVIDIA | Miika Aittala NVIDIA | Timo Aila NVIDIA | Samuli Laine NVIDIA



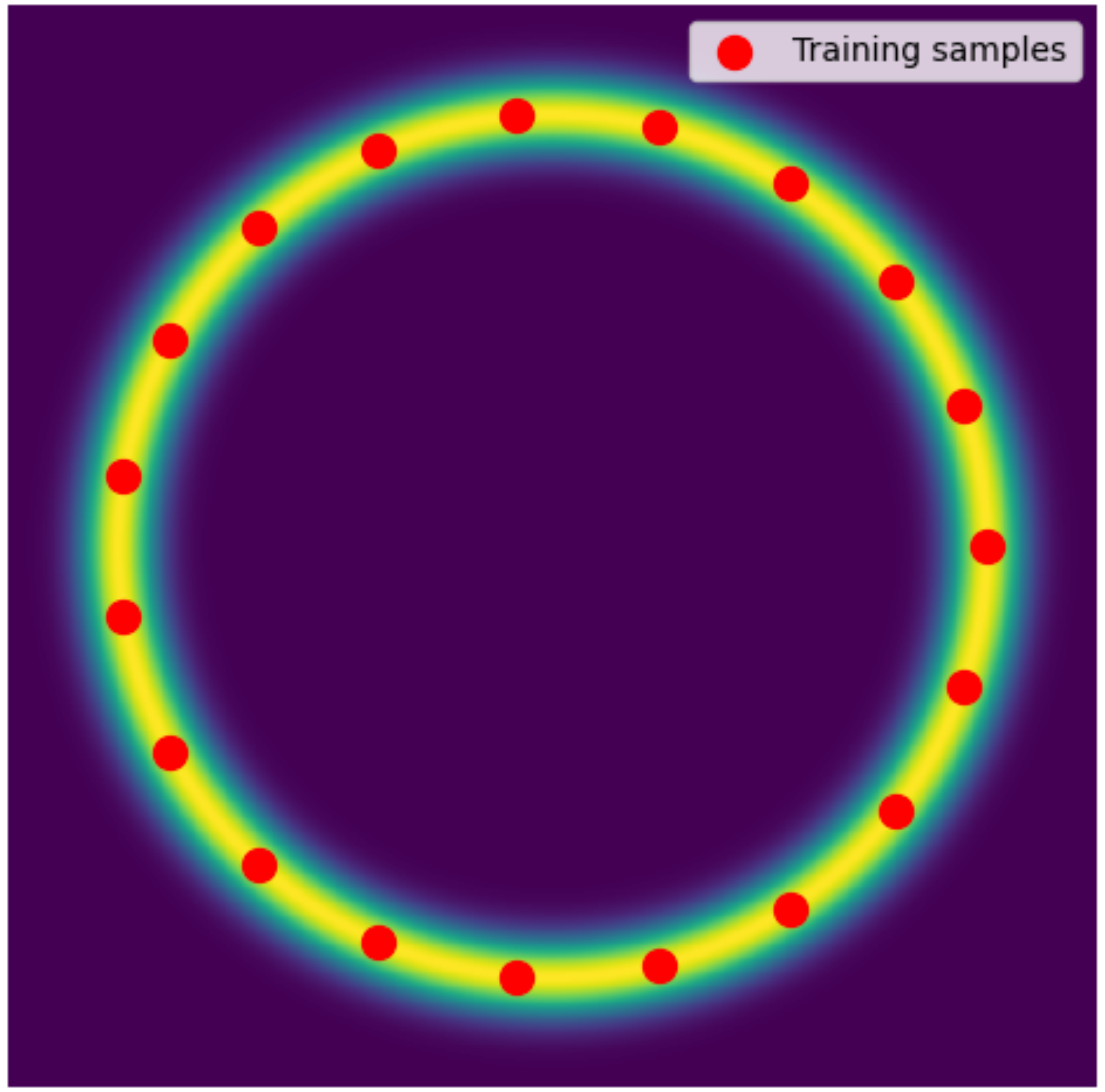


Diffúziós Modellek

Általánosító képesség



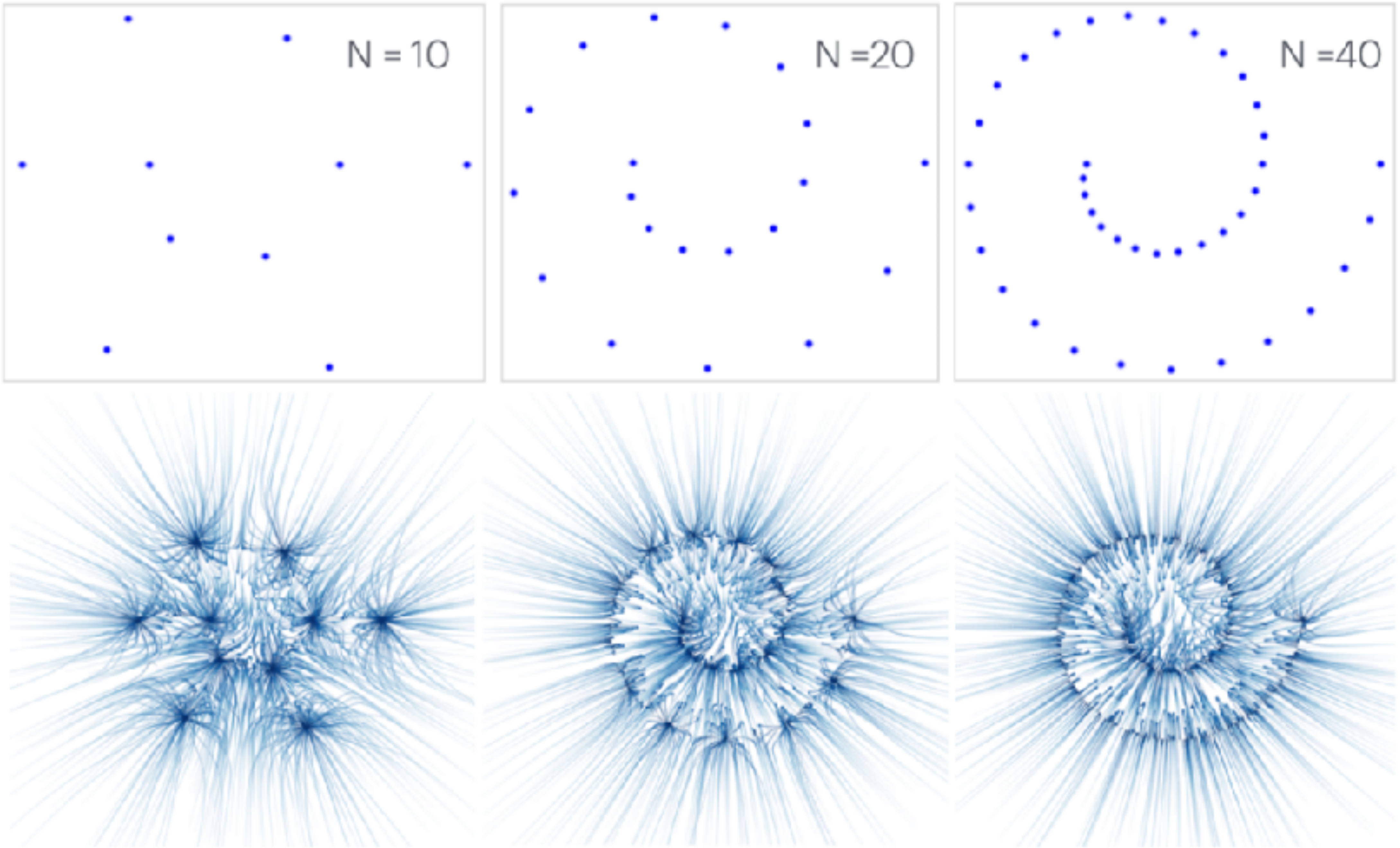
Rossz általánosító képesség (memorizálás)



Jó általánosító képesség

Diffúziós Modellek

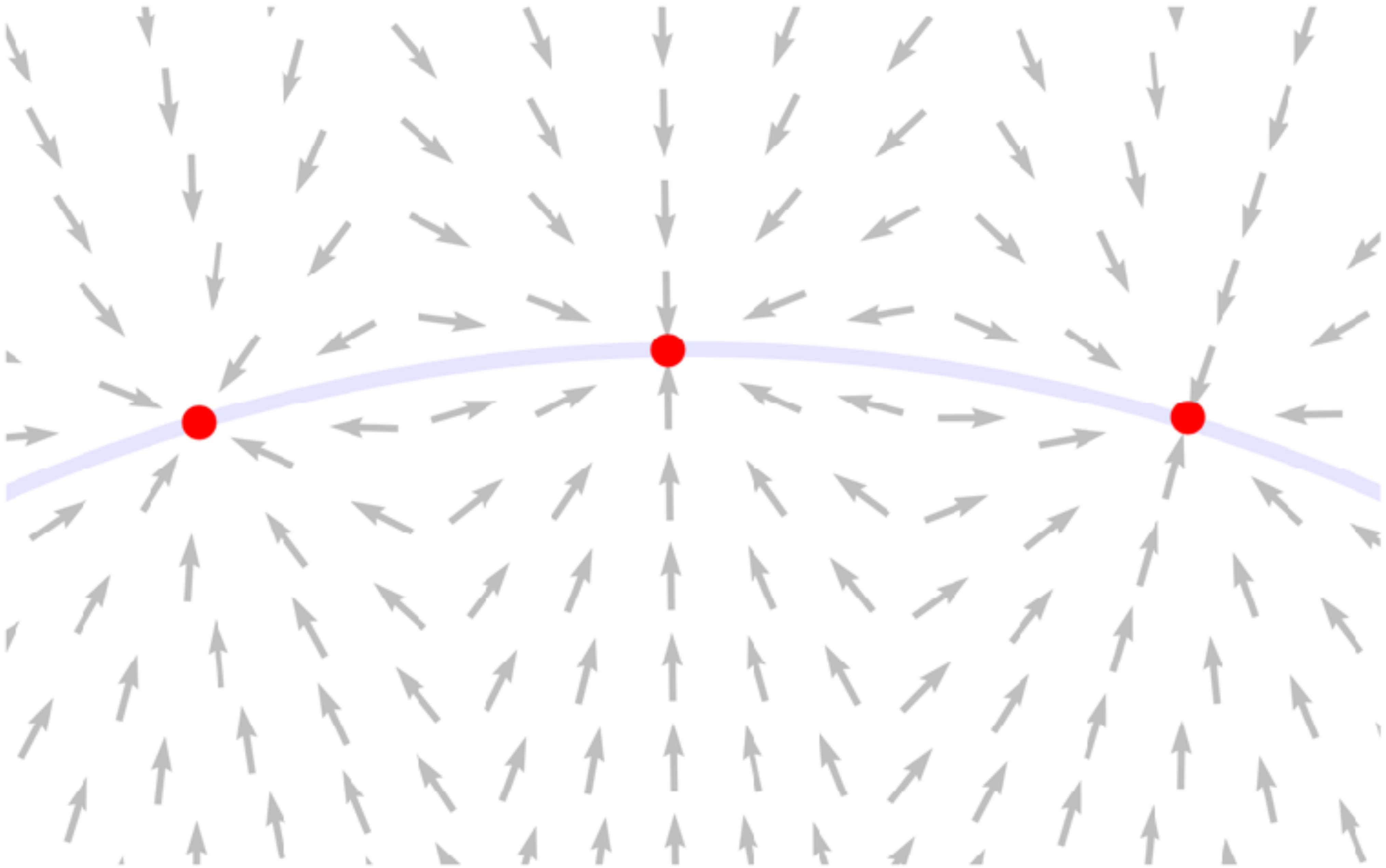
Általánosító képesség



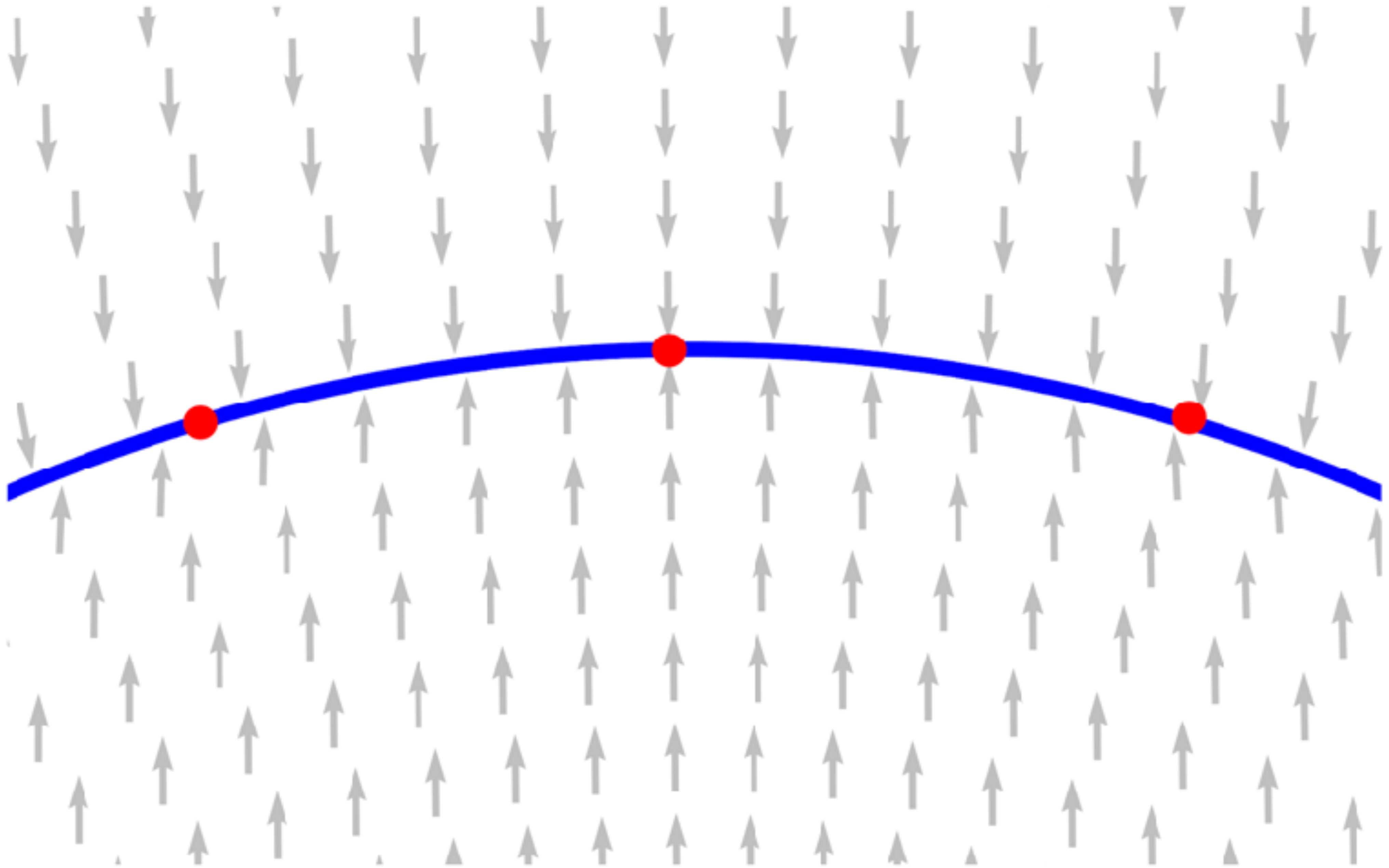
Kevés adatra a determinisztikus mintavétel memorizálja az adathalmazt!
Több adatpontra javul az általánosító képesség…

Diffúziós Modellek

Általánosító képesség



Ideális denoizer



Közelítő (neurális) denoizer

Az általánosítás kulcsa(?): a tökéletlen közelítés elsimítja az adateloszlást!

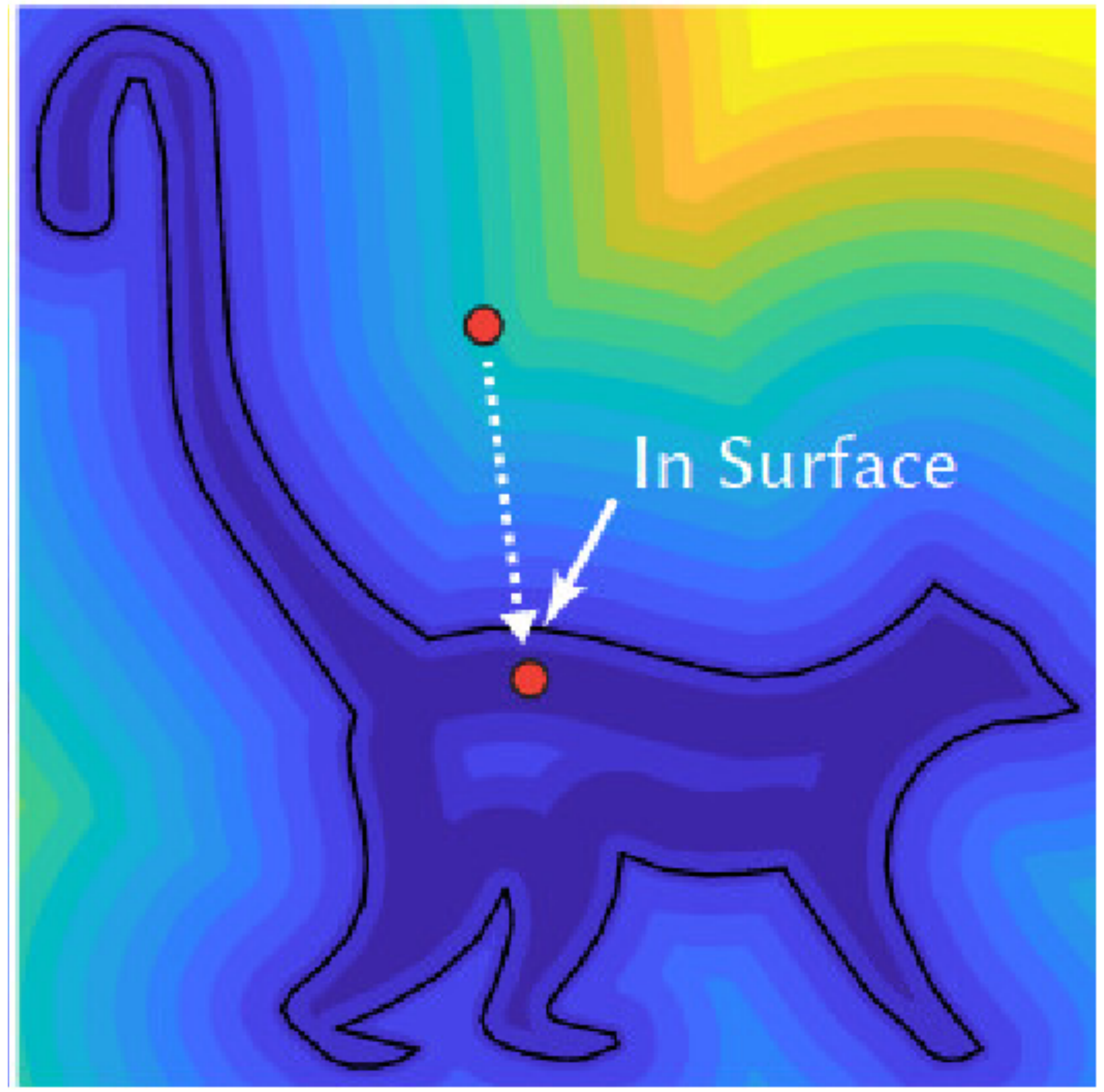
Diffúziós Modellek

Geometriai értelmezés

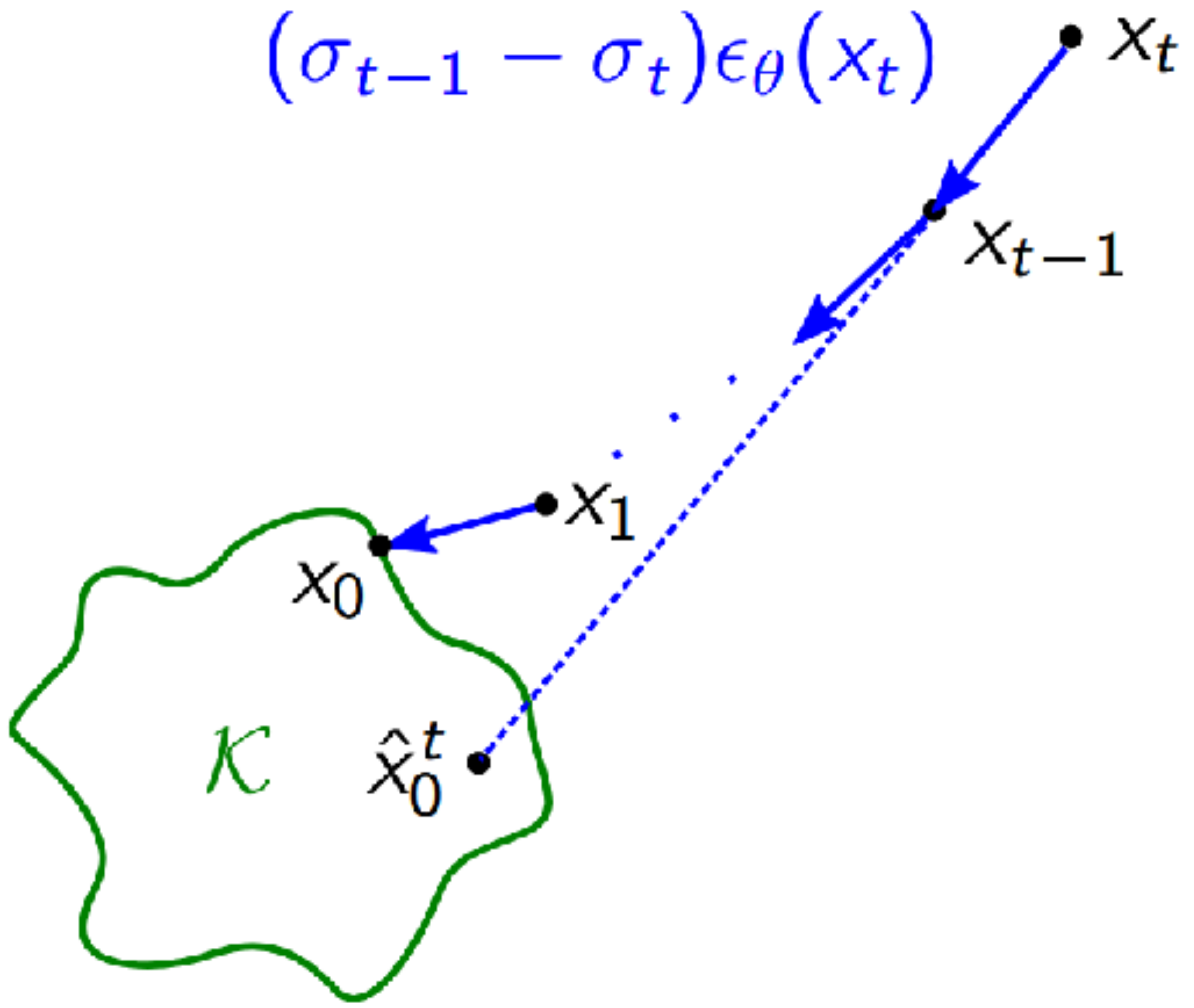
Interpreting and Improving Diffusion Models from an Optimization Perspective

