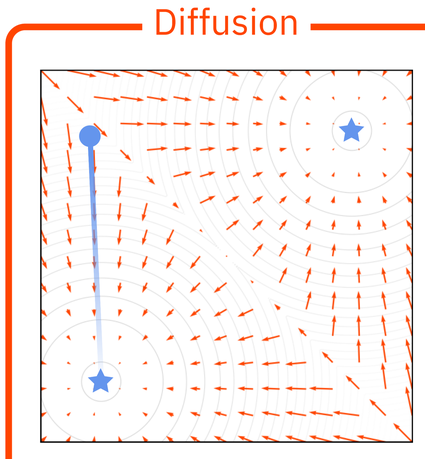
Memory in Plain Sight

Uncanny Resemblance of **Diffusion** & **Associative Memory**

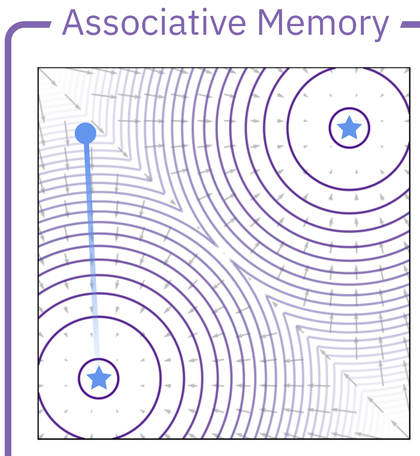
Benjamin Hoover **Hendrik** Strobelt **Dmitry** Krotov **Judy** Hoffman **Zsolt** Kira **Polo** Chau

What are Memories? 🤔

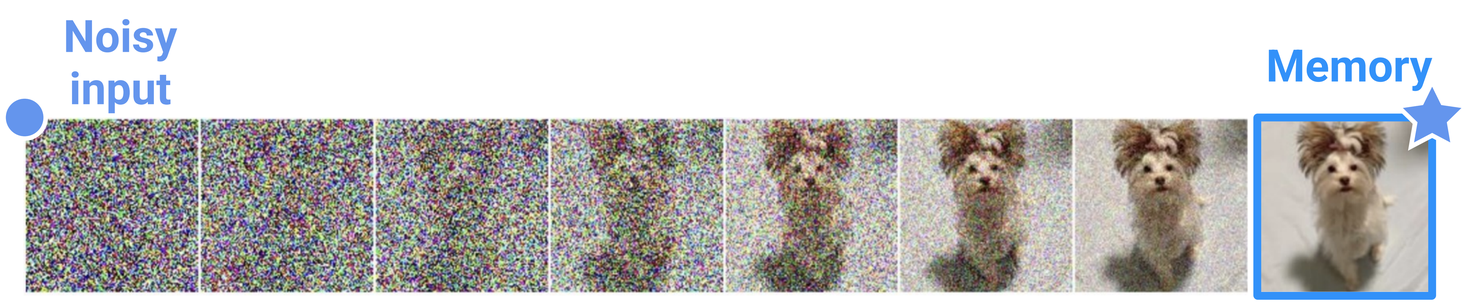
Energy unites **Diffusion** and **Associative Memory**



Learn a **score** function (arrows) point to **peaks in log-probability (★)**



Learn an **energy** function (contours) around **basins of attraction (★)**



Memory retrieval minimizes energy in *forward time*

$$\tau \frac{dx}{dt} = -\frac{\partial E_\theta}{\partial x} \qquad x^{t+1} = x^t - \alpha \frac{\partial E_\theta}{\partial x}$$

Memories live in **energy basins** and represent the **fixed points** of retrieval dynamics

Diffusion Models vs Associative Memories

1. Do I have an **energy**?
2. Is energy **bounded from below**?
3. $\frac{dE}{dt} \leq 0$ everywhere?

⇓

Energy is **Lyapunov**

You have an **Associative Memory**

	Diffusion Models	Associative Memories
Parameterizes	**Score Function**	**Energy Function**
Dynamics over	t in [0,T]	t > 0
Fixed point attractor	✗	✓
Constrained arch.	✗	✓
Lyapunov energy	✗	✓

Sandbox for general Associative Memories

Neurons are dynamical variables that evolve to minimize their contribution to a global energy

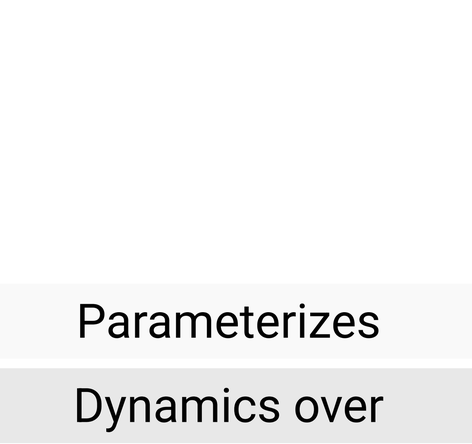
Each neuron has two properties
1. an *internal state* $\mathbf{x}$ that evolves in time
2. a convex, scalar *Lagrangian function* $\mathcal{L}$

The Lagrangian defines the *activations*

$$\mathbf{g} = \nabla_{\mathbf{x}} \mathcal{L}$$

and the neuron's energy

$$E = \left(\mathbf{g}^\mathsf{T} \mathbf{x}\right) - \mathcal{L}$$



Synapses describe relationships between the activations of dynamic variables (neurons)

Synapses can be any function, but are often parameterized by weights $\mathbf{W}$

$$E = -\mathbf{g}_m^\mathsf{T} \mathbf{W} \mathbf{g}_f$$

Neuron states evolve to descend **total energy**

$$E = E_{\text{features}} + E_{\text{memories}} + E$$

IBM Research Georgia Tech