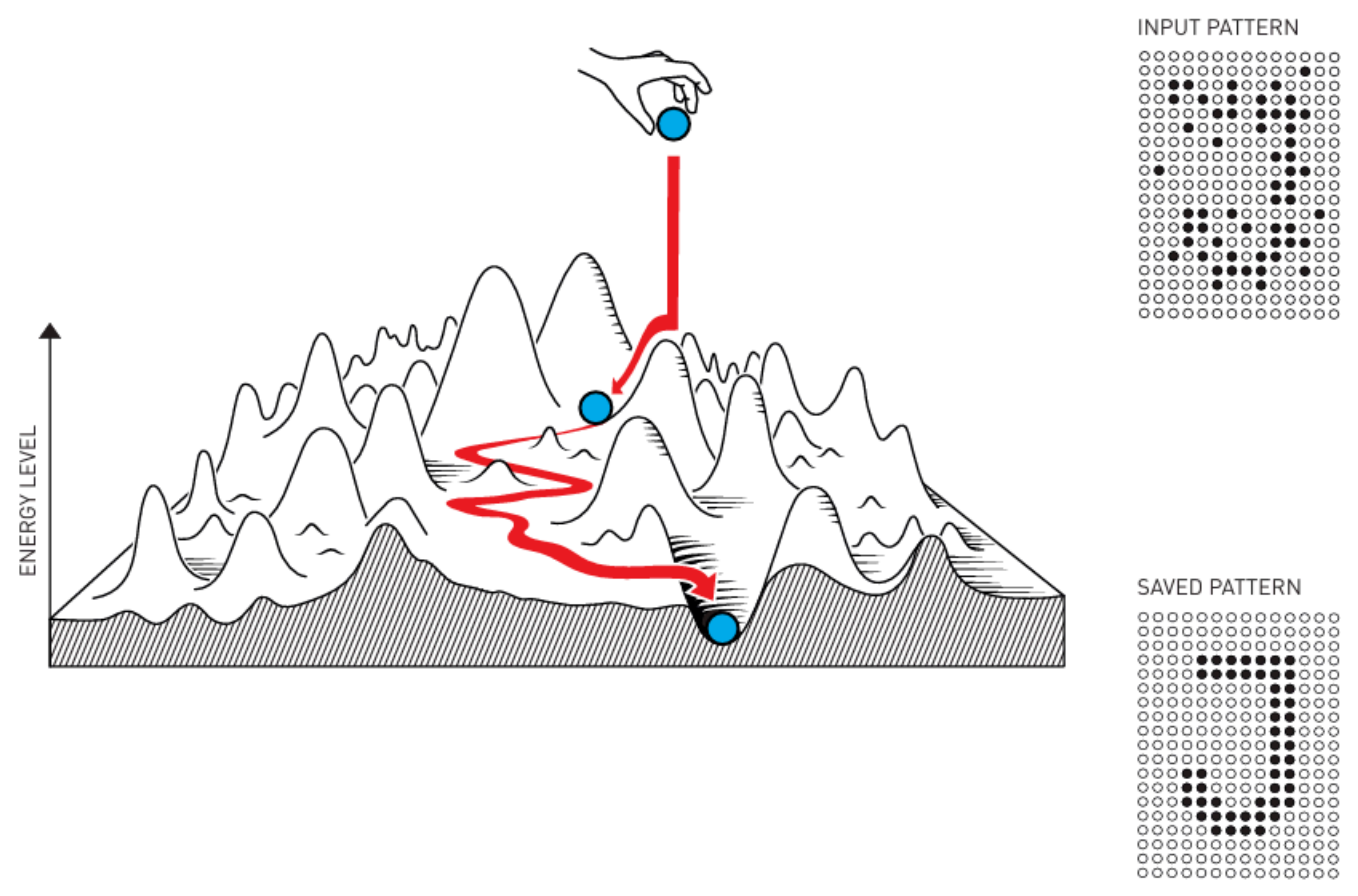
Low-dimensional representations & associative memory energy functions

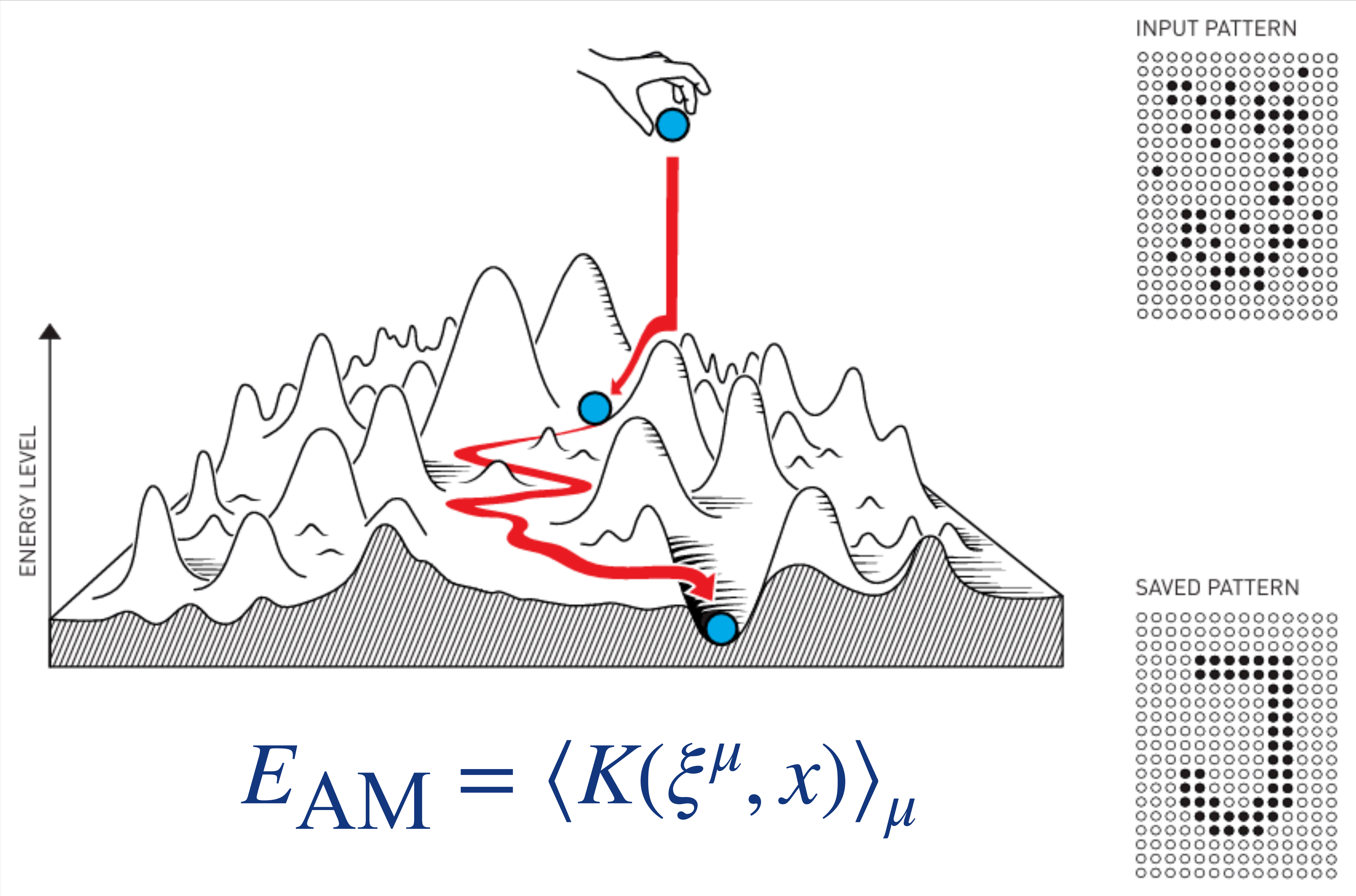
David Lipshutz
New Frontiers in Associative Memory