Frank Permenter [*1] Chenyang Yuan [*1]

https://www.chenyang.co/diffusion.html

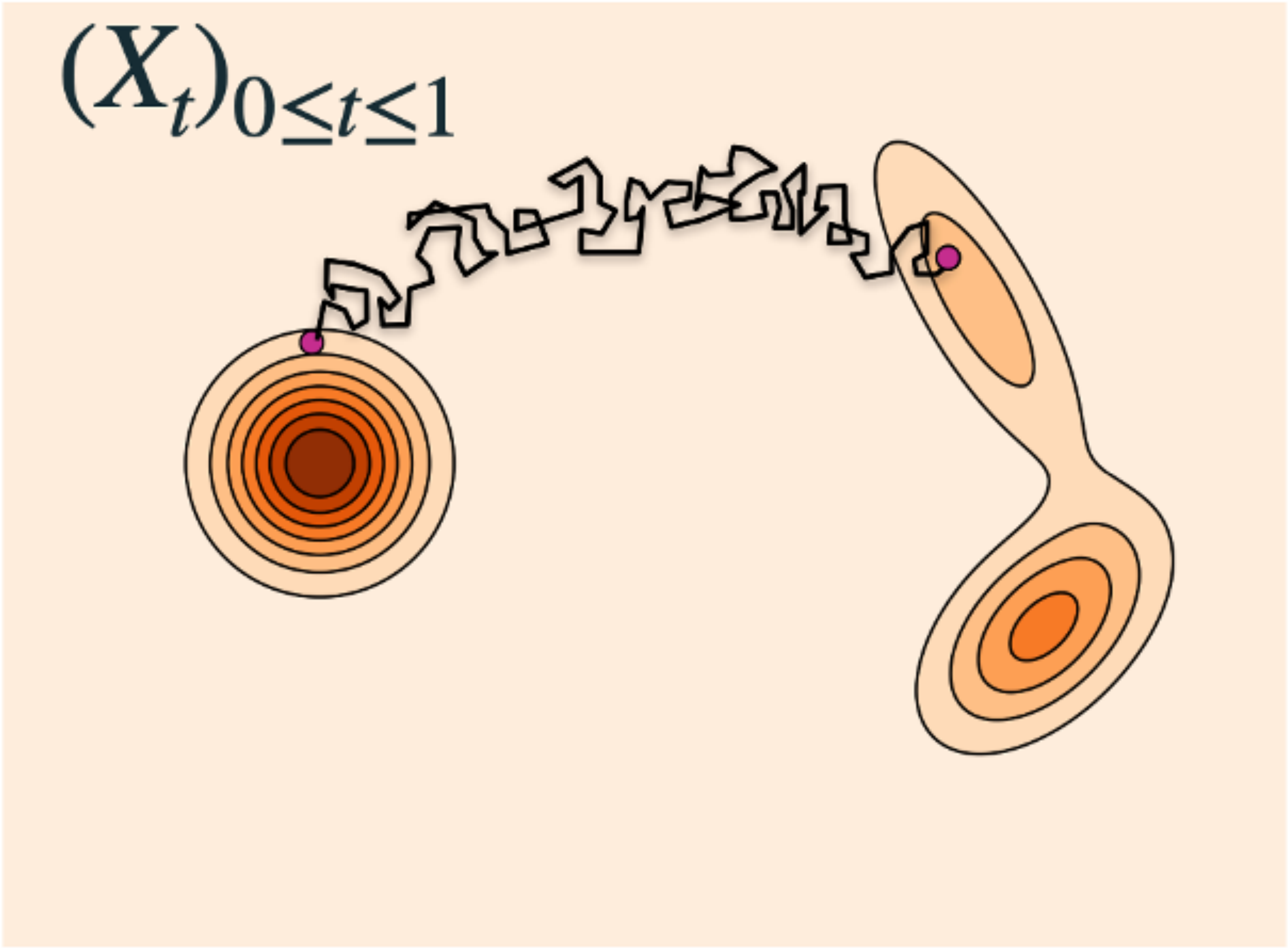


Score függvény ≈ simított távolság(négyzet) az adatsokaságtól

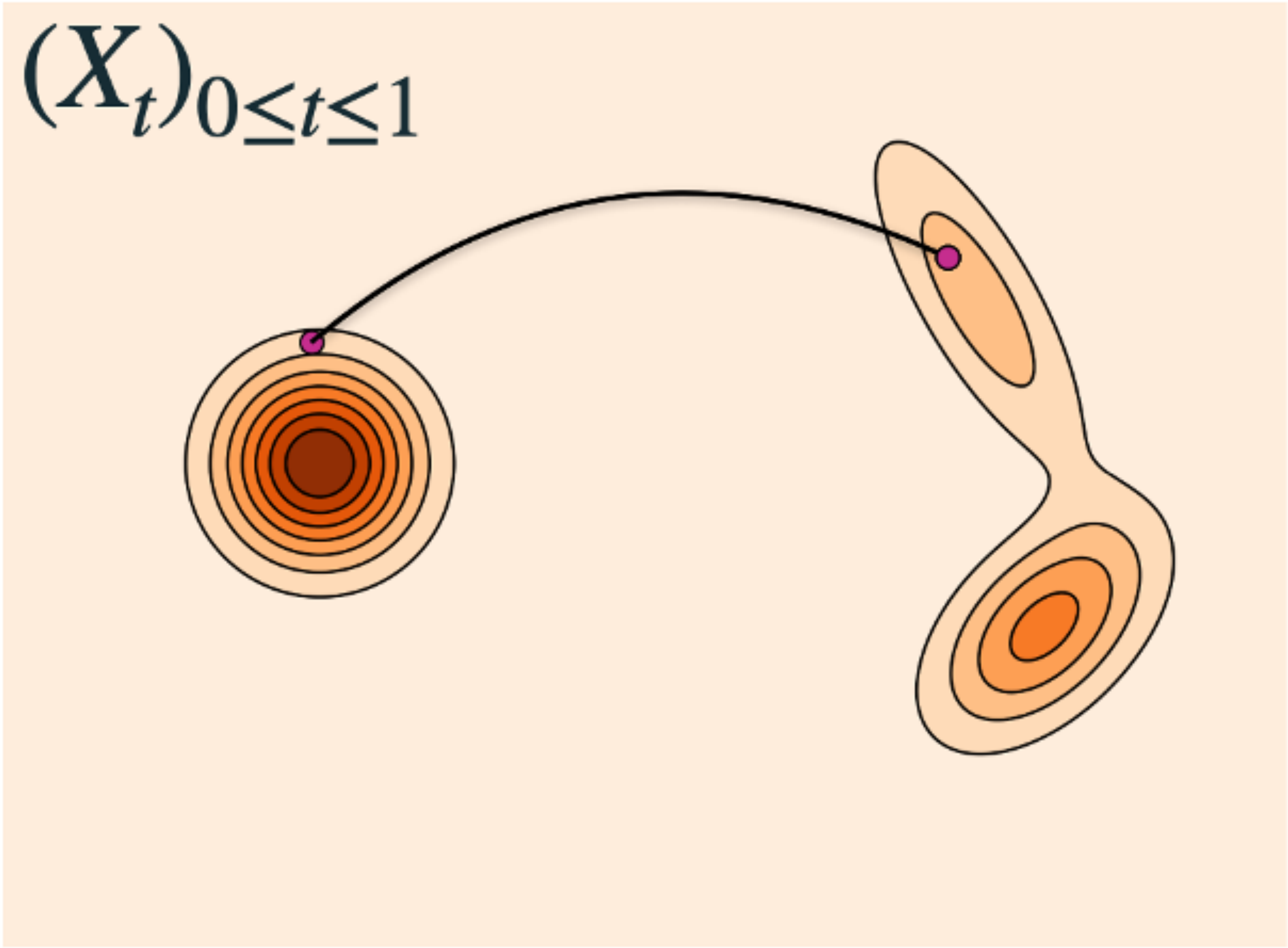


Diffúziós generálás (pl. DDIM) ≈ gradient descent a simított távolságfüggvényre

Folyamillesztés

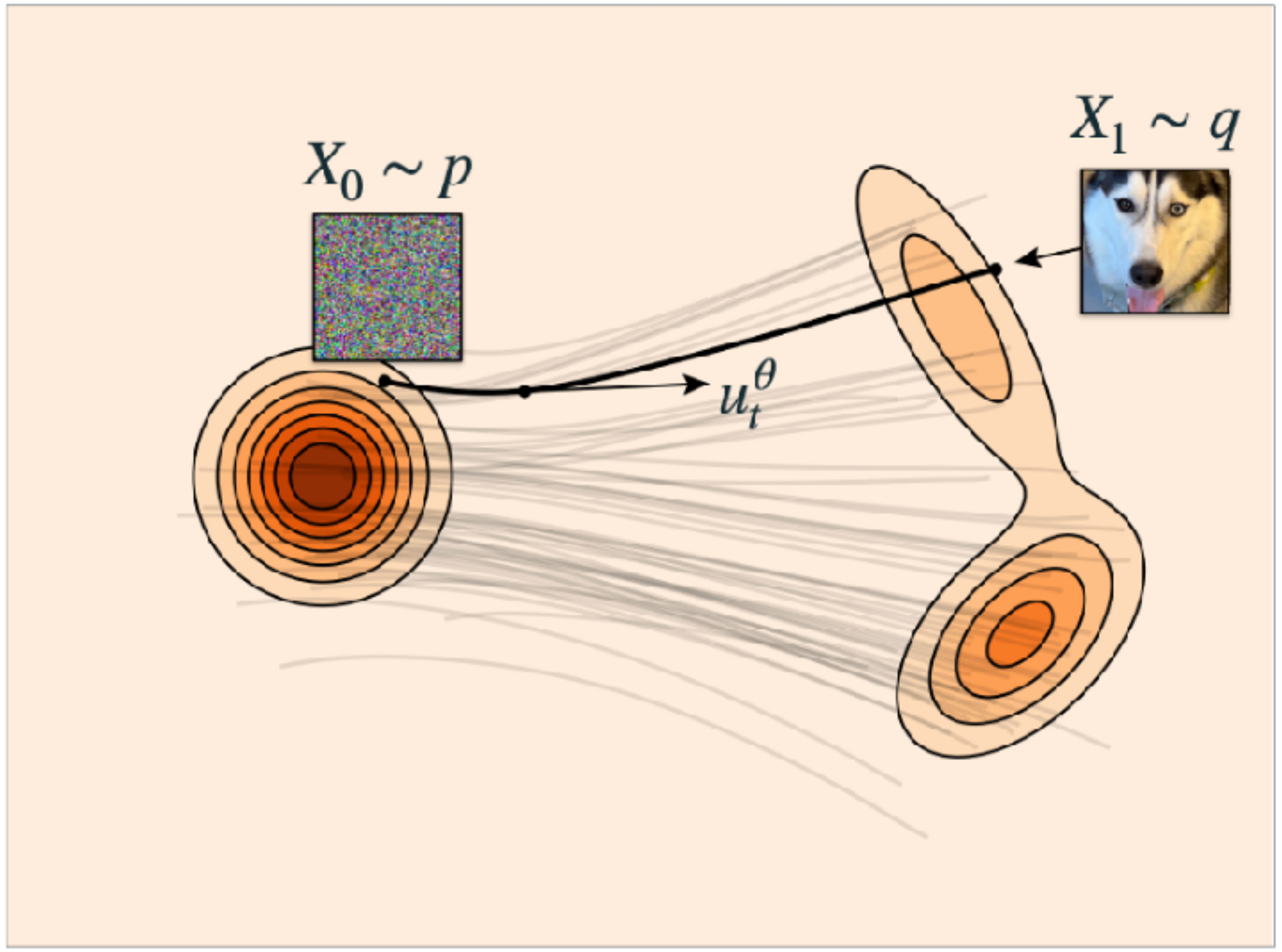


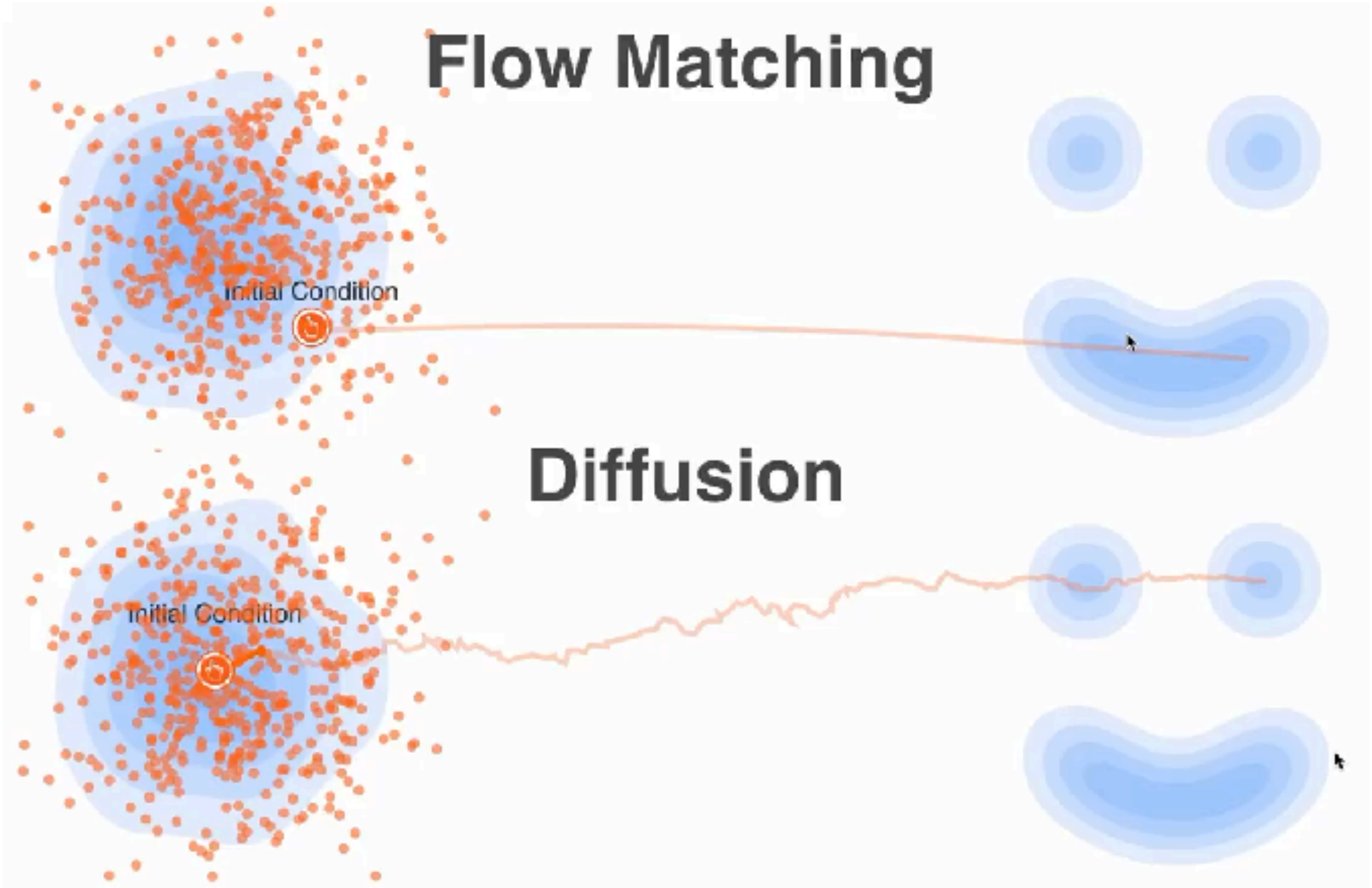
Diffúzió (SDE)



Folyam (ODE)

Folyamillesztés

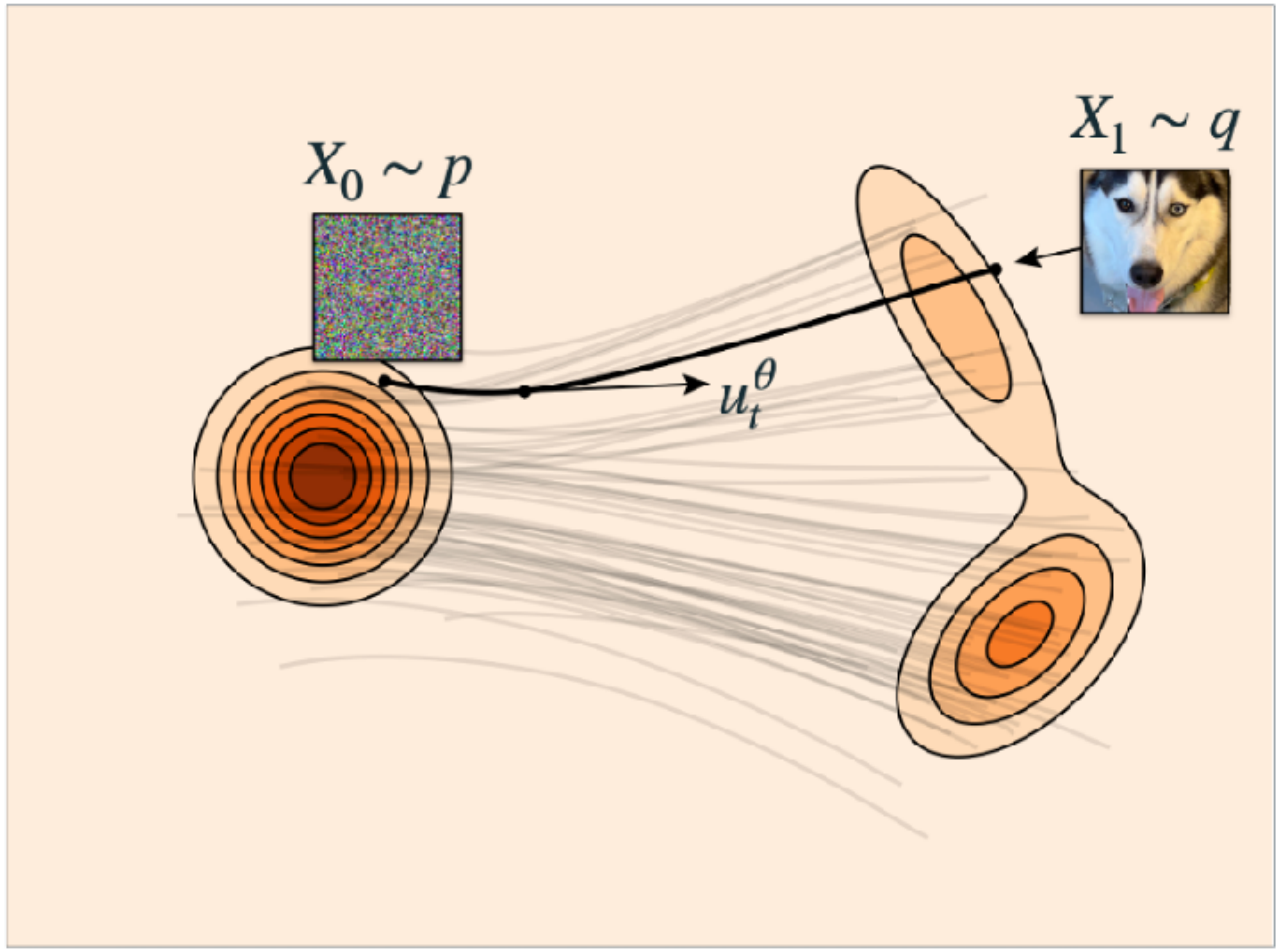


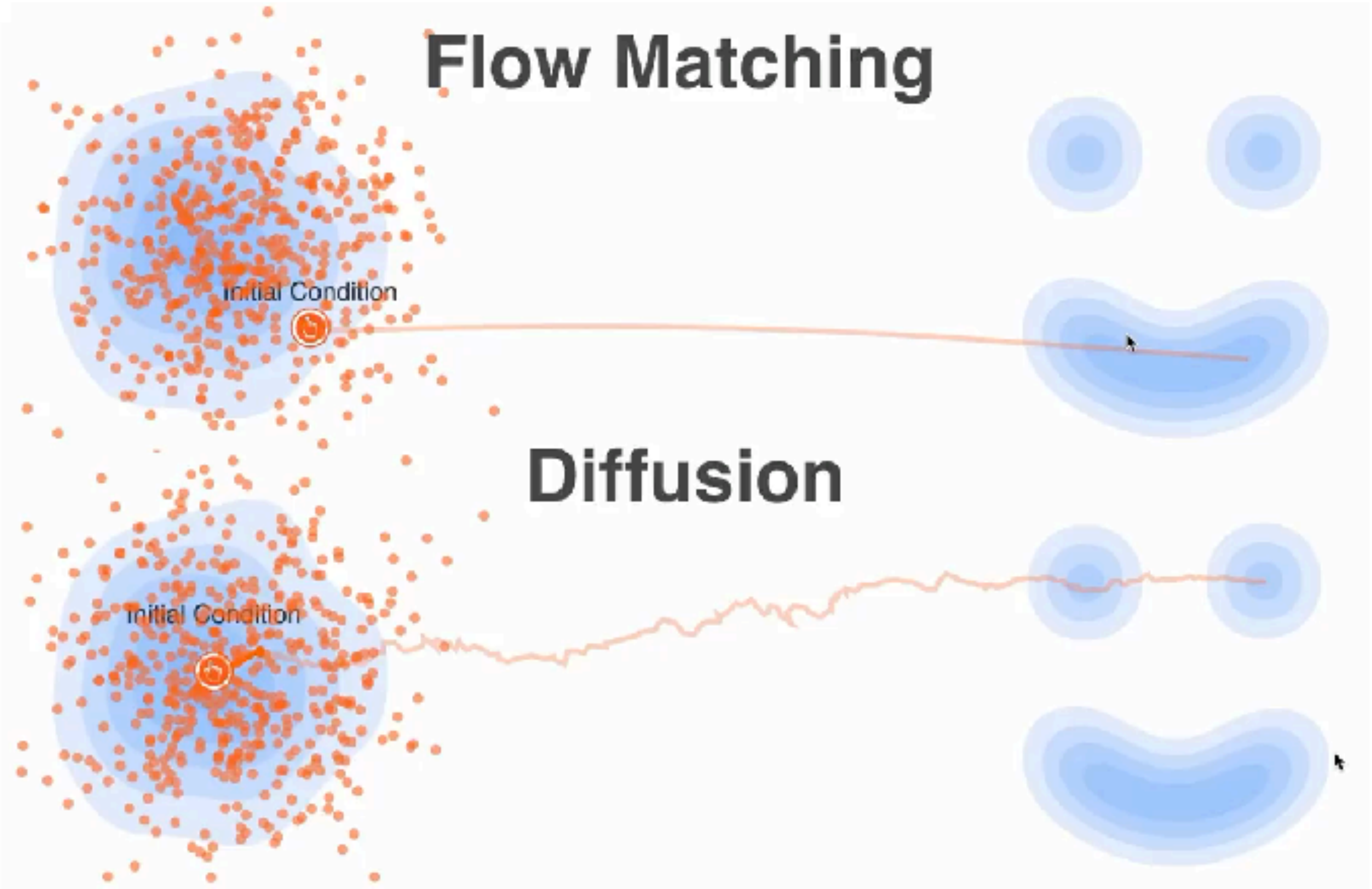


https://alechelbling.com/Diffusion-Explorer/

A diffúziós modellek valójában egy ODE/folyam, azaz egy *vektormező* alapján generálnak
Ötlet: definiáljuk a feladatot egy **vektormező** illesztéseként!

Folyamillesztés





https://alechelbling.com/Diffusion-Explorer/

A diffúziós modellek valójában egy ODE/folyam, azaz egy *vektormező* alapján generálnak
Ötlet: definiáljuk a feladatot egy **vektormező** illesztéseként!

Folyamillesztés
Valószínűségi folyam

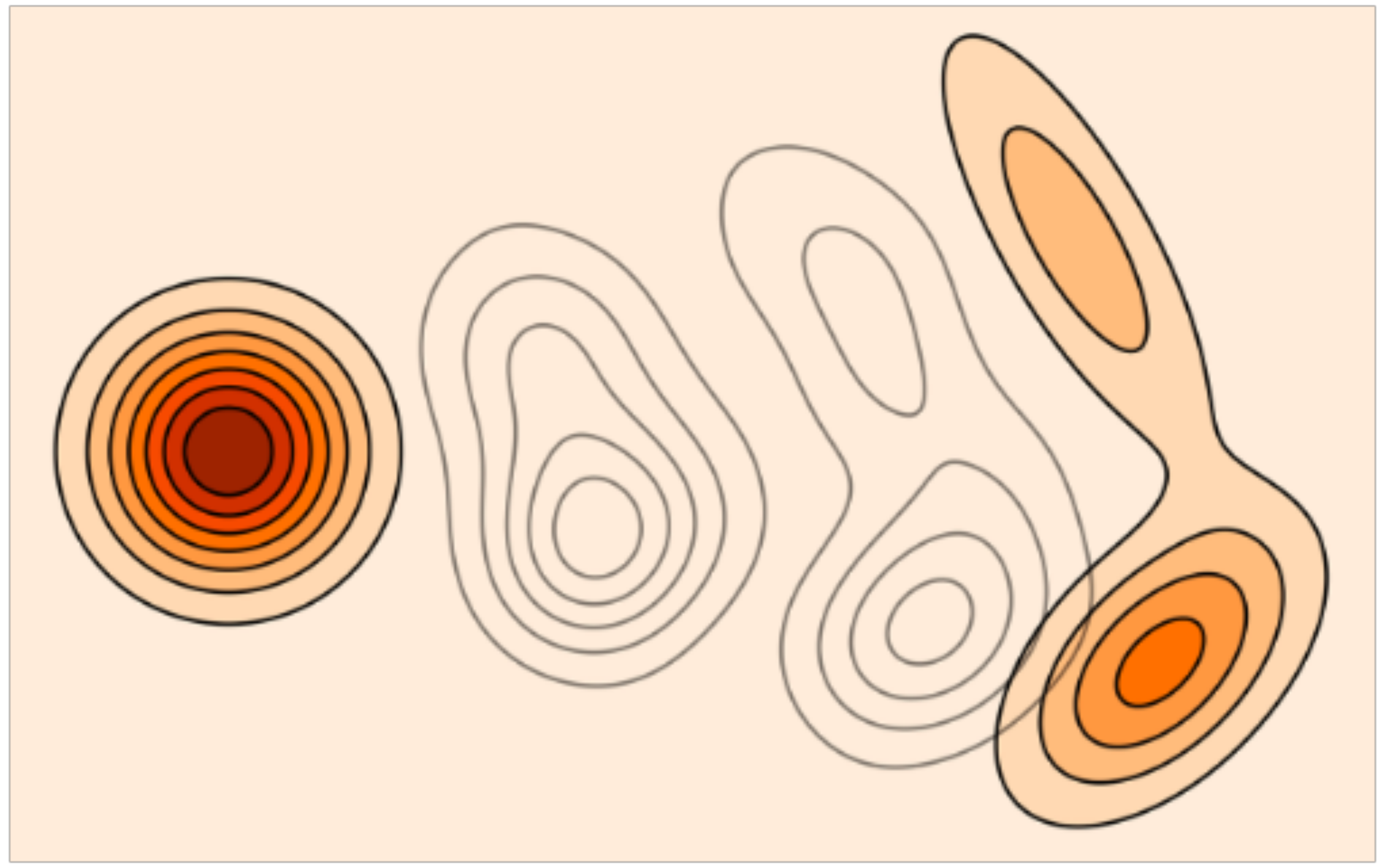
- Cél: interpolálni a p_0 Gaussi eloszlás (zaj) és a p_1 adat eloszlás között – p_t eloszlás-család $(0 \leq t \leq 1)$

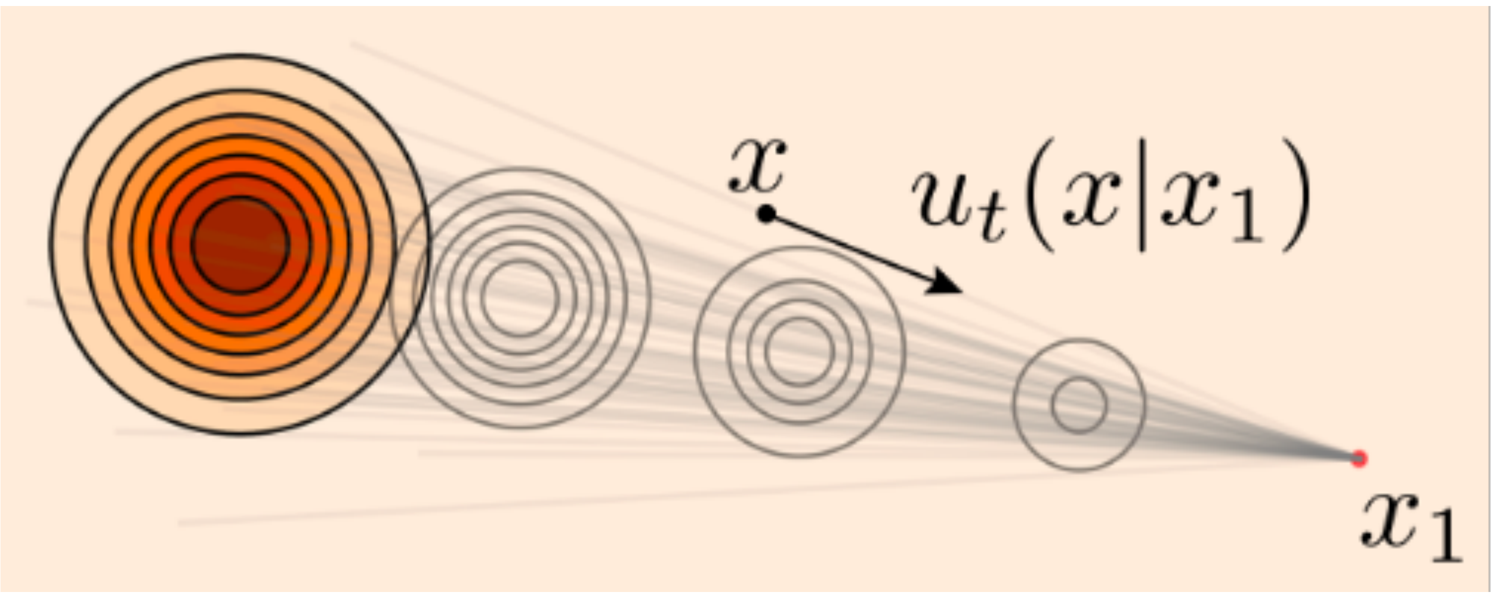
- Adott $x_0 \sim p_0$, $x_1 \sim p_1$ mintákra:

$$(1-t)x_0 + tx_1 \sim p_t$$

- Tegyük fel, hogy egyetlen x_1 adatpontunk van – haladjon minden x_t pont egyenletes sebességgel x_1 irányába (**<u>feltételes vektormező</u>**):

$$v(x_t \mid x_1) = \frac{dx_t}{dt} = x_1 - x_0$$





Folyamillesztés

Marginális és feltételes folyamillesztés

- Komplexebb p_1 adateloszlásra a **marginális folyamot** kapjuk:

$$v(x_t) = \int v(x_t \mid x) \cdot p_1(x) dx$$

- Közelítése neurális hálóval — **folyamillesztés (flow matching)**:

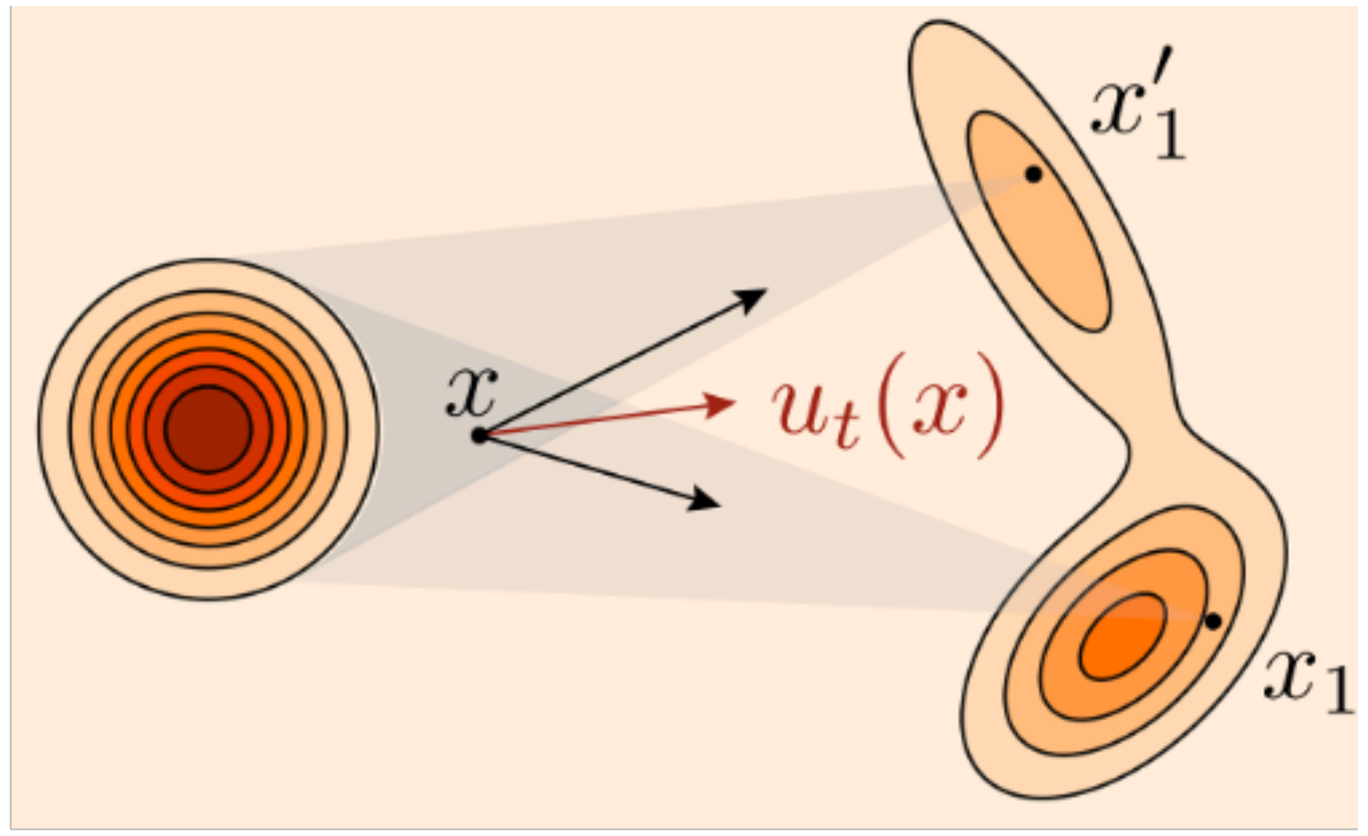
$$L_{FM}(\theta) = \mathbb{E}_{x_0 \sim p_0, x_1 \sim p_1, t \sim U(0,1)}[\, \| \, v_\theta(x_t, t) - v(x_t) \, \|^2] \text{ ☠}$$

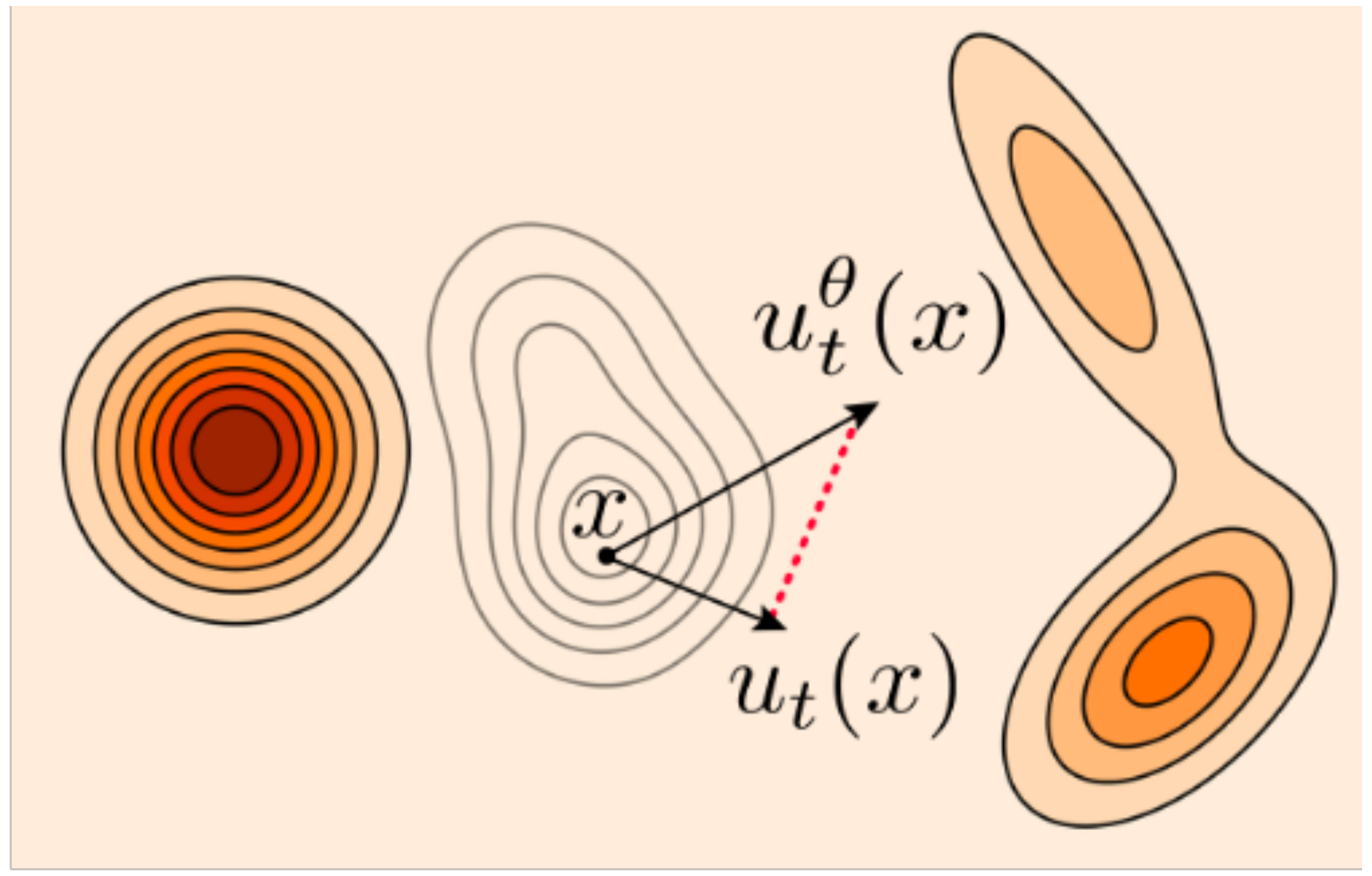
- A marginális folyam helyettesíthető a feltételes folyammal!

$$L_{CFM}(\theta) = \mathbb{E}_{x_0 \sim p_0, x_1 \sim p_1, t \sim U(0,1)}[\, \| \, v_\theta(x_t, t) - v(x_t \mid x_1) \, \|^2]$$

$$= \boxed{\mathbb{E}_{x_0 \sim p_0, x_1 \sim p_1, t \sim U(0,1)}[\, \| \, v_\theta((1-t)x_0 + t x_1, t) - (x_1 - x_0) \, \|^2]}$$

  - A két loss lényegében ekvivalens: $L_{CFM} = L_{FM} + \text{const}$.