Associative memory models



Royal Swedish Academy of Sciences, nobelprize.org

Associative memory models

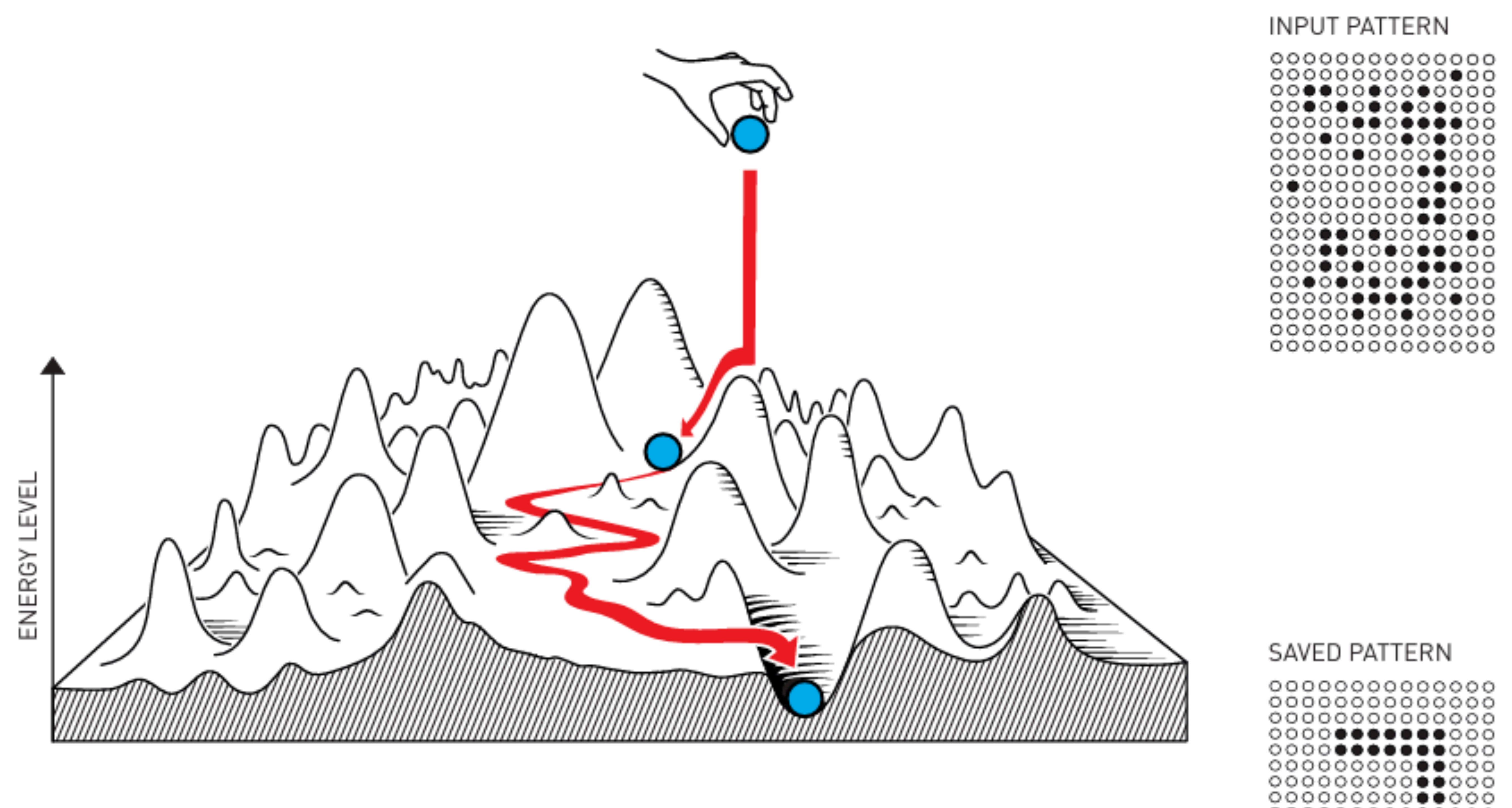


$$E_{\mathrm{AM}} = \langle K(\xi^{\mu}, x) \rangle_{\mu}$$

Energy function depends on clean storage patterns $\{\xi^{\mu}\}$