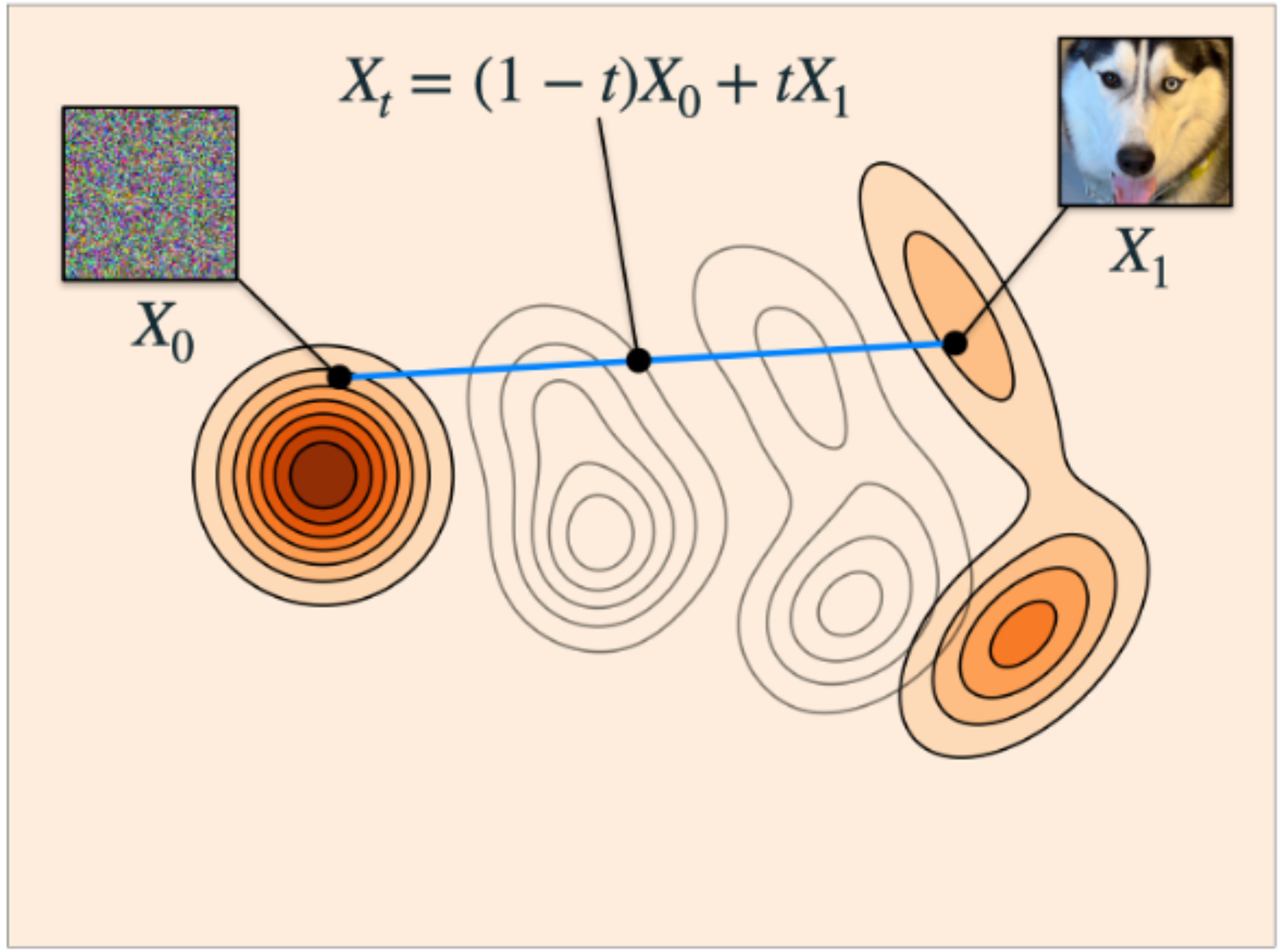
Folyamillesztés

Tanítás

- **<u>Folyamillesztés (flow matching)</u>** előnyei:
  - **<u>Nagyon egyszerű tanítás (regresszió)!</u>**
  - Determinisztikus generálás (ODE)!
  - Rugalmasság! (Tetszőleges összerendelés, cél/forrás eloszlás, interpoláció, stb.)

```
for _ in range(10000):
    x_1  = Tensor(make_moons(256, noise=0.05)[0])
    x_0  = torch.randn_like(x_1)
    t    = torch.rand(len(x_1), 1)
    x_t  = (1 - t) * x_0 + t * x_1
    dx_t = x_1 - x_0
    optimizer.zero_grad()
    loss_fn(flow(x_t, t), dx_t).backward()
    optimizer.step()
```

Igazi Pytorch, nem pszeudokód!

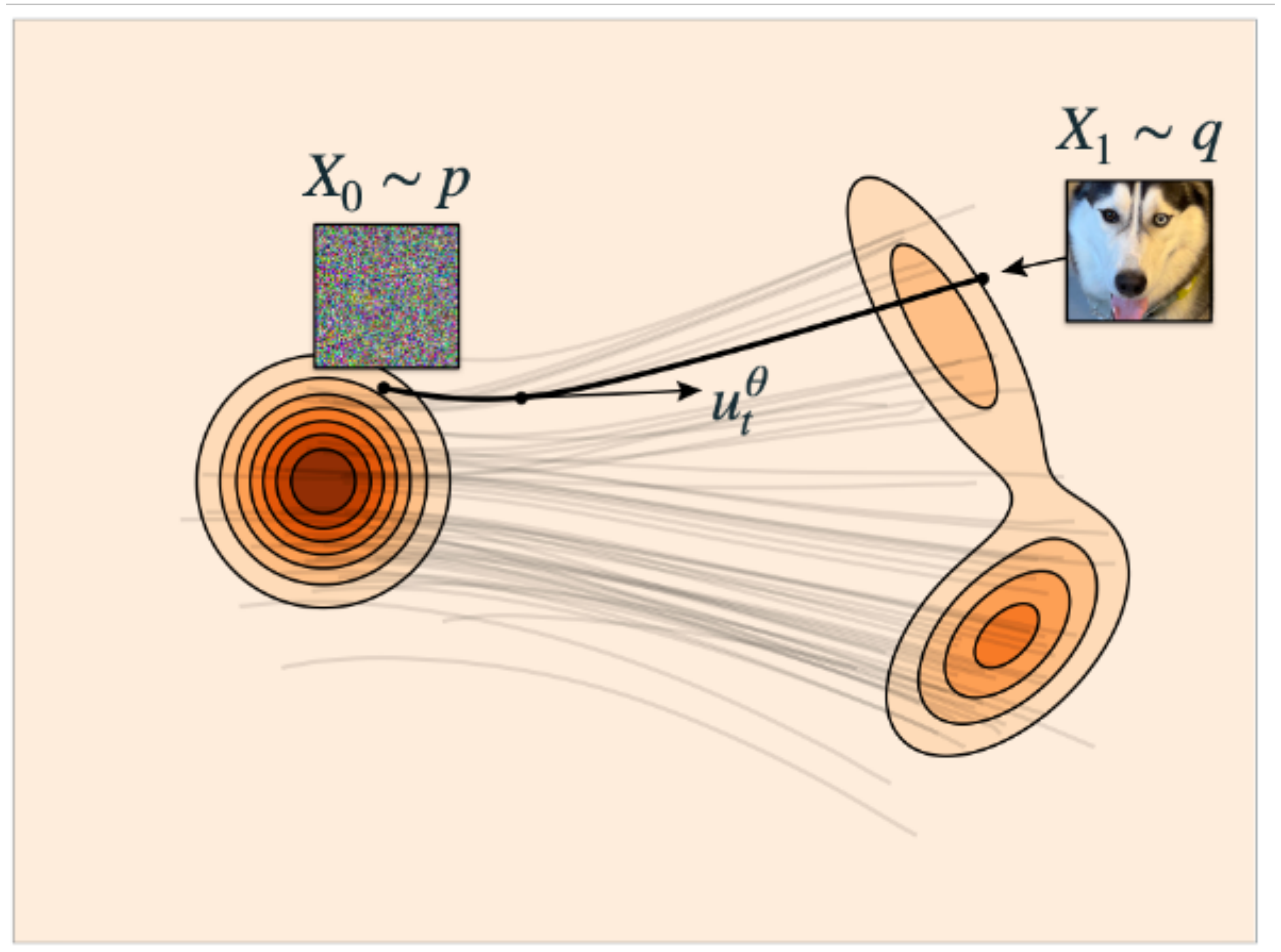


$$\mathbb{E}_{t,X_0,X_1}\left\| u_t^\theta(X_t) - (X_1 - X_0) \right\|^2$$

Folyamillesztés

Generálás

- **<u>Folyamillesztés (flow matching)</u>** előnyei:
  - Nagyon egyszerű tanítás (regresszió)!
  - **<u>Determinisztikus generálás (ODE)!</u>**
  - Rugalmasság! (Tetszőleges összerendelés, cél/forrás eloszlás, interpoláció, stb.)

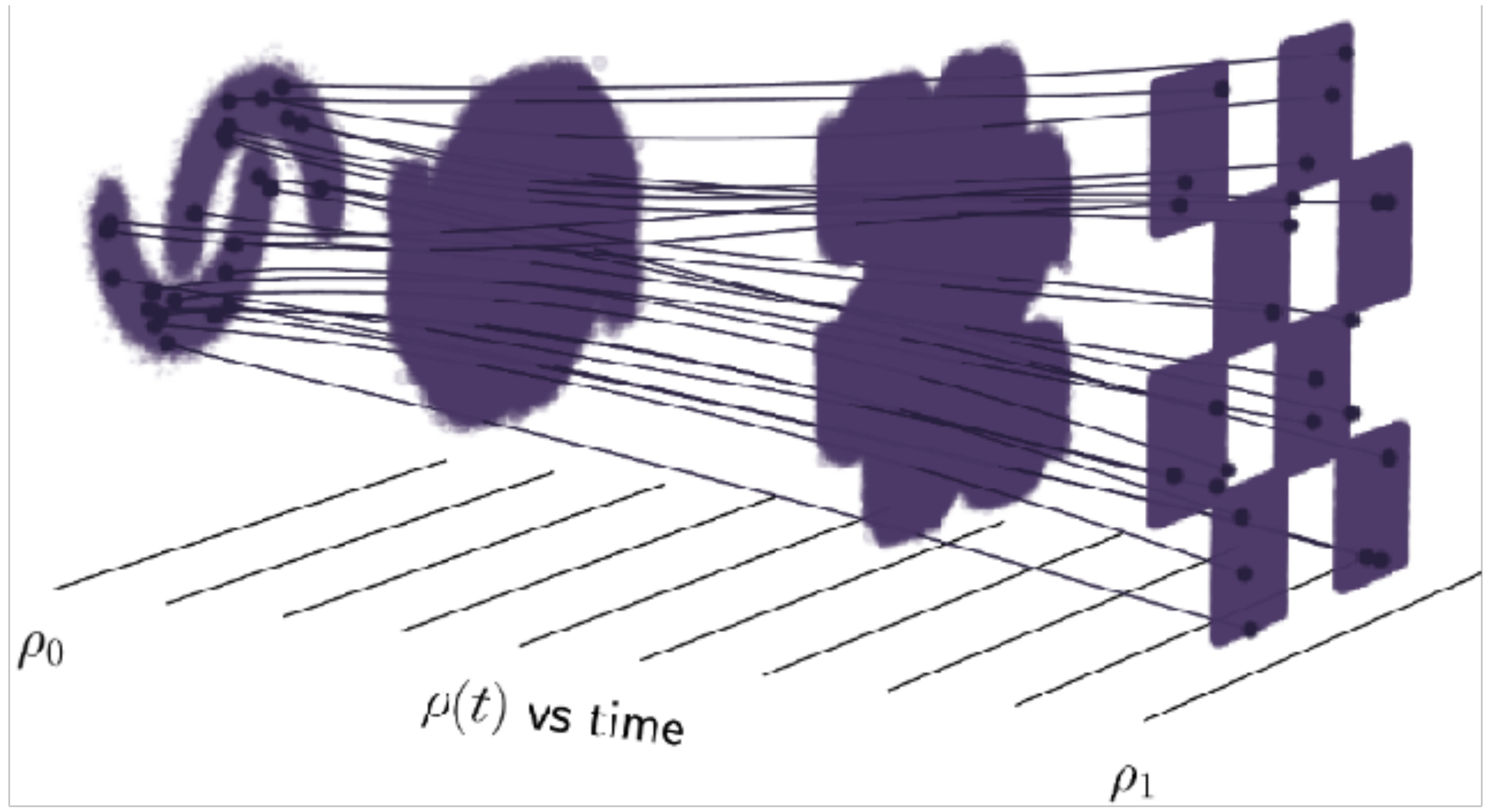


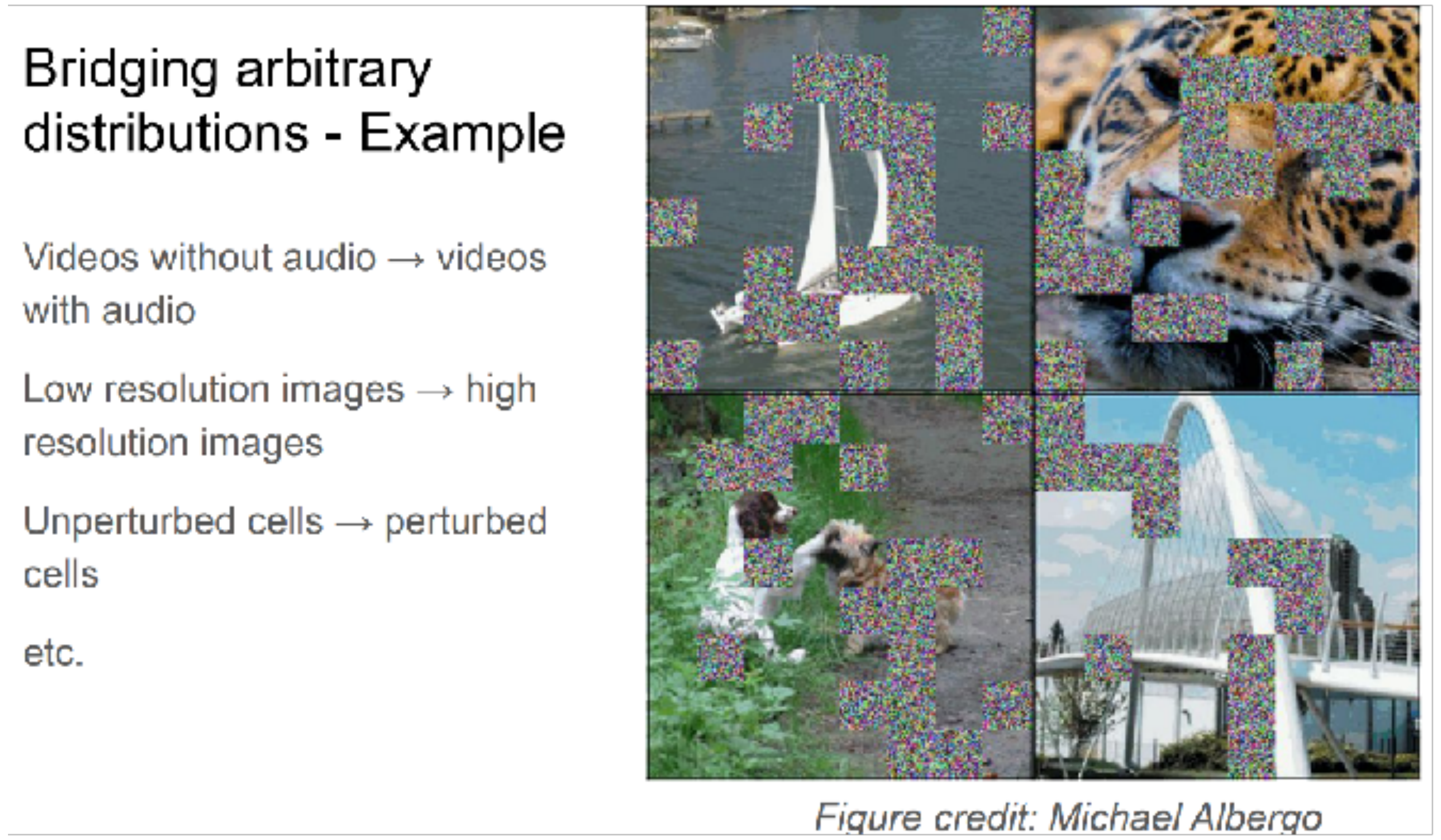
<u>Bármilyen ODE solver (integrátor) használható!</u>

Folyamillesztés

Rugalmasság

- **<u>Folyamillesztés (flow matching)</u>** előnyei:
    - Nagyon egyszerű tanítás (regresszió)!
    - Determinisztikus generálás (ODE)!
    - **<u>Rugalmasság!</u>** (Tetszőleges összerendelés, cél/ forrás eloszlás, interpoláció, stb.)

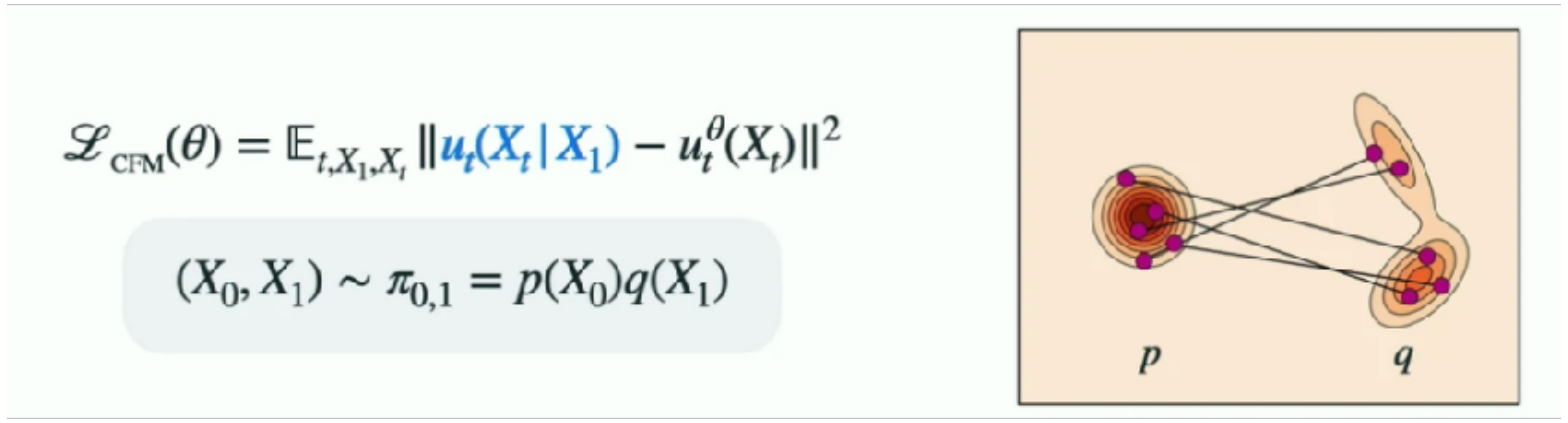




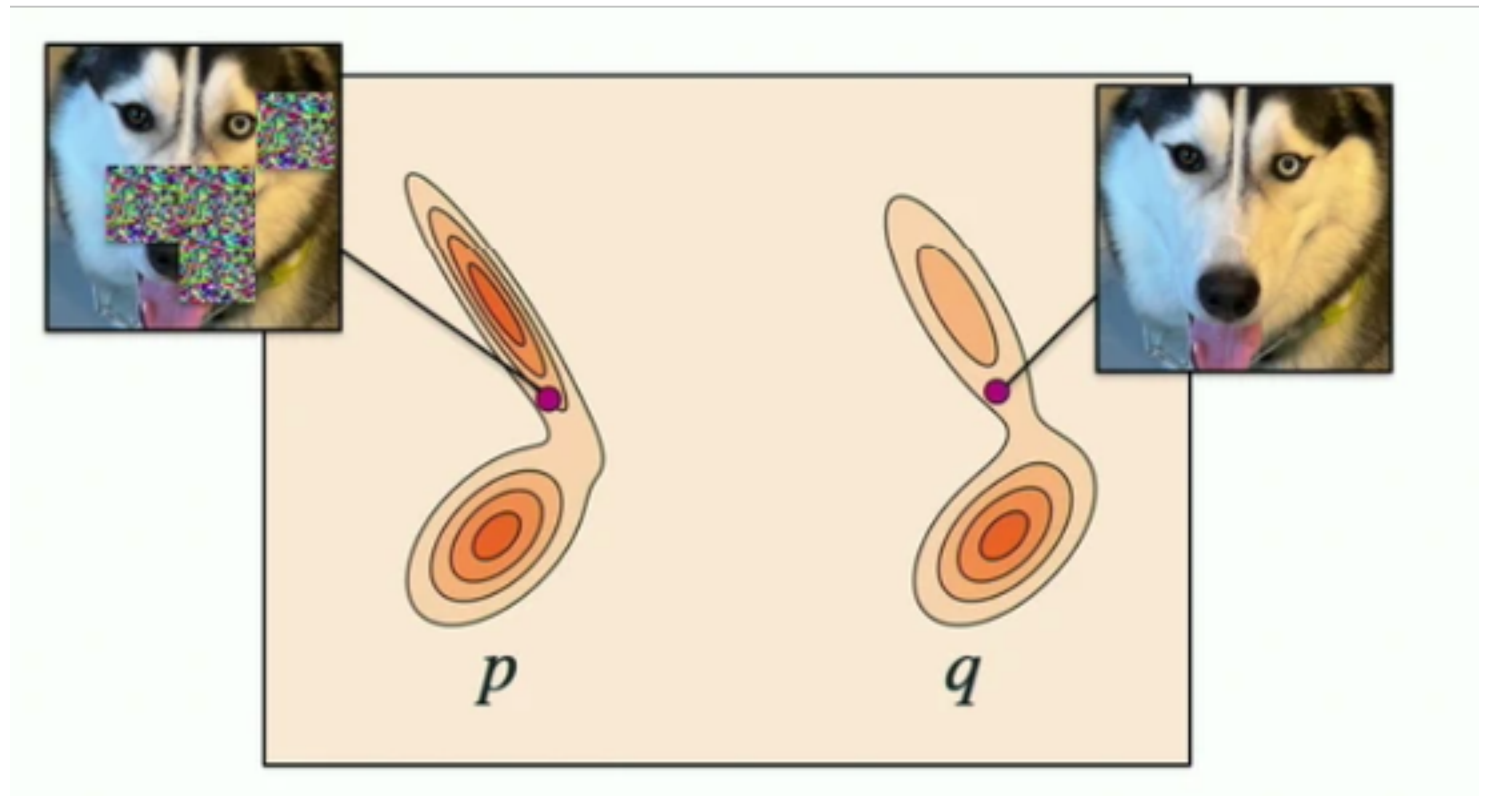
<u>Tetszőleges eloszlások között interpolálhatunk!</u>