Royal Swedish Academy of Sciences, nobelprize.org

Associative memory models

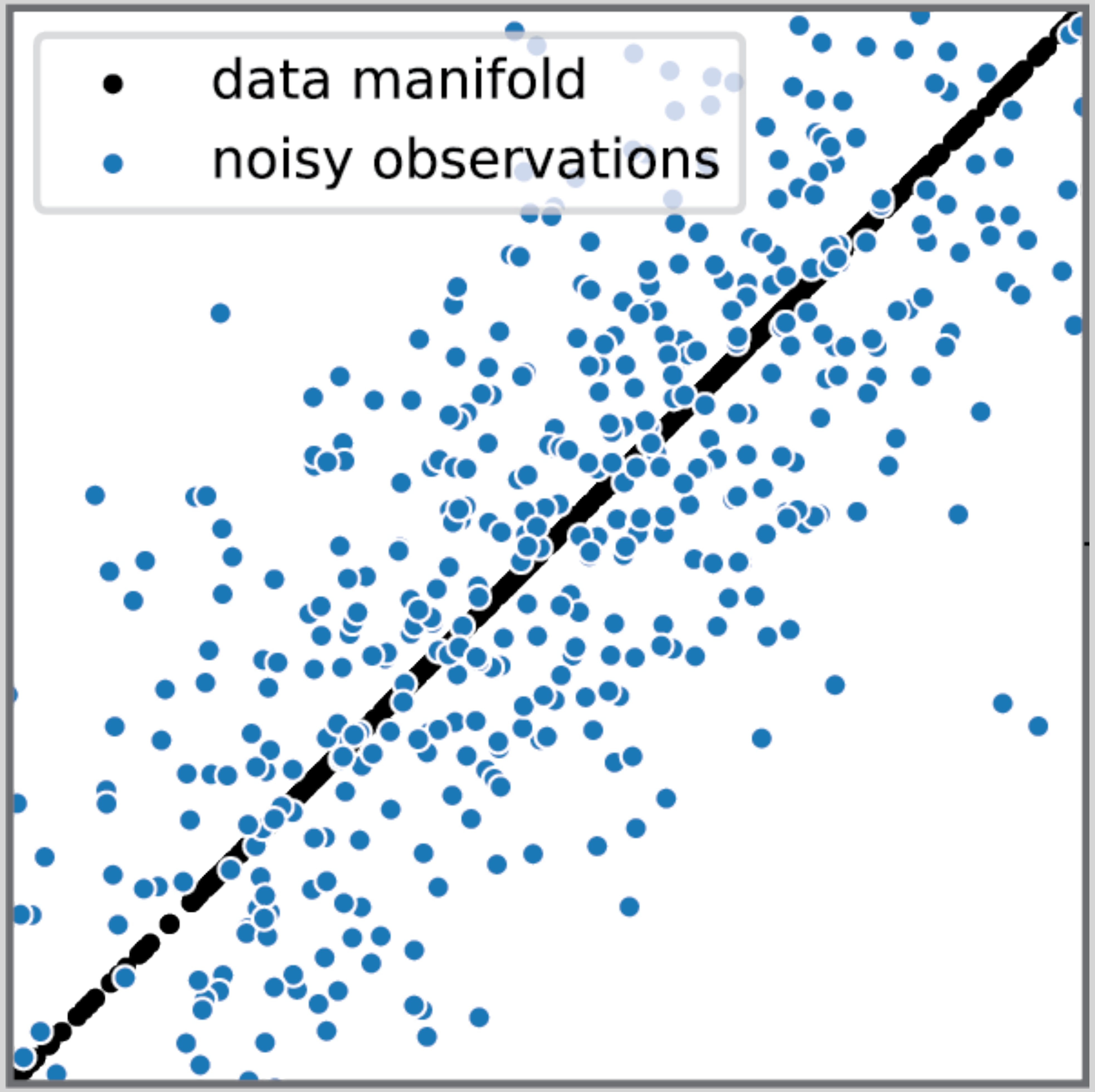


$$E = \langle ? \rangle$$

What if the energy function has to be learned from noisy input patterns?

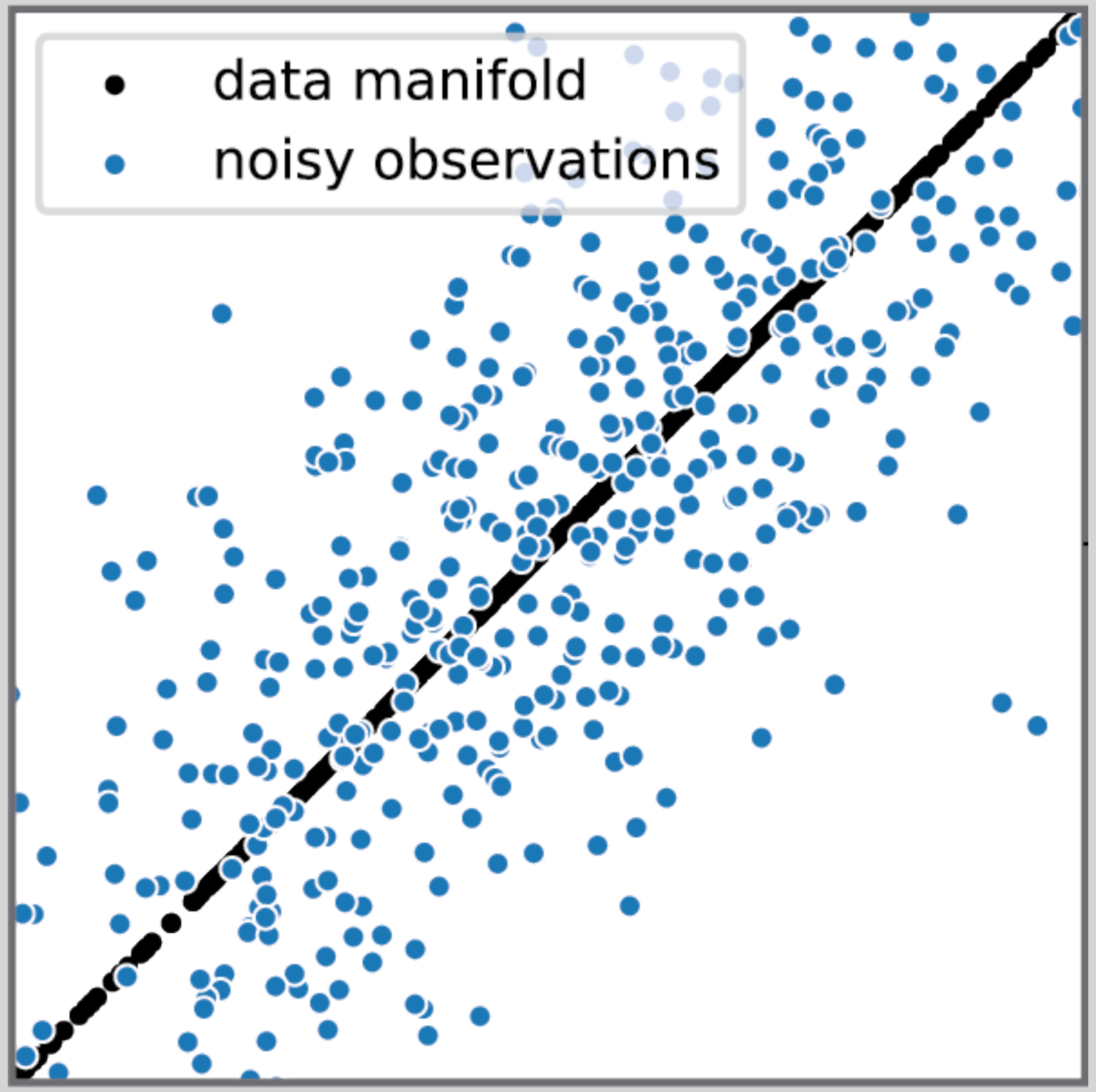
Royal Swedish Academy of Sciences, nobelprize.org