Folyamillesztés

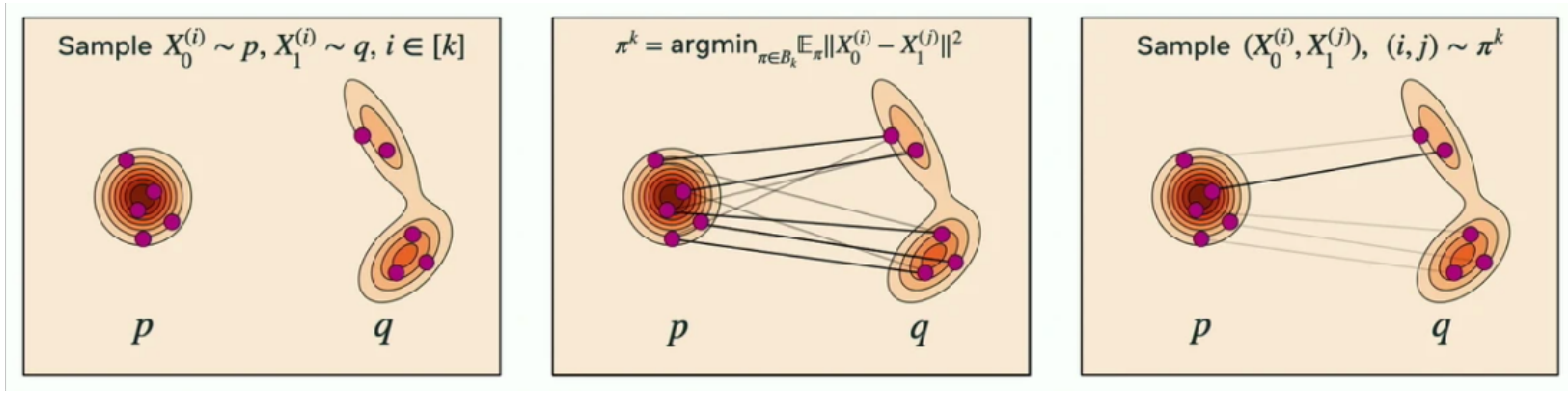
Adatok összerendelése



<u>Független összerendelés</u>



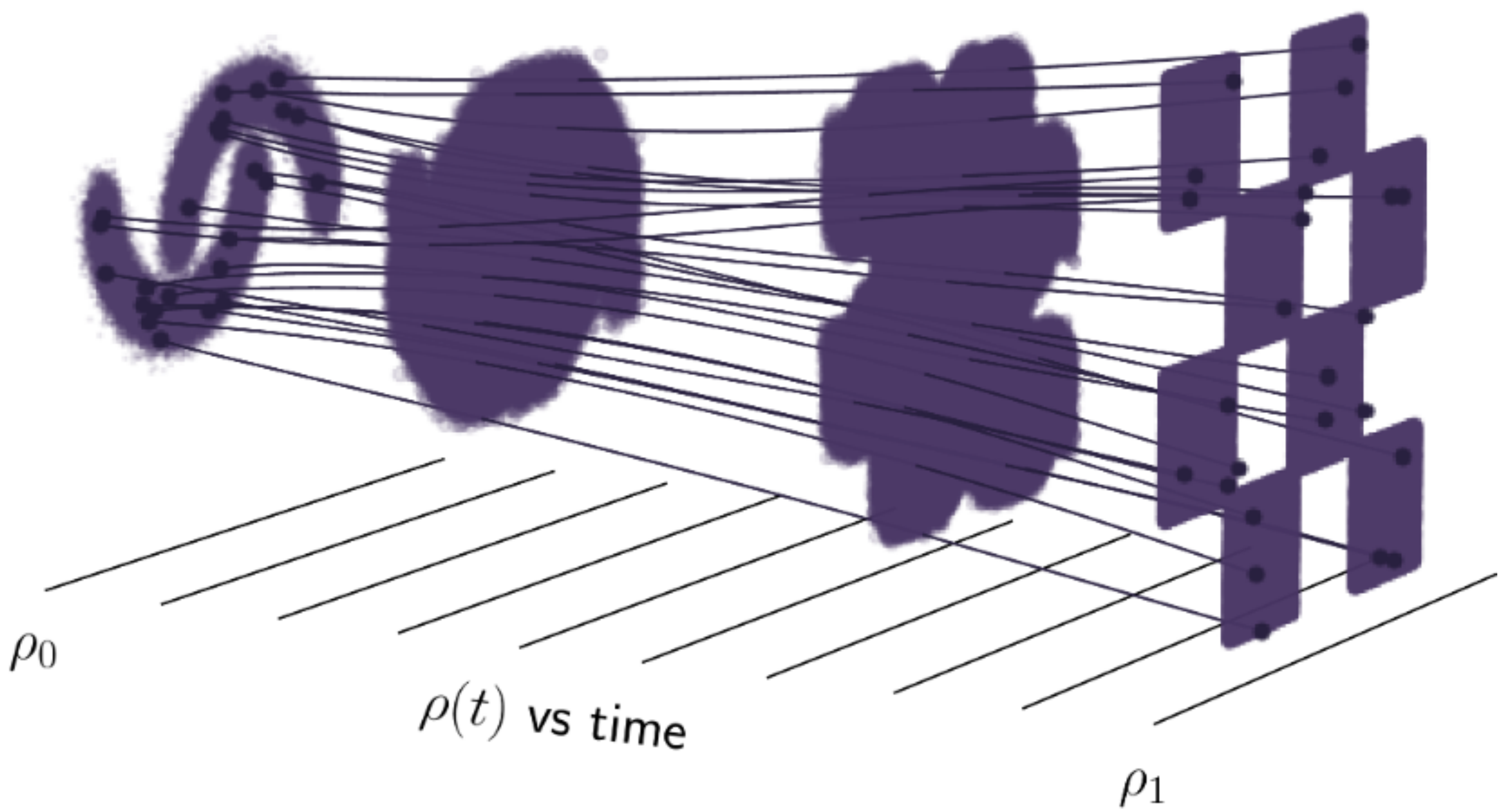
<u>Párosított adatok</u>



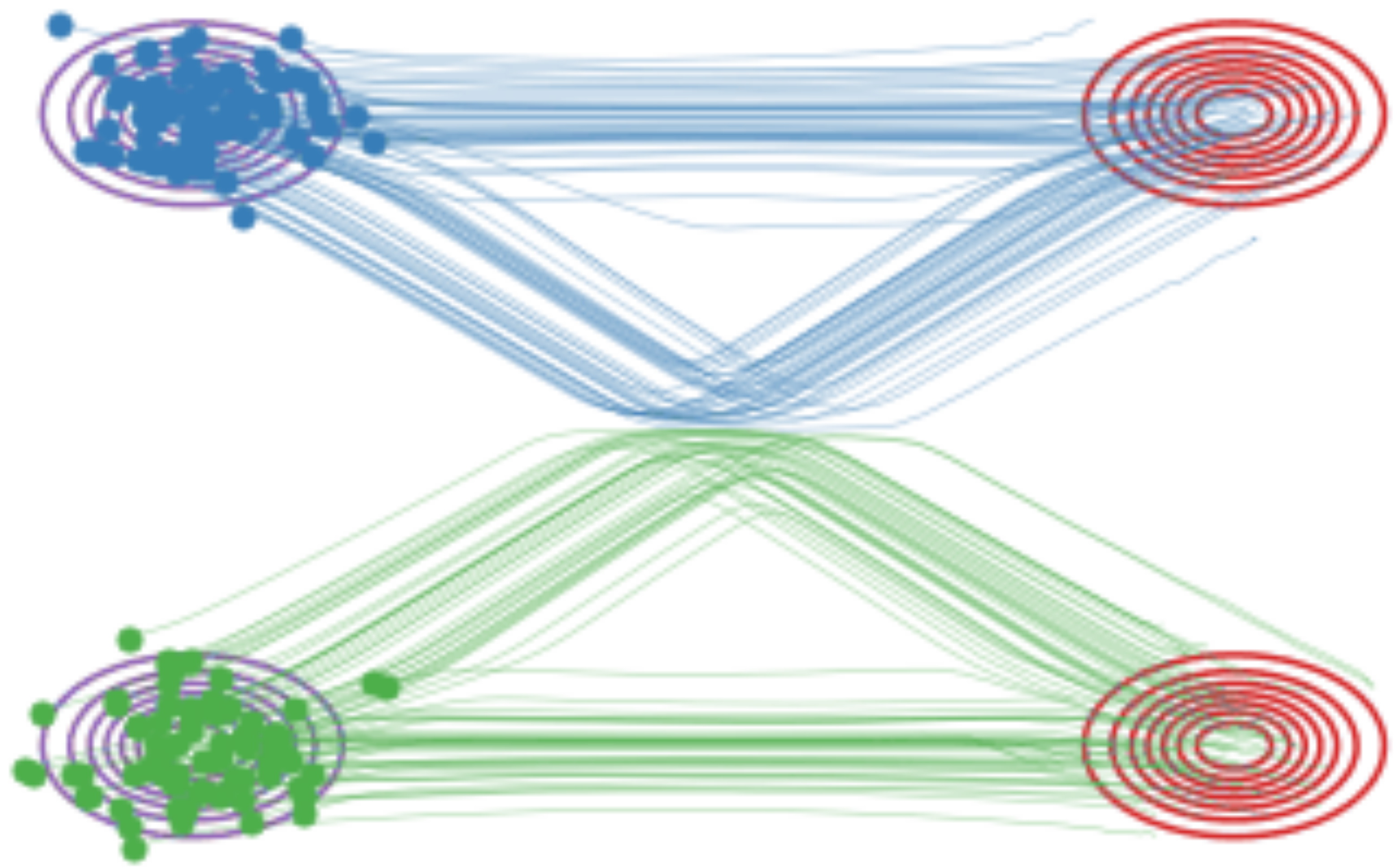
<u>(Mini-Batch) Optimal Transport</u>

Folyamillesztés

Kapcsolódó formalizmusok



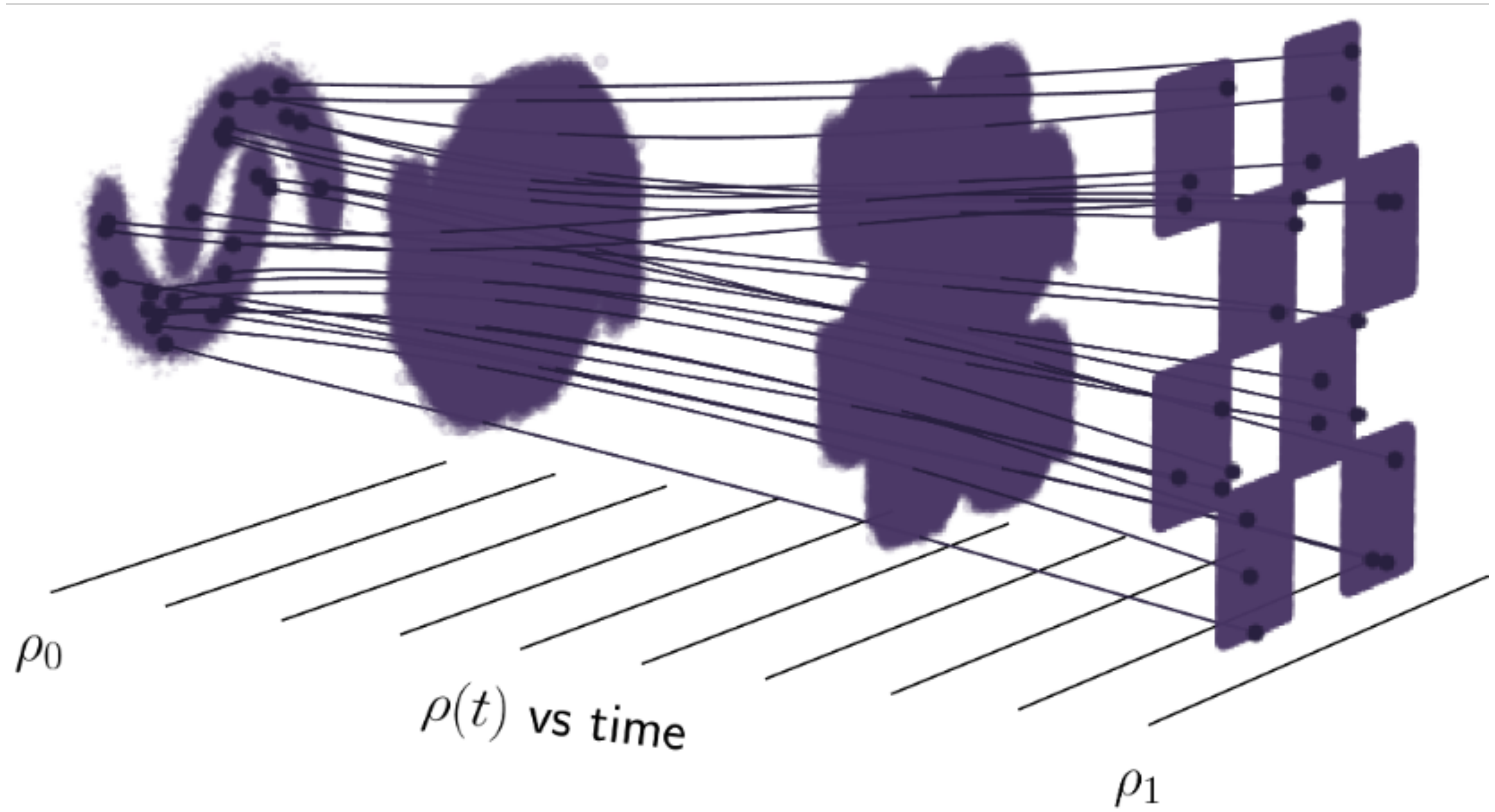
Sztochasztikus Interpolánsok



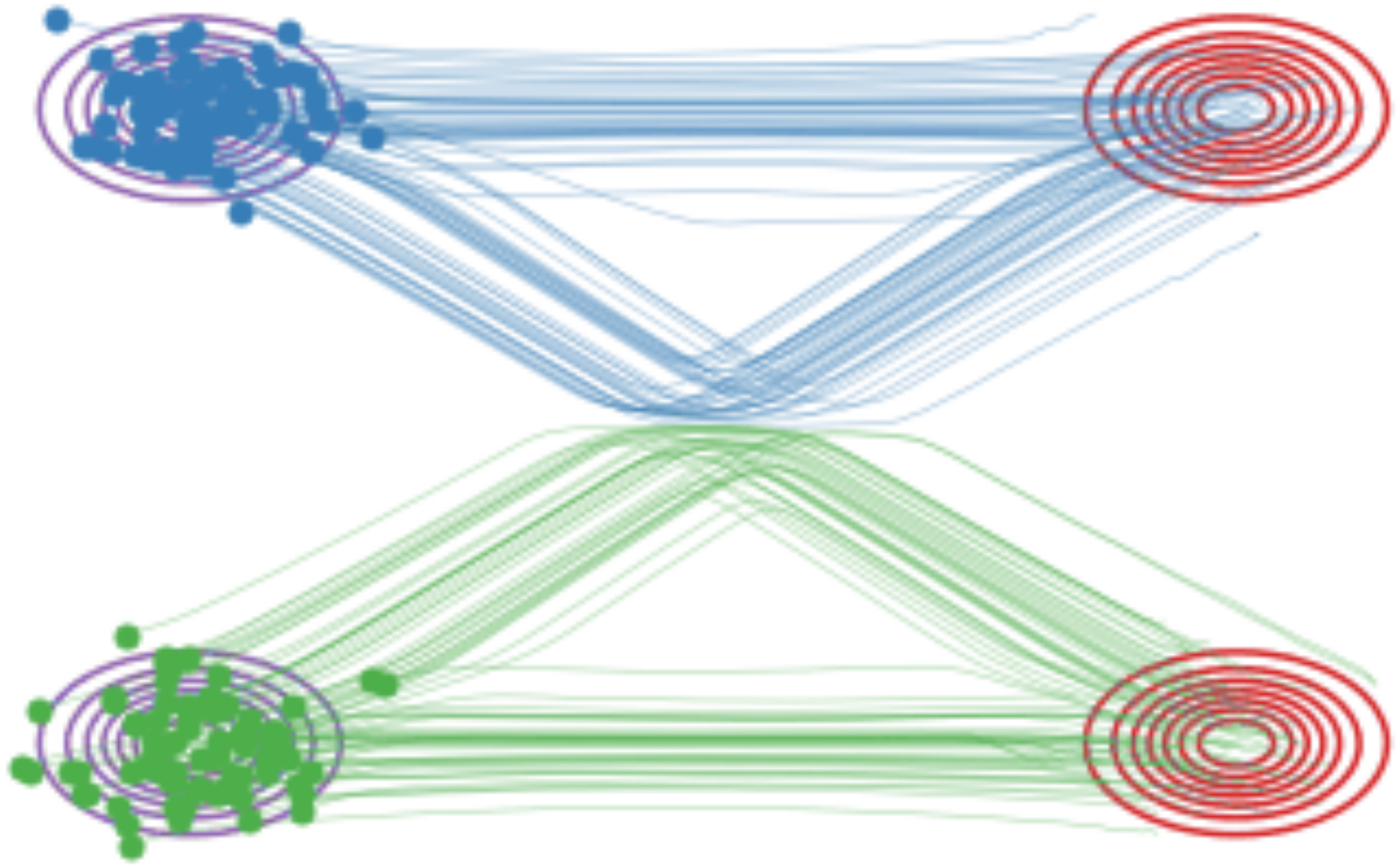
Rektifikált folyamok

Folyamillesztés

Kapcsolódó formalizmusok



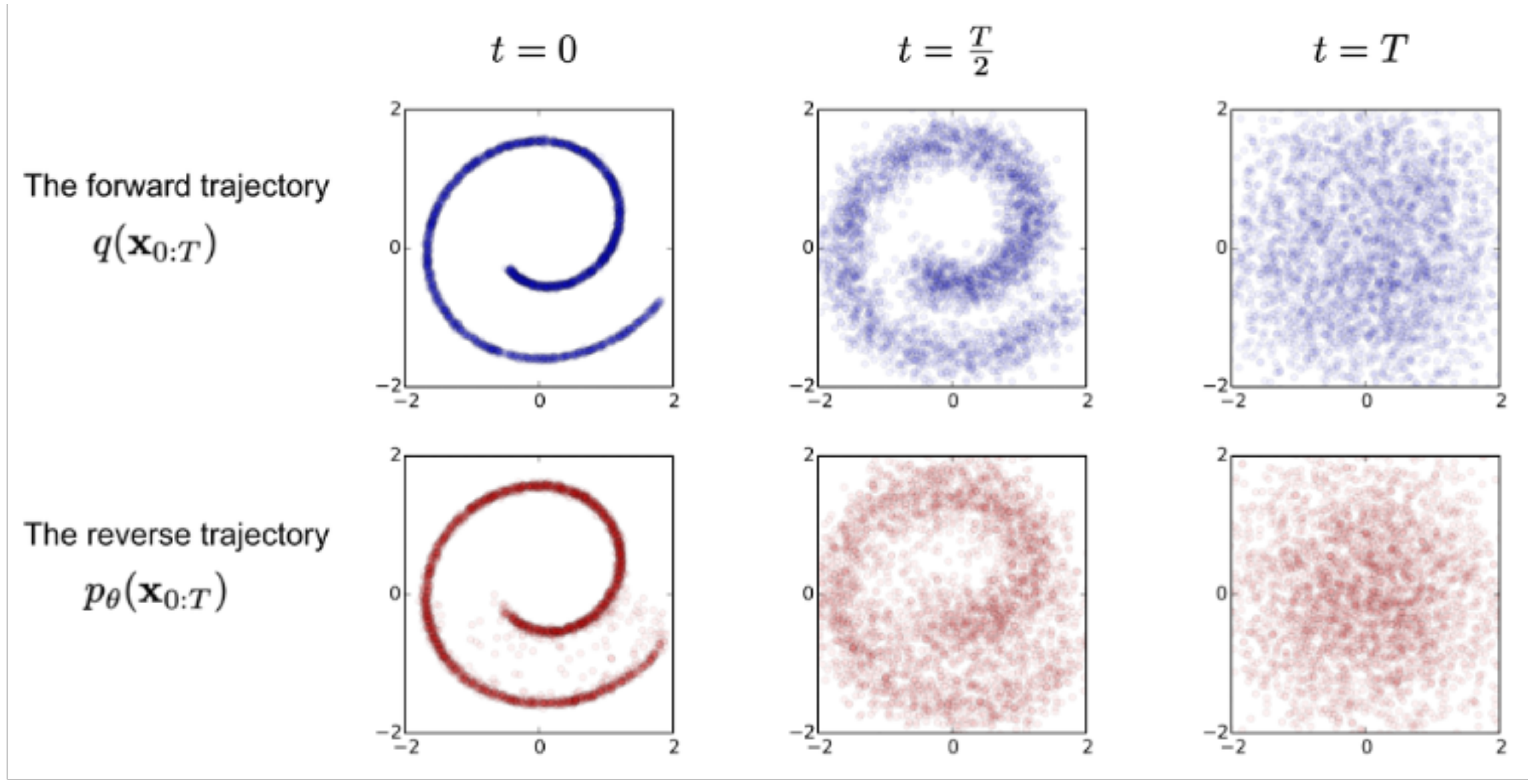
Sztochasztikus Interpolánsok



Rektifikált folyamok

Diffúziós Modellek

Történelem



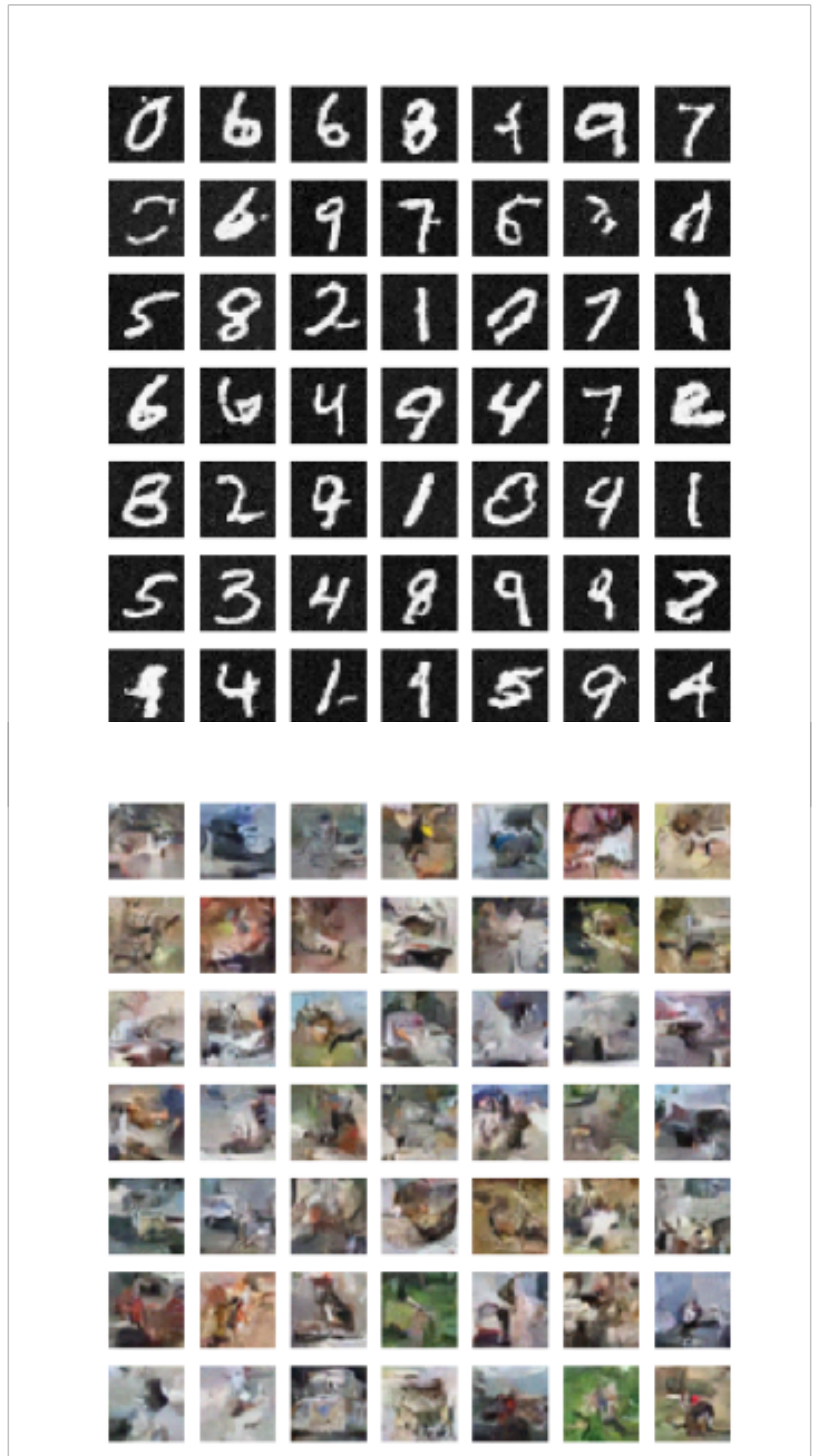
Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein
Stanford University

Eric A. Weiss
University of California, Berkeley

Niru Maheswaranathan
Stanford University

Surya Ganguli
Stanford University



2015: még csak egy generatív módszer a sok közül…

Diffúziós Modellek

Történelem

Generative Modeling by Estimating Gradients of the Data Distribution

Yang Song
Stanford University

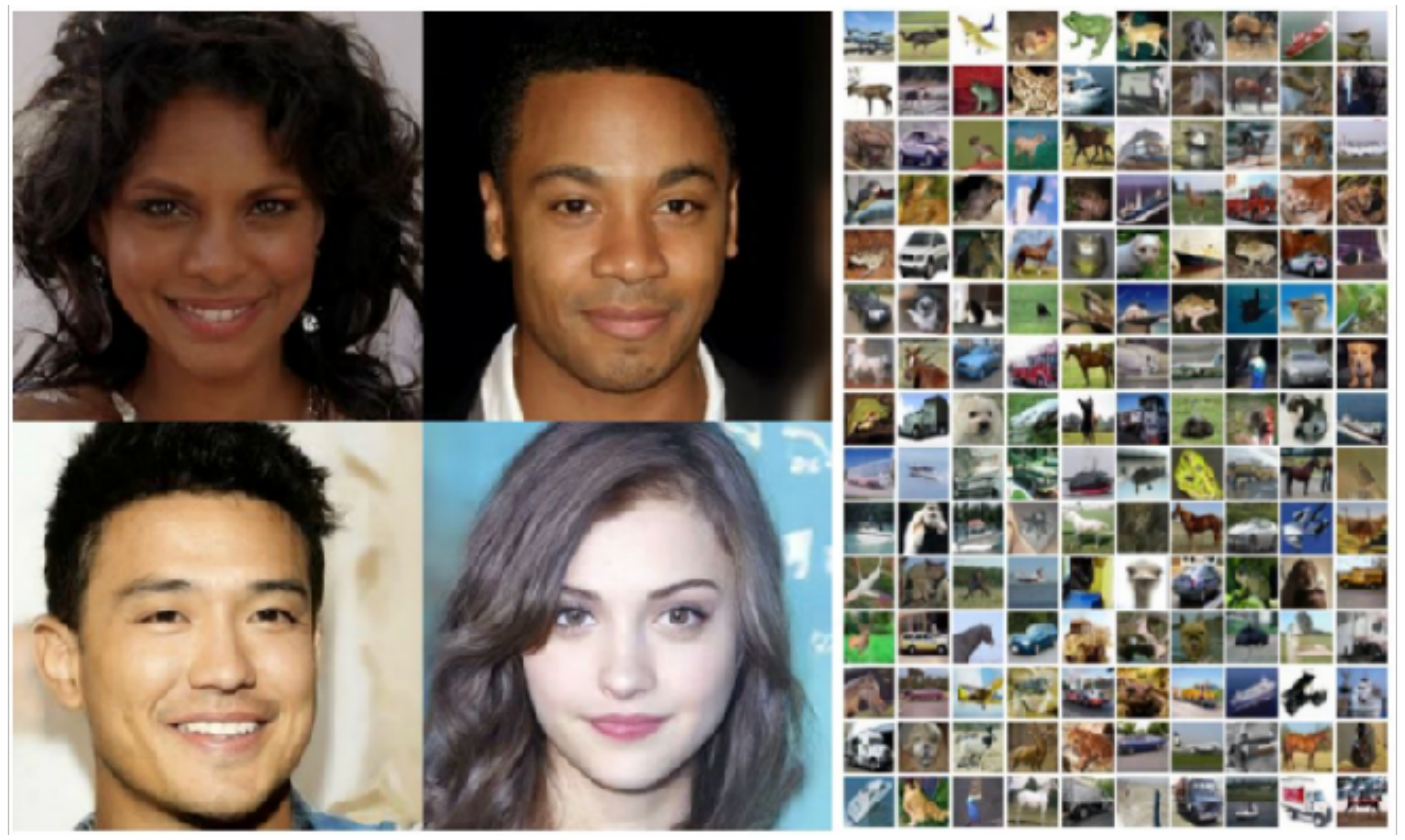
Stefano Ermon
Stanford University

Denoising Diffusion Probabilistic Models

Jonathan Ho
UC Berkeley

Ajay Jain
UC Berkeley

Pieter Abbeel
UC Berkeley



2019-2020: a diffúziós modellek újjáéledése

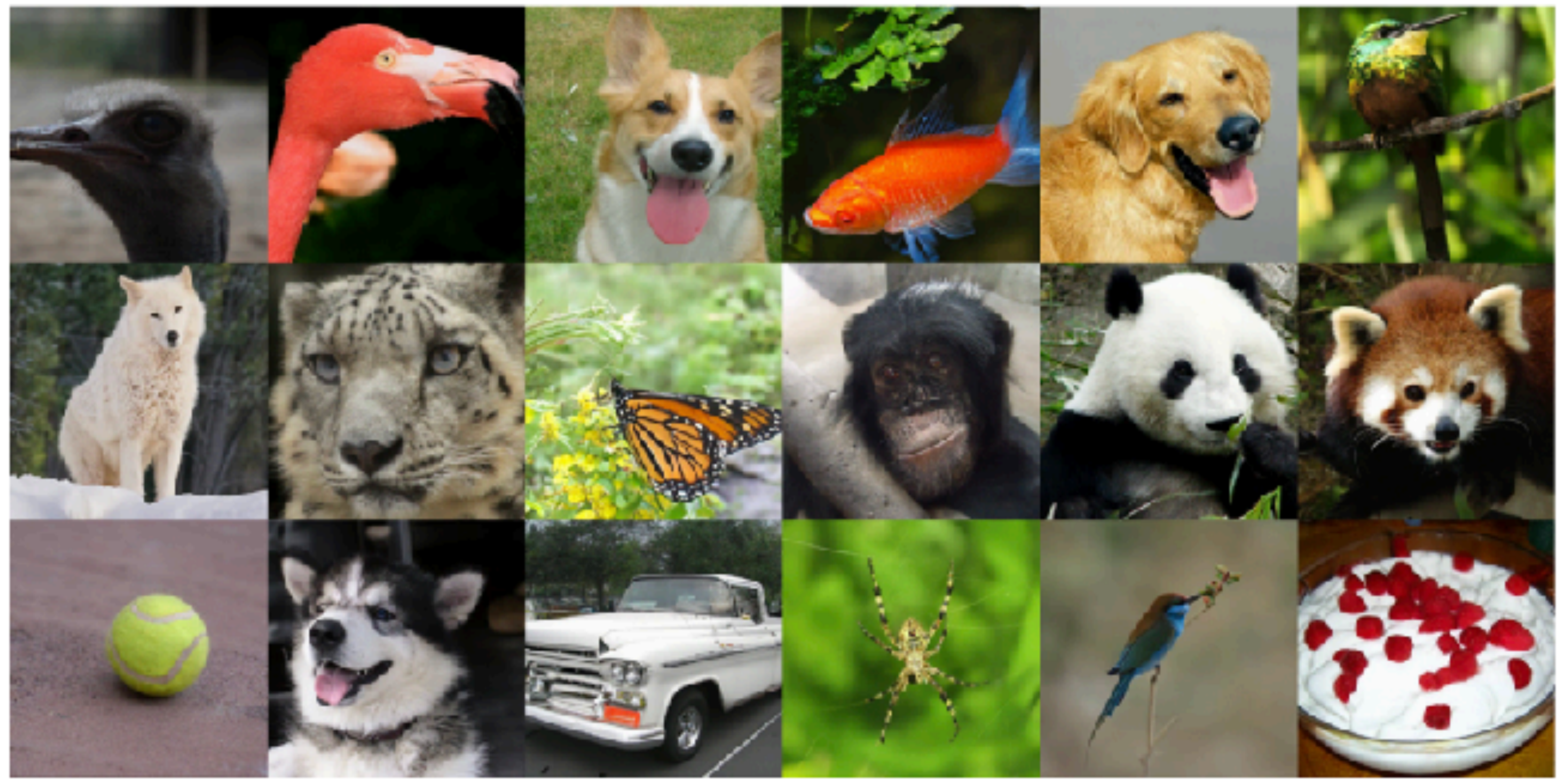
Diffúziós Modellek

Történelem

Diffusion Models Beat GANs on Image Synthesis

Prafulla Dhariwal*
OpenAI
prafulla@openai.com

Alex Nichol*
OpenAI
alex@openai.com



High-Resolution Image Synthesis with Latent Diffusion Models

Robin Rombach[1] * Andreas Blattmann[1] * Dominik Lorenz[1] Patrick Esser[R] Björn Ommer[1]
[1]Ludwig Maximilian University of Munich & IWR, Heidelberg University, Germany [R]Runway ML
https://github.com/CompVis/latent-diffusion

2021-22: növekvő felbontás és modell méretek

Diffúziós Modellek

Történelem

DALL·E 2

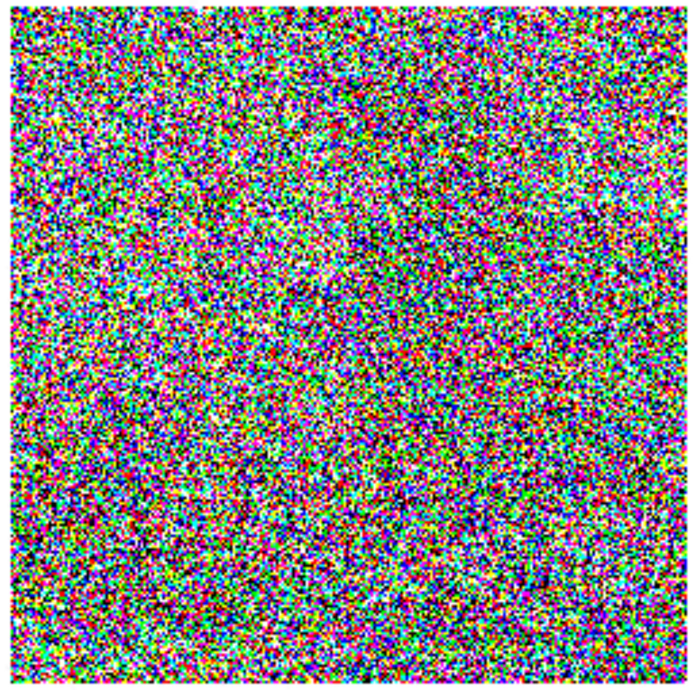
GOOGLE IMAGEN AI

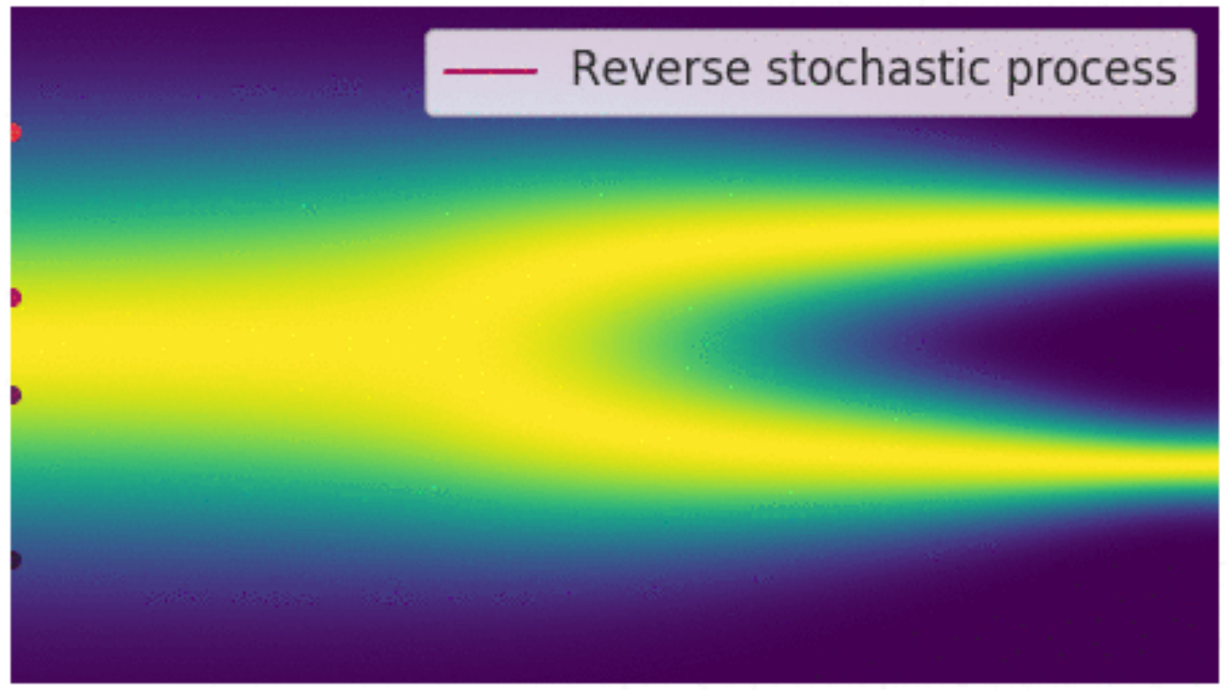
2022-: a diffúziós modelleket az ébredő nagy nyelvi modellekkel vezérelve megszületnek az első igazán általános képgenerátorok!