Learning from noisy observations

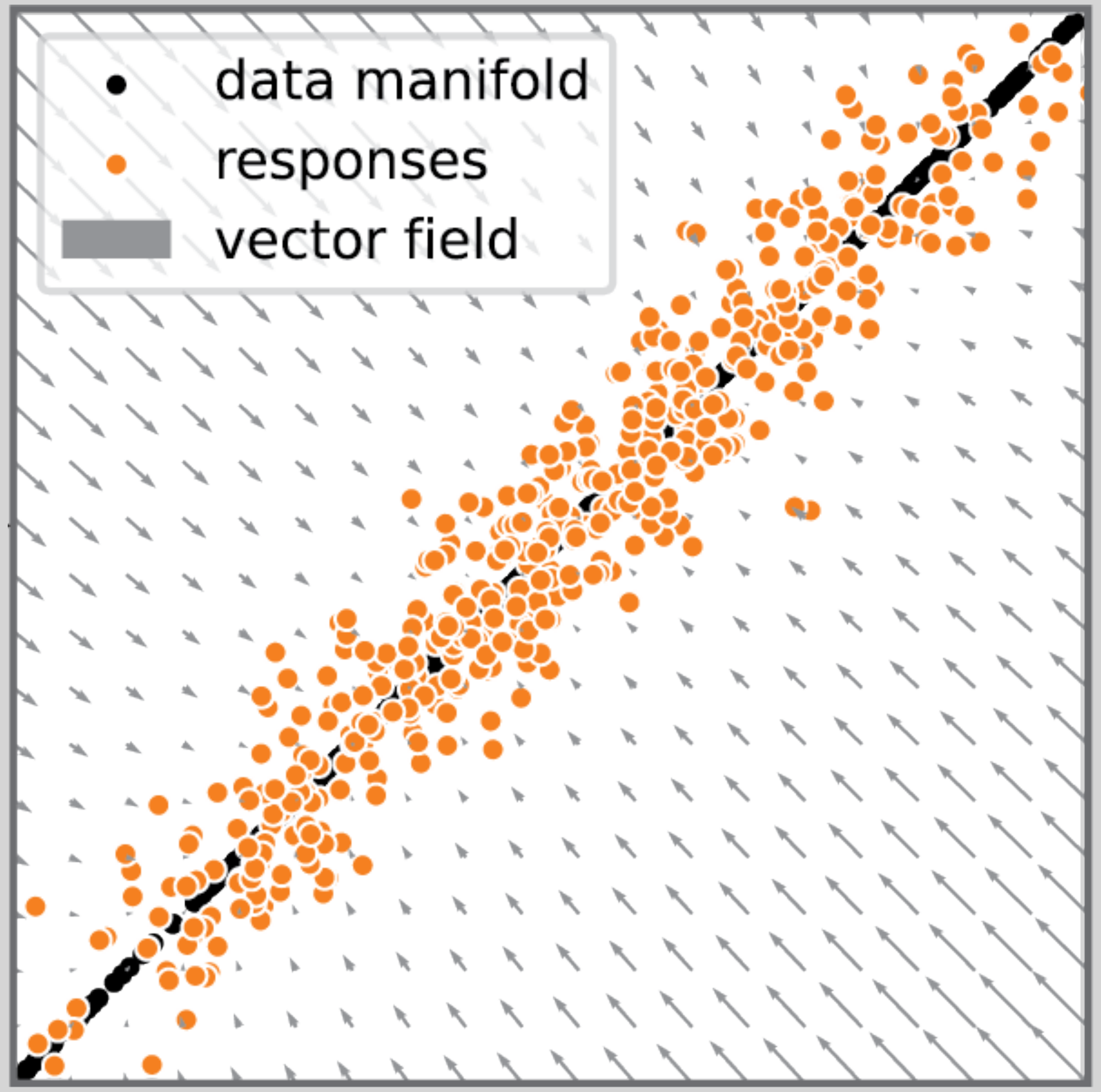


Learning from noisy observations

Inputs: noisy observations x



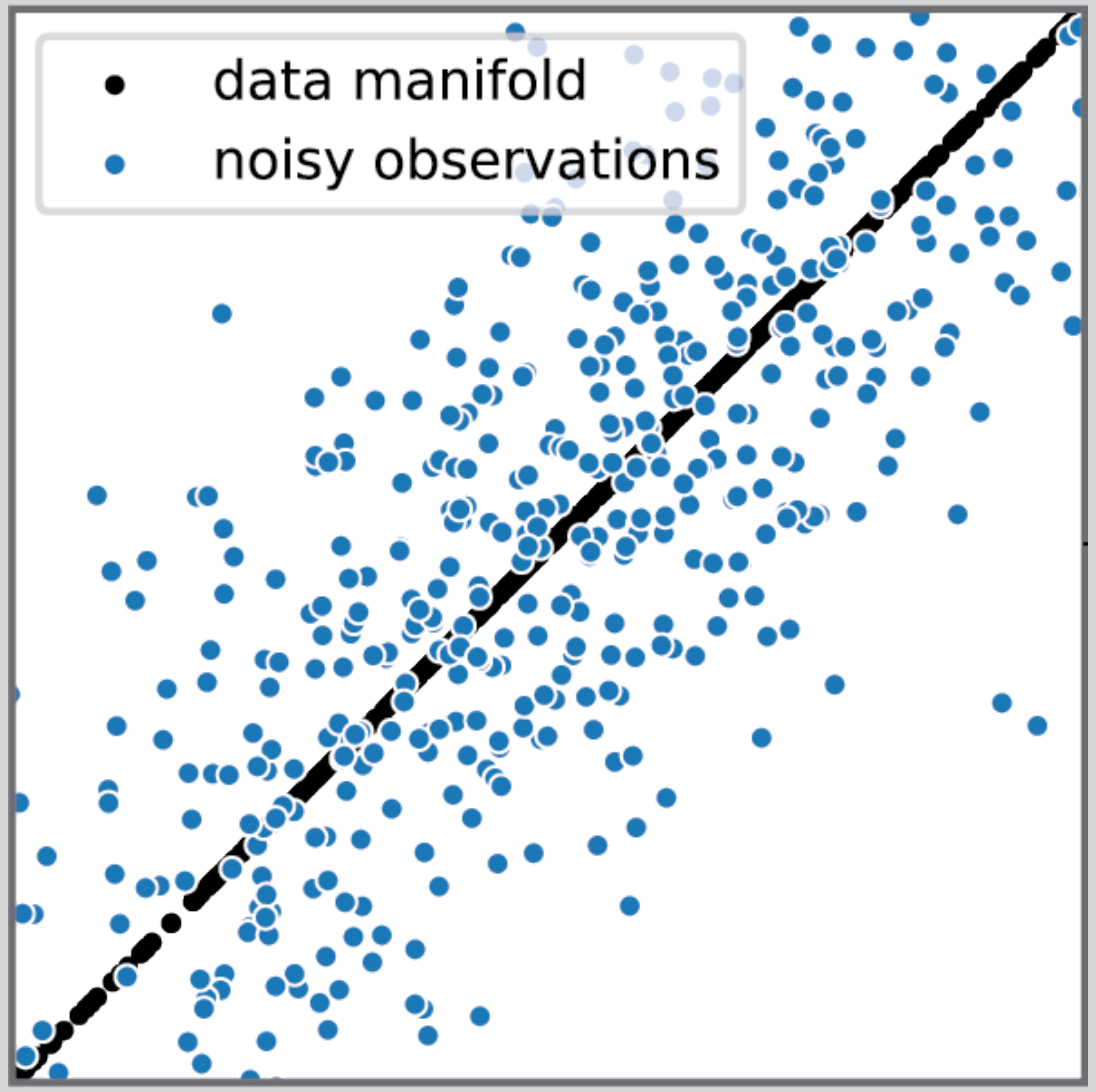
Output: low-dim responses y



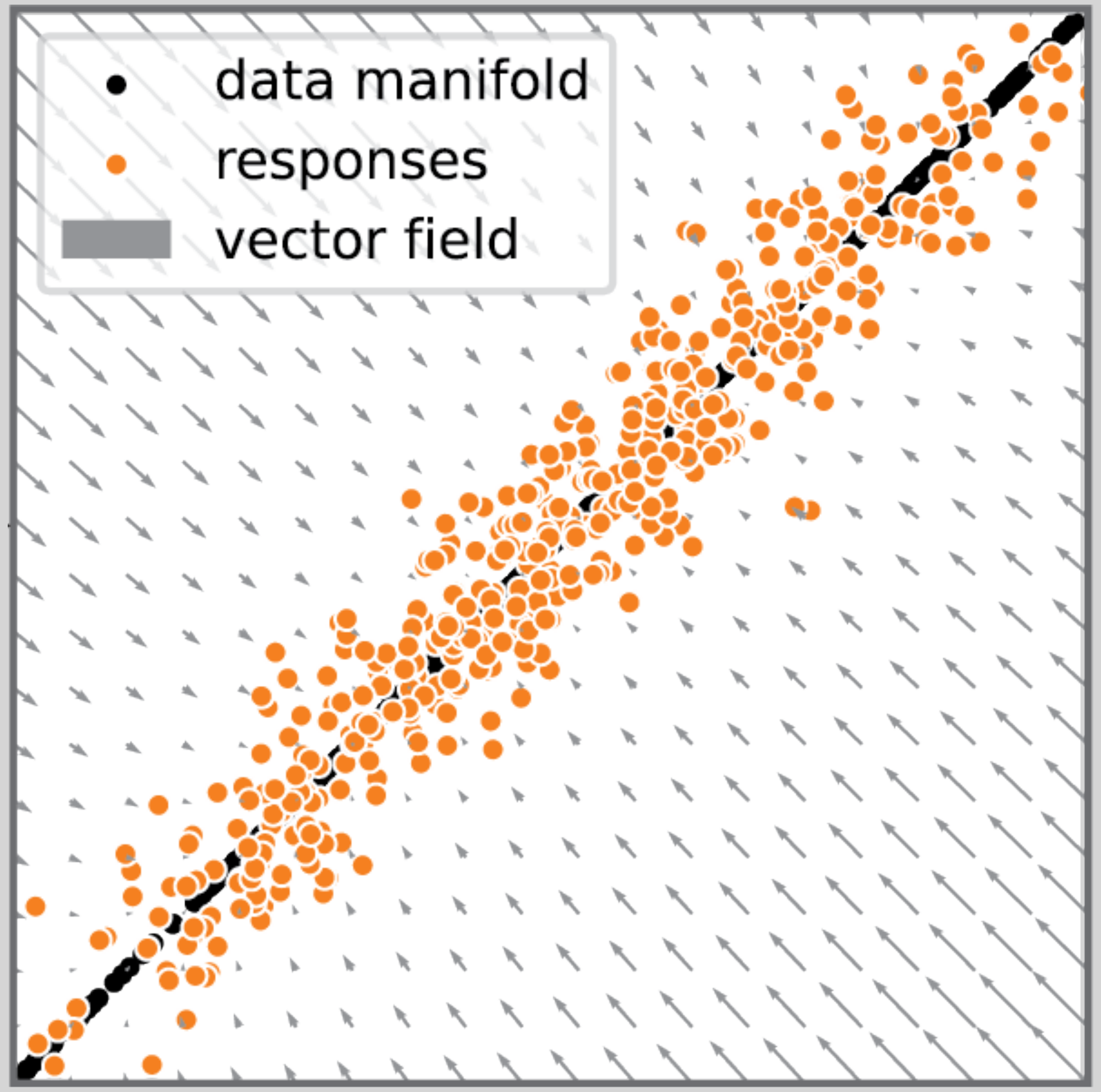
Duong *et al.* COSYNE 2026

Learning from noisy observations

Inputs: noisy observations x



Output: low-dim responses y

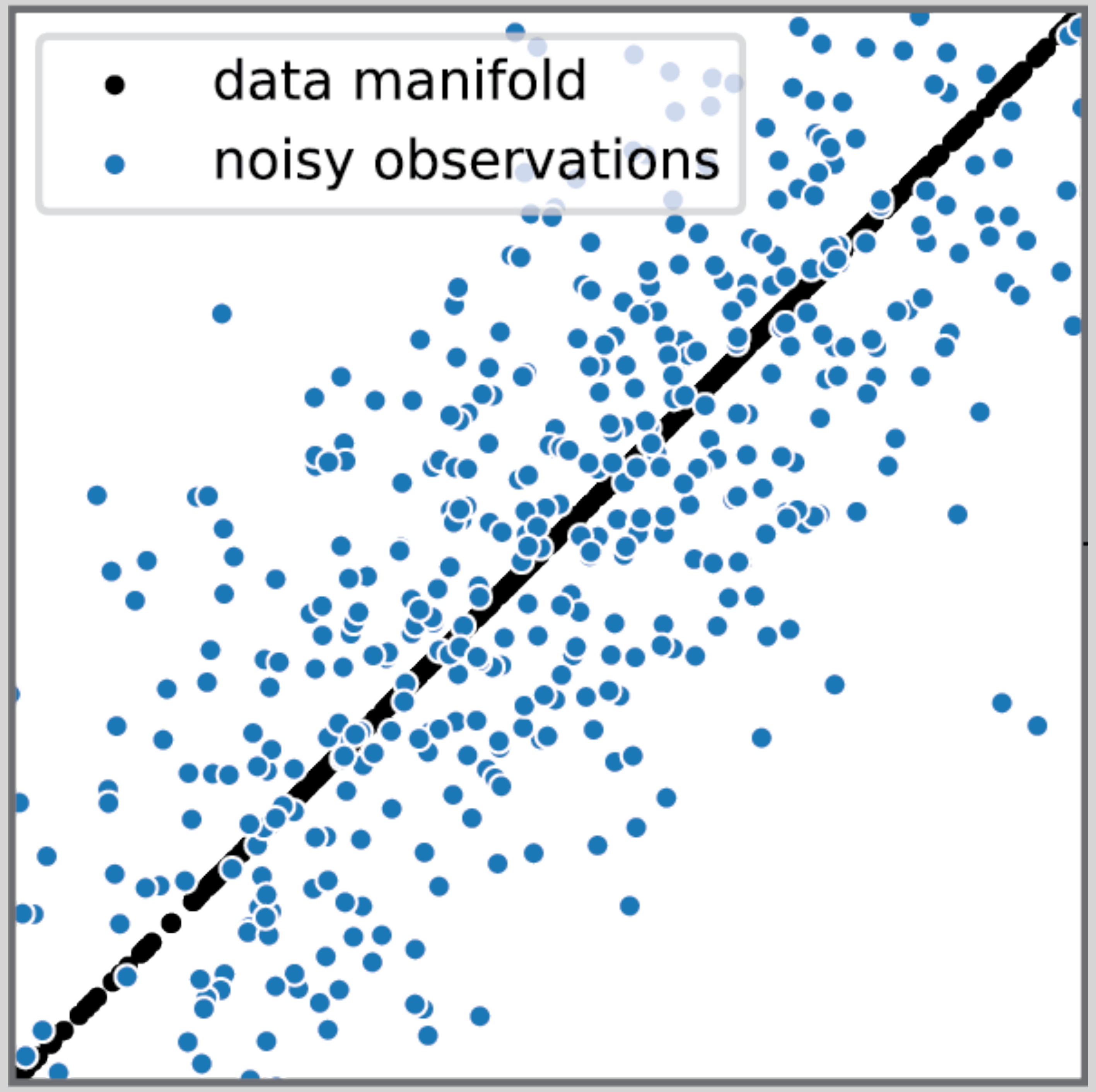


$$\text{Loss} = \left\langle \|x - y\|^2 \right\rangle - \frac{N^2}{2} D_{\text{IPR}} \quad \text{s.t.} \quad \left\langle y_i^2 \right\rangle = 1$$