Diffúziós Modellek
Történelem

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

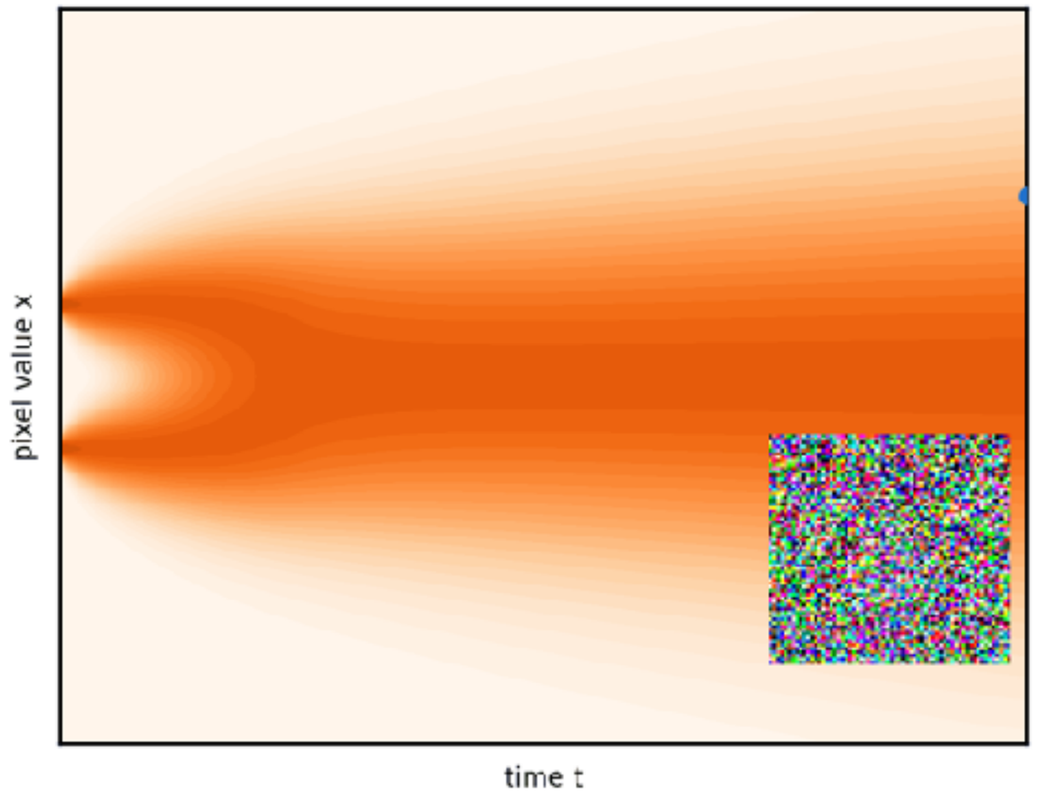
Yang Song* Stanford University yangsong@cs.stanford.edu
Jascha Sohl-Dickstein Google Brain jaschasd@google.com
Diederik P. Kingma Google Brain durk@google.com
Abhishek Kumar Google Brain abhishk@google.com
Stefano Ermon Stanford University ermon@cs.stanford.edu
Ben Poole Google Brain pooleb@google.com





Elucidating the Design Space of Diffusion-Based Generative Models

Tero Karras NVIDIA
Miika Aittala NVIDIA
Timo Aila NVIDIA
Samuli Laine NVIDIA



2022-: SDE/ODE értelmezések,
fokozatos áttérés hatékonyabb *determinisztikus* generálási módszerekre

Diffúziós Modellek

Történelem

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

Yang Song*
Stanford University
yangsong@cs.stanford.edu

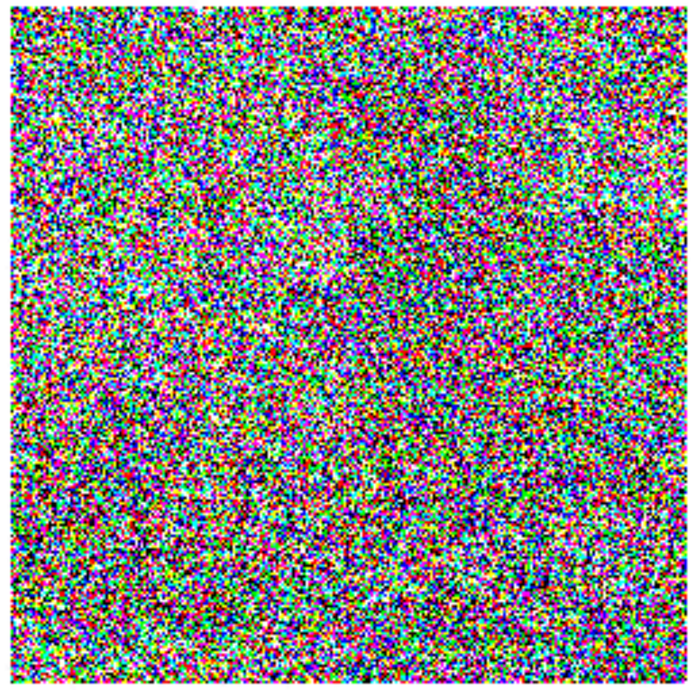
Jascha Sohl-Dickstein
Google Brain
jaschasd@google.com

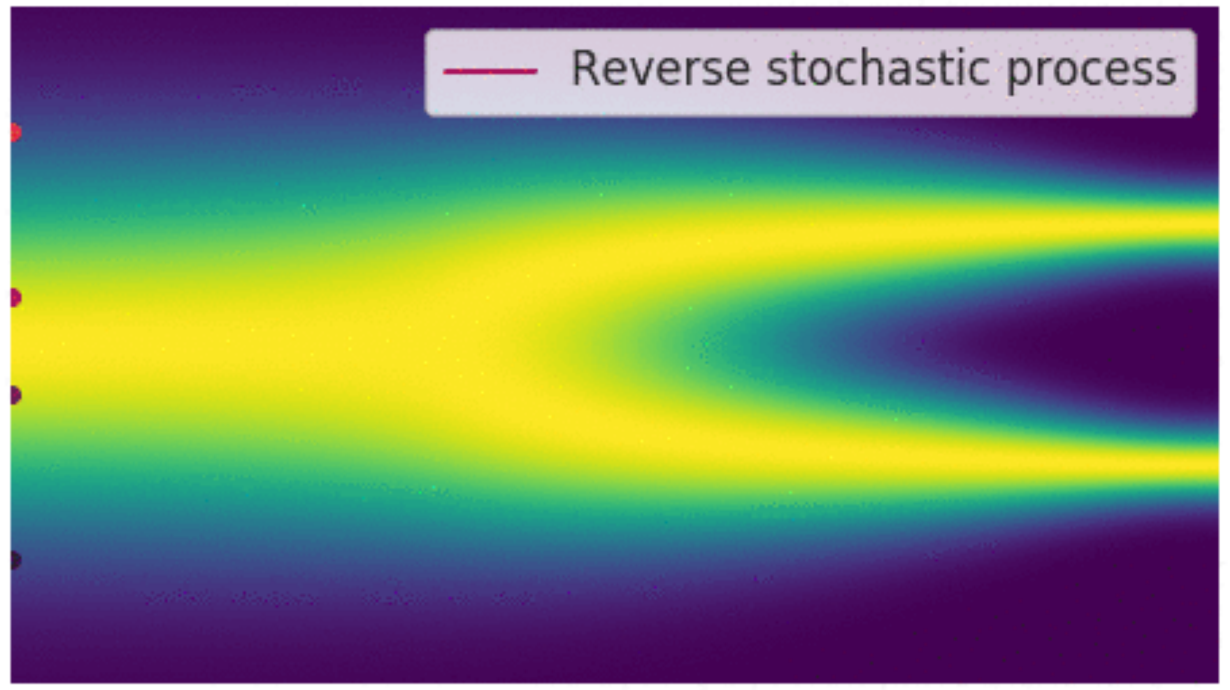
Diederik P. Kingma
Google Brain
durk@google.com

Abhishek Kumar
Google Brain
abhishk@google.com

Stefano Ermon
Stanford University
ermon@cs.stanford.edu

Ben Poole
Google Brain
pooleb@google.com





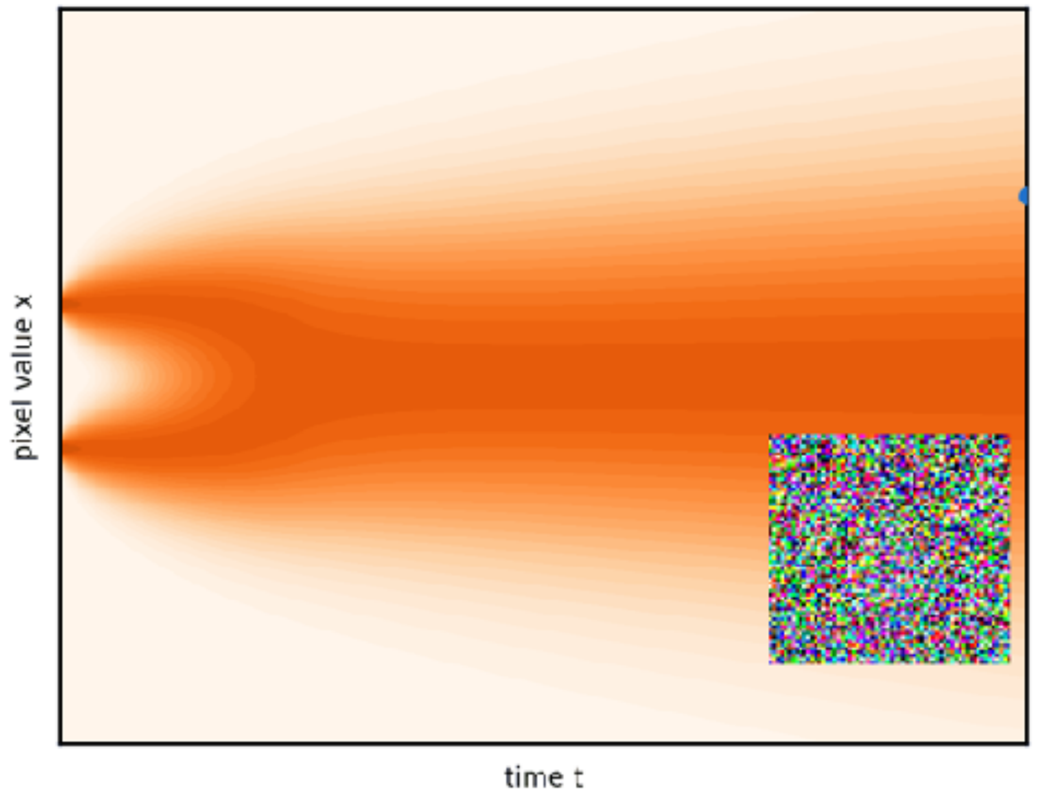
Elucidating the Design Space of Diffusion-Based Generative Models

Tero Karras
NVIDIA

Miika Aittala
NVIDIA

Timo Aila
NVIDIA

Samuli Laine
NVIDIA



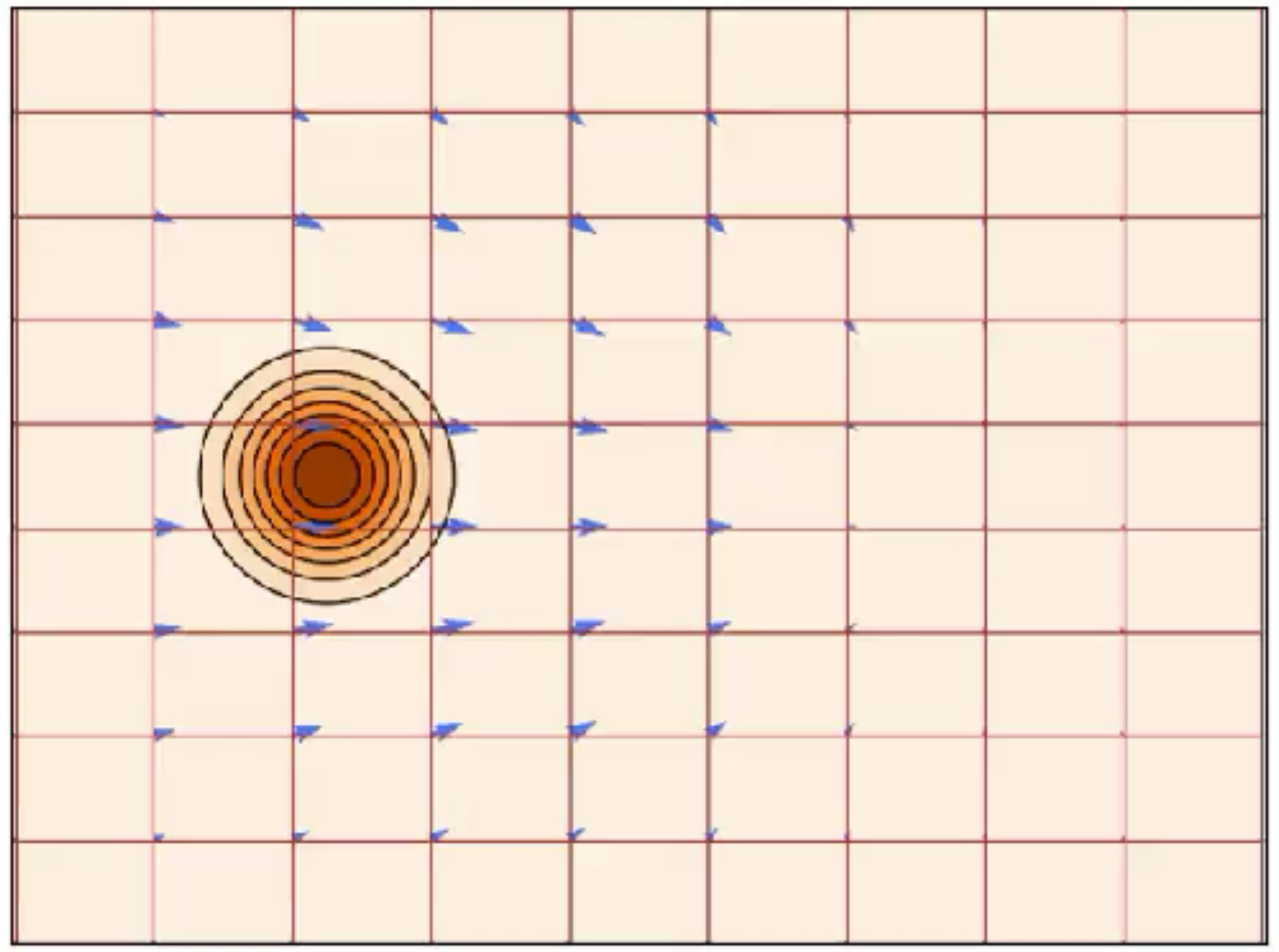
2022-: SDE/ODE értelmezések,
fokozatos áttérés hatékonyabb *determinisztikus* generálási módszerekre

Diffúziós Modellek

Történelem

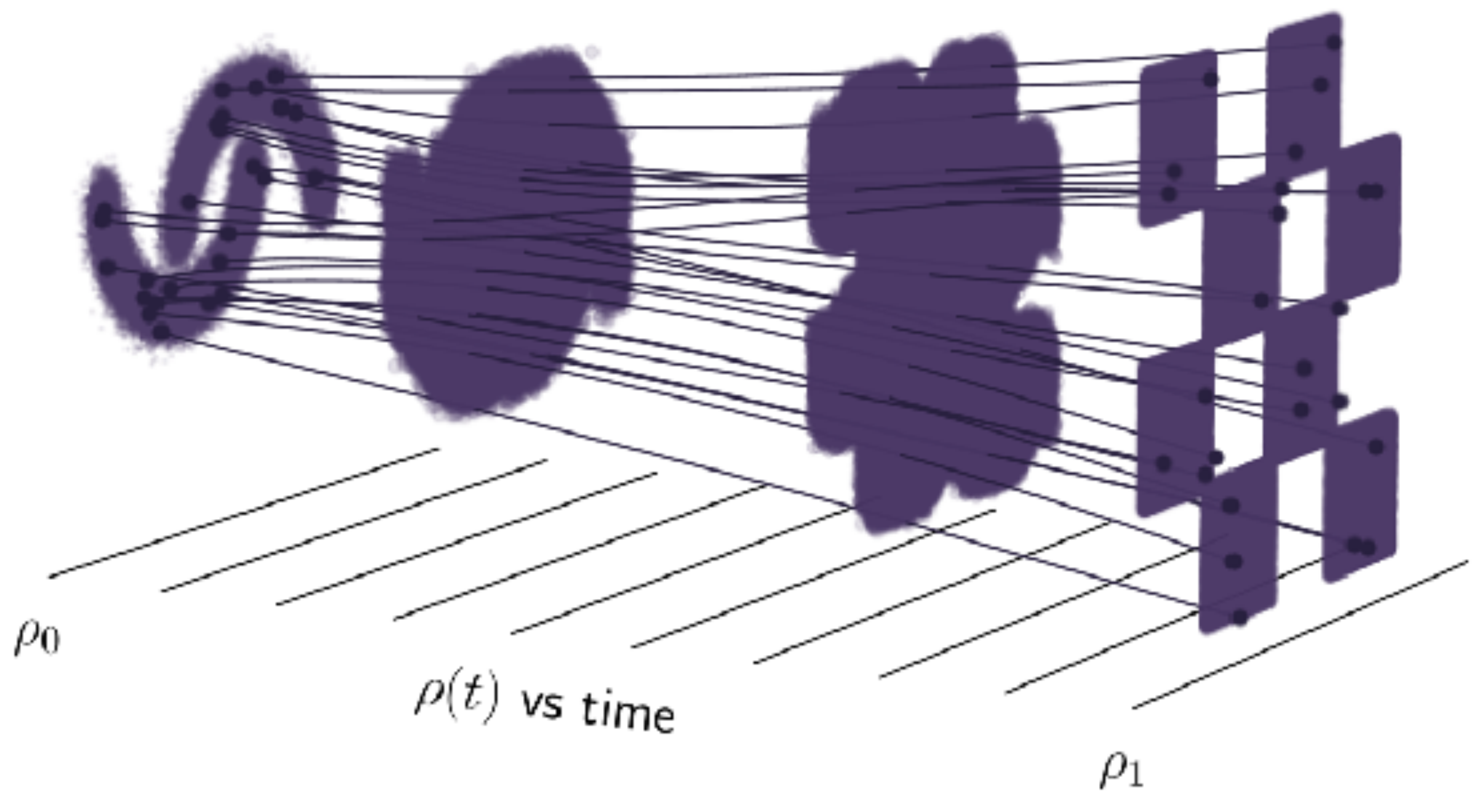
FLOW MATCHING FOR GENERATIVE MODELING

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



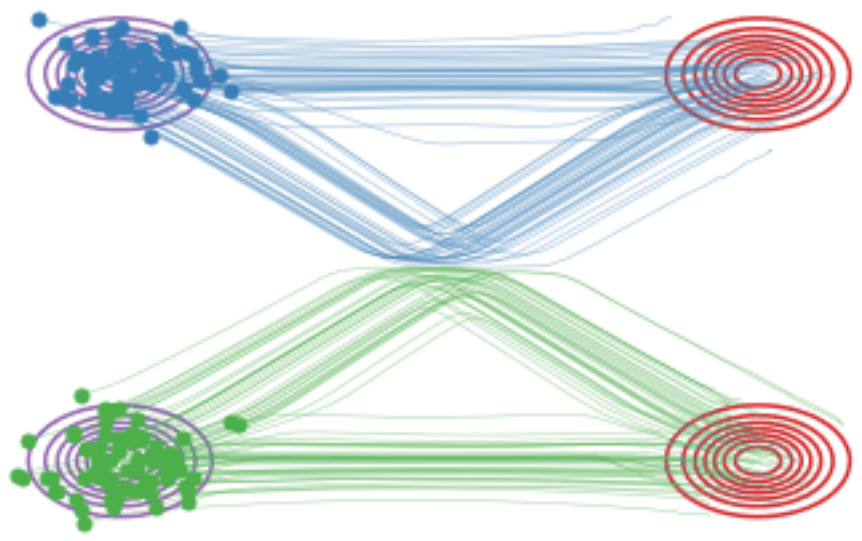
**Stochastic Interpolants:
A Unifying Framework for Flows and Diffusions**

Michael S. Albergo* ALBERGO@NYU.EDU
*Center for Cosmology and Particle Physics
New York University
New York, NY 10012, USA*
Nicholas M. Boffi* BOFFI@CIMS.NYU.EDU
*Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA*
Eric Vanden-Eijnden EVE2@CIMS.NYU.EDU
*Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA*



Flow Straight and Fast:
Learning to Generate and Transfer Data with Rectified Flow

Xingchao Liu*
University of Texas at Austin
xcliu@utexas.edu
Chengyue Gong*
University of Texas at Austin
cygong@cs.utexas.edu
Qiang Liu
University of Texas at Austin
lqiang@cs.utexas.edu

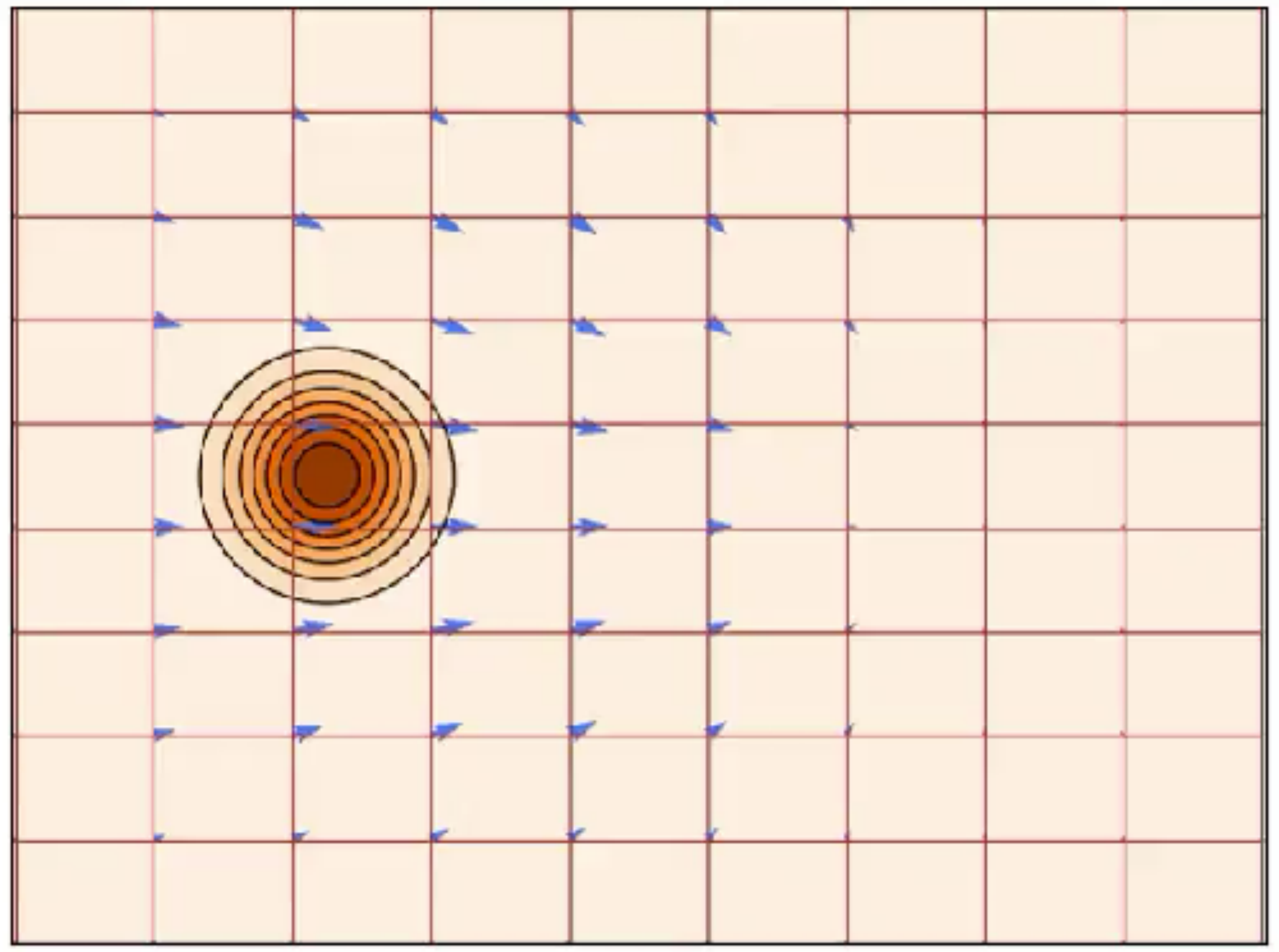


2022-23: ODE központú megközelítések, folyamillesztés

Diffúziós Modellek
Történelem

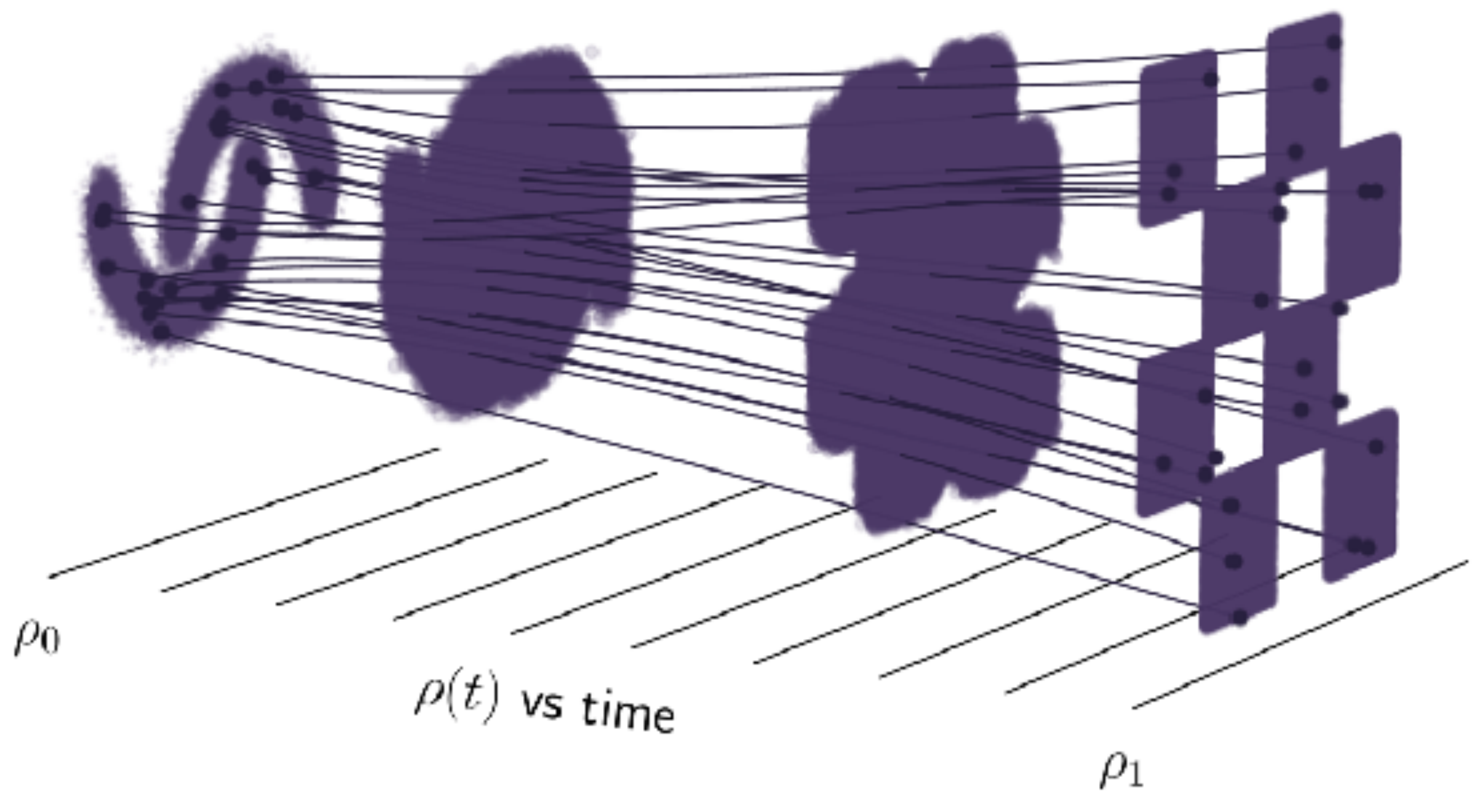
FLOW MATCHING FOR GENERATIVE MODELING

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



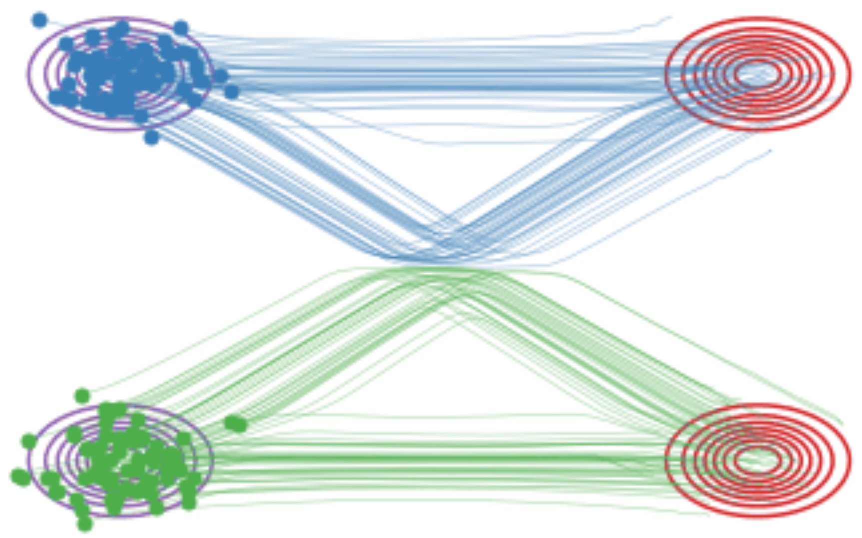
Stochastic Interpolants:
A Unifying Framework for Flows and Diffusions

Michael S. Albergo* ALBERGO@NYU.EDU
Center for Cosmology and Particle Physics
New York University
New York, NY 10012, USA
Nicholas M. Boffi* BOFFI@CIMS.NYU.EDU
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA
Eric Vanden-Eijnden EVE2@CIMS.NYU.EDU
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012, USA



Flow Straight and Fast:
Learning to Generate and Transfer Data with Rectified Flow

Xingchao Liu*
University of Texas at Austin
xcliu@utexas.edu
Chengyue Gong*
University of Texas at Austin
cygong@cs.utexas.edu
Qiang Liu
University of Texas at Austin
lqiang@cs.utexas.edu



2022-23: ODE központú megközelítések, folyamillesztés

Diffúziós Modellek

Történelem

FLUX.1 Kontext

腾讯混元-Video

Isaac GROOT

π_0: A Vision-Language-Action Flow Model for General Robot Control

Physical Intelligence

SimpleFold: Folding Proteins is Simpler than You Think

Yuyang Wang, Jiarui Lu*, Navdeep Jaitly, Josh Susskind, Miguel Angel Bautista

A jelenlegi trend: **folyamillesztés** mindenhol, mindenre!

Diffúziós Modellek

Történelem

Stable Diffusion

Midjourney

FLUX.1
A new era of creation

Black Forest Labs

Diffusers

Napjainkra már komoly ökoszisztéma épült a diffúziós modellek köré!

Diffúziós Modellek

Jegyzetajánló

The Principles of Diffusion Models

From Origins to Advances

Chieh-Hsin Lai
Sony AI

Yang Song
OpenAI

Dongjun Kim
Stanford University

Yuki Mitsufuji
Sony Corporation, Sony AI

Stefano Ermon
Stanford University

https://arxiv.org/abs/2510.21890

https://the-principles-of-diffusion-models.github.io/#/blog

An Introduction to Flow Matching and Diffusion Models

Peter Holderrieth and Ezra Erives

Website: https://diffusion.csail.mit.edu/

STEP-BY-STEP DIFFUSION: AN ELEMENTARY TUTORIAL

Preetum Nakkiran[1], Arwen Bradley[1], Hattie Zhou[1,2], Madhu Advani[1]

[1]Apple, [2]Mila, Université de Montréal

https://arxiv.org/abs/2406.08929

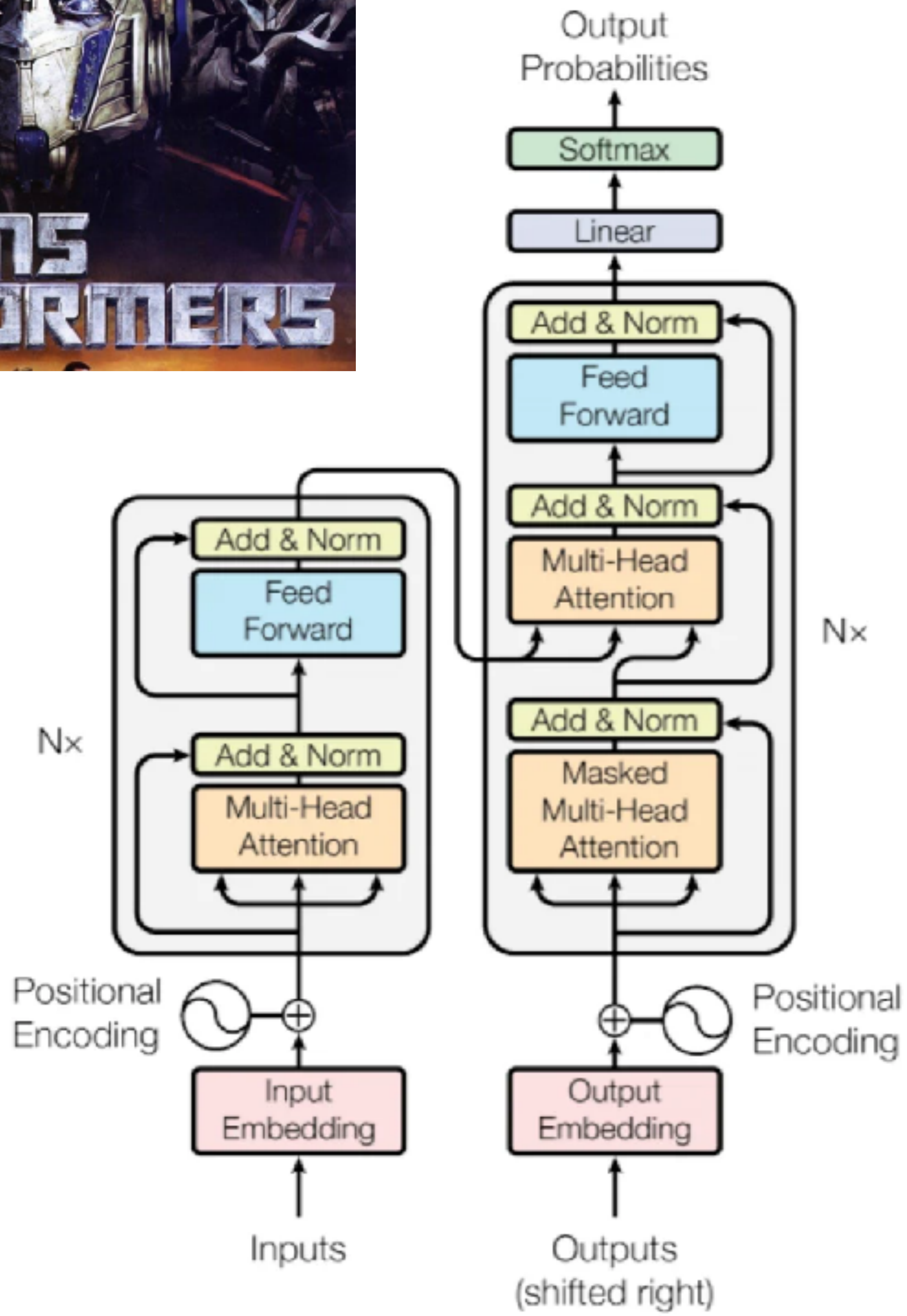
Generative Modeling by Estimating Gradients of the Data Distribution

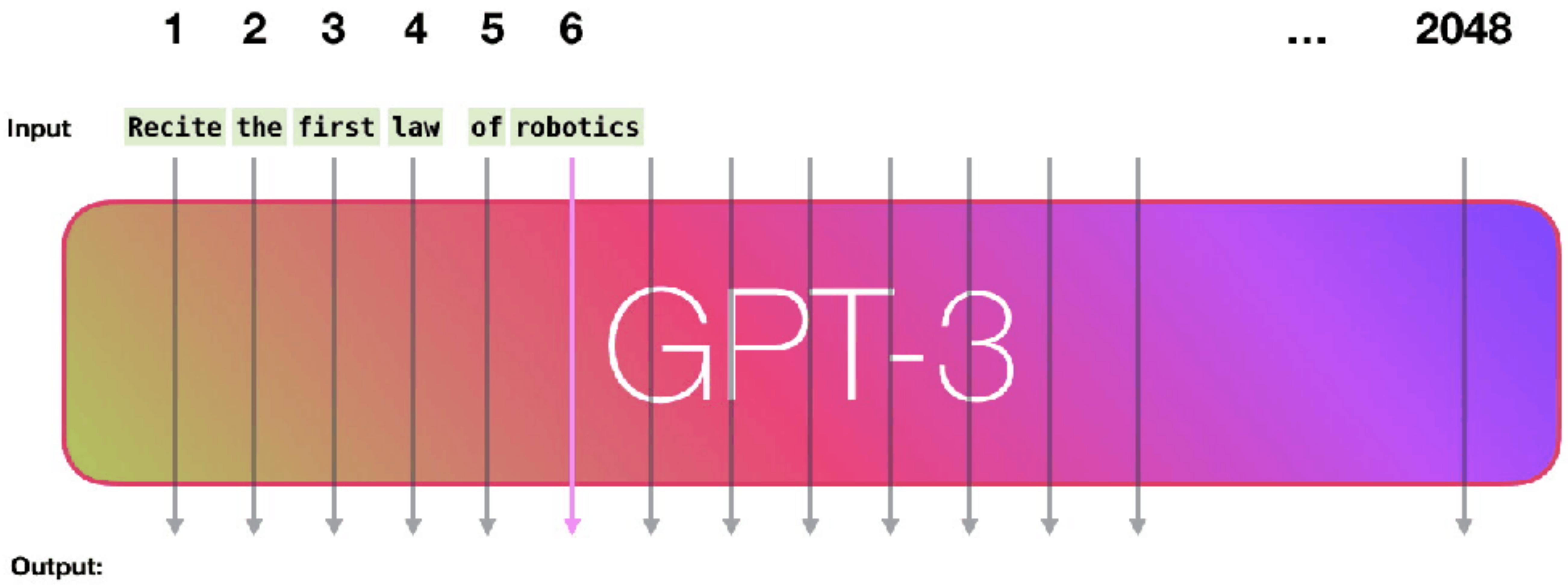
AUTHORS
Yang Song

https://yang-song.net/blog/2021/score/

Következő előadás:
Autoregresszív Generálás

- Transformer architektúra / figyelem
- Nagy nyelvi modellek alapjai
- Multimodalitás





Következő előadás:

Autoregresszív Generálás

- Transformer architektúra / figyelem
- Nagy nyelvi modellek alapjai
- Multimodalitás