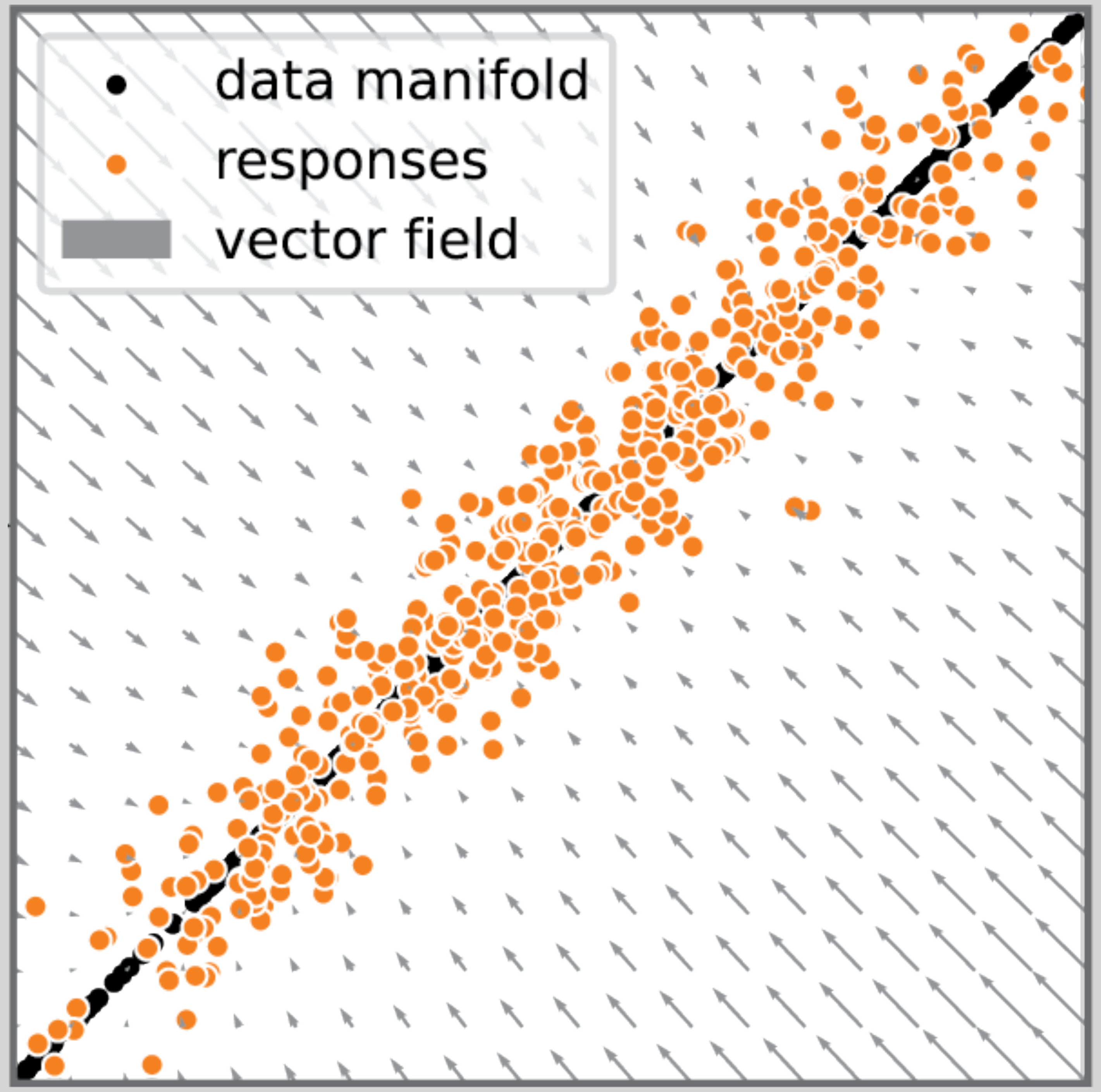
Duong *et al.* COSYNE 2026

Learning from noisy observations

Inputs: noisy observations x



Output: low-dim responses y



$$\boxed{D_{\mathrm{IPR}} = \frac{\mathrm{Tr}(C_y^2)}{\mathrm{Tr}(C_y)^2} = \frac{\sum_i \lambda_i^2}{(\sum_i \lambda_i)^2}}$$

$$\boxed{\mathrm{Loss} = \left\langle \|x - y\|^2 \right\rangle - \frac{N^2}{2} \boxed{D_{\mathrm{IPR}}} \quad \text{s.t.} \quad \left\langle y_i^2 \right\rangle = 1}$$

Duong *et al.* COSYNE 2026

Relation to AM energy function

$$\boxed{\text{Loss} = \left\langle \|x - y\|^2 \right\rangle - \frac{N^2}{2} D_{\text{IPR}} \quad \text{s.t.} \quad \left\langle y_i^2 \right\rangle = 1}$$

Hopfield 1984; Pehlevan *et al.* 2018

Relation to AM energy function

minimizing the loss
with respect to y

$$\boxed{\text{Loss} = \left\langle \|x - y\|^2 \right\rangle - \frac{N^2}{2} D_{\text{IPR}} \quad \text{s.t.} \quad \left\langle y_i^2 \right\rangle = 1}$$

Hopfield 1984; Pehlevan *et al.* 2018

Relation to AM energy function

minimizing the loss
with respect to y

$$\text{Loss} = \left\langle \|x - y\|^2 \right\rangle - \frac{N^2}{2} D_{\text{IPR}} \quad \text{s.t.} \quad \left\langle y_i^2 \right\rangle = 1$$

is equivalent to the
nested optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\text{Hopfield}} \right\rangle + \frac{1}{2} \|W\|_F^2 - \|c\|_1 \right\}$$

Hopfield 1984; Pehlevan *et al.* 2018

Relation to AM energy function

minimizing the loss
with respect to y

$$\boxed{\text{Loss} = \left\langle \|x - y\|^2 \right\rangle - \frac{N^2}{2} D_{\text{IPR}} \quad \text{s.t.} \quad \left\langle y_i^2 \right\rangle = 1}$$

is equivalent to the
nested optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\text{Hopfield}} \right\rangle + \frac{1}{2}\|W\|_F^2 - \|c\|_1 \right\}$$

$$E_{\text{Hopfield}} = -2\langle x, y \rangle - y^\top W y + \sum_i c_i y_i^2$$

Hopfield 1984; Pehlevan *et al.* 2018

Relation to AM energy function

minimizing the loss
with respect to y

$$\text{Loss} = \boxed{\left\langle \|x-y\|^2 \right\rangle} - \frac{N^2}{2} D_{\text{IPR}} \quad \text{s.t.} \quad \left\langle y_i^2 \right\rangle = 1$$

is equivalent to the
nested optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\text{Hopfield}} \right\rangle + \frac{1}{2}\|W\|_F^2 - \|c\|_1 \right\}$$

$$E_{\text{Hopfield}} = \boxed{-2\langle x, y\rangle} - y^\top W y + \sum_i c_i y_i^2$$

Hopfield 1984; Pehlevan *et al.* 2018

Relation to AM energy function

minimizing the loss
with respect to y

$$\text{Loss} = \left\langle \|x - y\|^2 \right\rangle \boxed{- \frac{N^2}{2} D_{\text{IPR}}} \quad \text{s.t.} \quad \left\langle y_i^2 \right\rangle = 1$$

is equivalent to the
nested optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\text{Hopfield}} \right\rangle + \boxed{\frac{1}{2} \|W\|_F^2} - \|c\|_1 \right\}$$

$$E_{\text{Hopfield}} = -2\langle x, y \rangle \boxed{- y^\top W y} + \sum_i c_i y_i^2$$

Under the capacity constraint, $N^2 D_{\text{IPR}} = \text{Tr}(C_y^2) = \max_W \text{Tr}\left(2WC_y - W^2\right)$

Hopfield 1984; Pehlevan *et al.* 2018

Relation to AM energy function

minimizing the loss with respect to y

$$\mathrm{Loss} = \left\langle \|x - y\|^2 \right\rangle - \frac{N^2}{2} D_{\mathrm{IPR}} \quad \text{s.t.} \quad \boxed{\langle y_i^2 \rangle = 1}$$

is equivalent to the nested optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\mathrm{Hopfield}} \right\rangle + \frac{1}{2}\|W\|_F^2 \boxed{- \|c\|_1} \right\}$$

$$E_{\mathrm{Hopfield}} = -2\langle x, y\rangle - y^\top W y + \boxed{\sum_i c_i y_i^2}$$

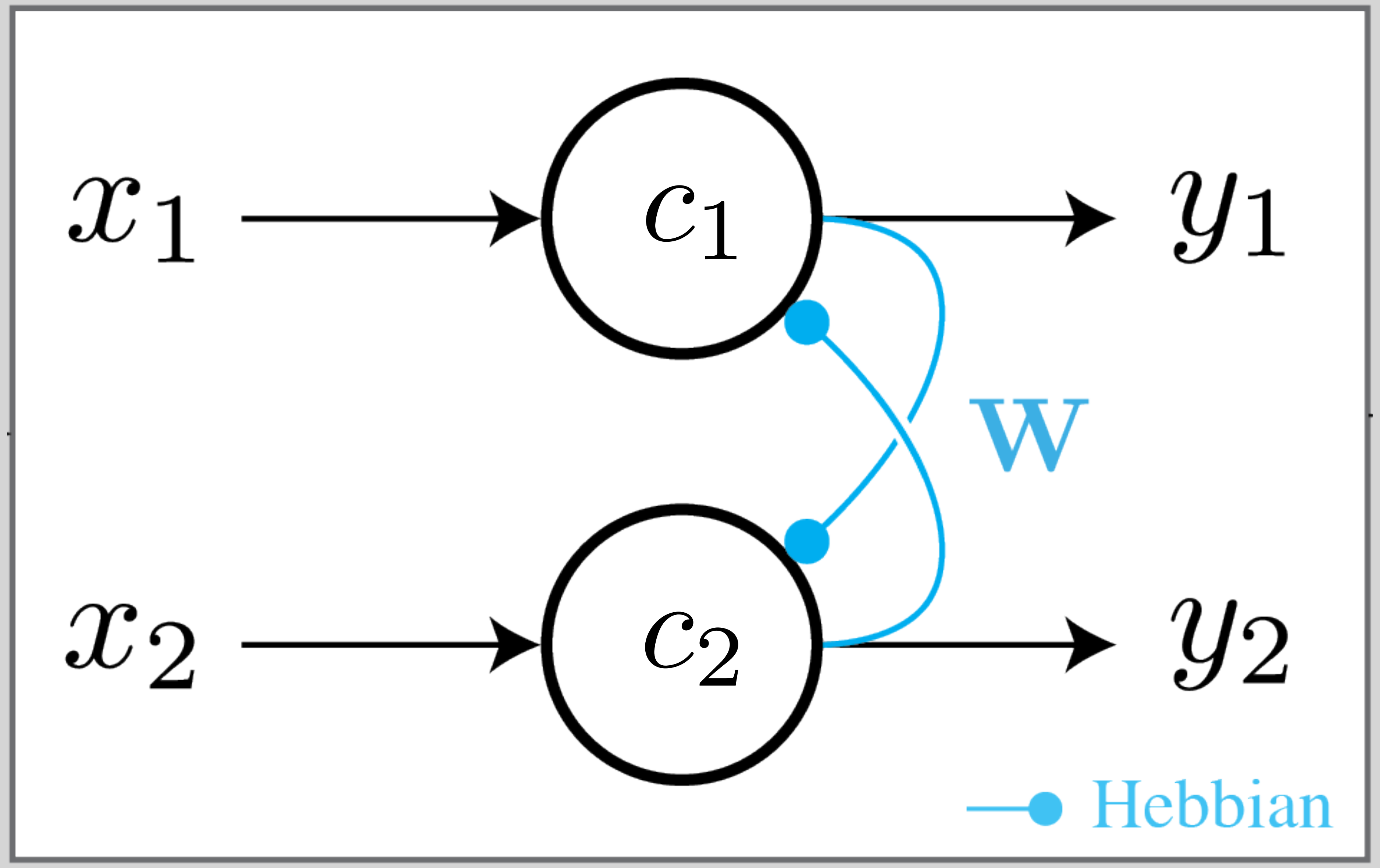
Lagrange multipliers $c = (c_1, \ldots, c_N)$ enforce the capacity constraints

Hopfield 1984; Pehlevan *et al.* 2018

Two-timescale optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\text{Hopfield}} \right\rangle + \frac{1}{2}\|W\|_F^2 - \|c\|_1 \right\}$$

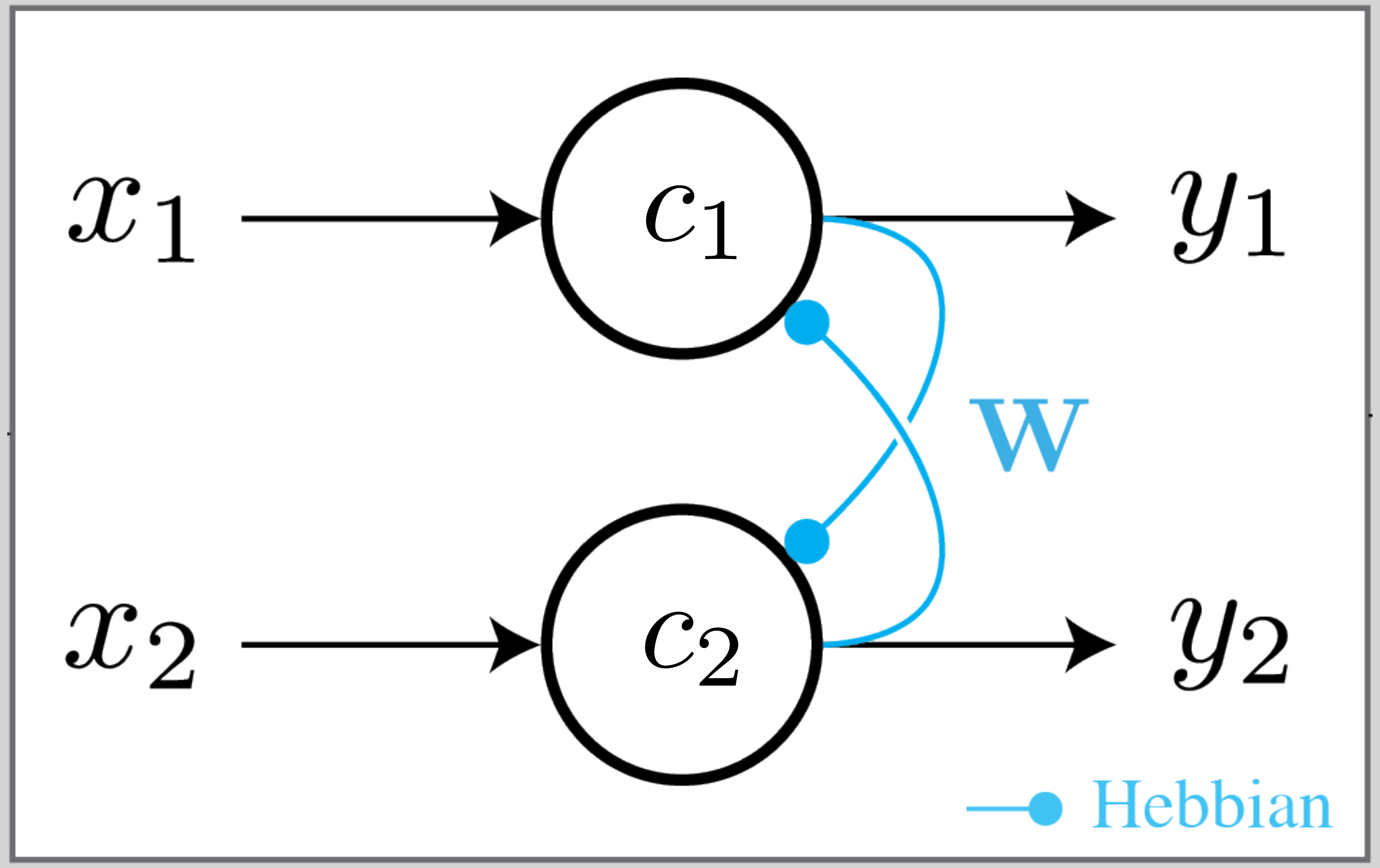
$$E_{\text{Hopfield}} = -2\langle x, y\rangle - y^\top W y + \sum_i c_i y_i^2$$



Two-timescale optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\text{Hopfield}} \right\rangle + \frac{1}{2}\|W\|_F^2 - \|c\|_1 \right\}$$

$$E_{\text{Hopfield}} = -2\langle x, y\rangle - y^\top W y + \sum_i c_i y_i^2$$

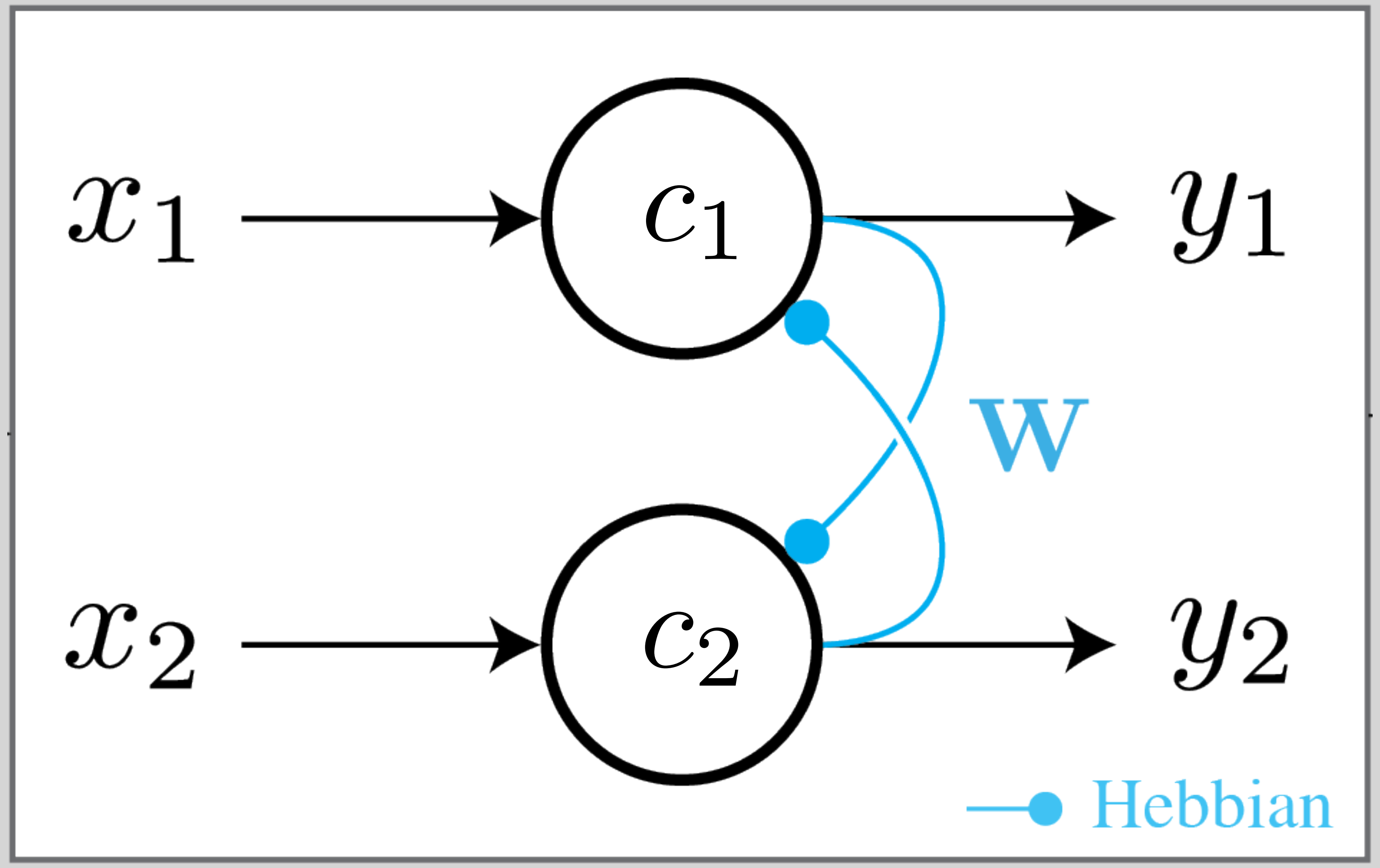


Two-timescale gradient-optimization:

Two-timescale optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\text{Hopfield}} \right\rangle + \frac{1}{2}\|W\|_F^2 - \|c\|_1 \right\}$$

$$E_{\text{Hopfield}} = -2\langle x, y\rangle - y^\top W y + \sum_i c_i y_i^2$$



Two-timescale gradient-optimization:

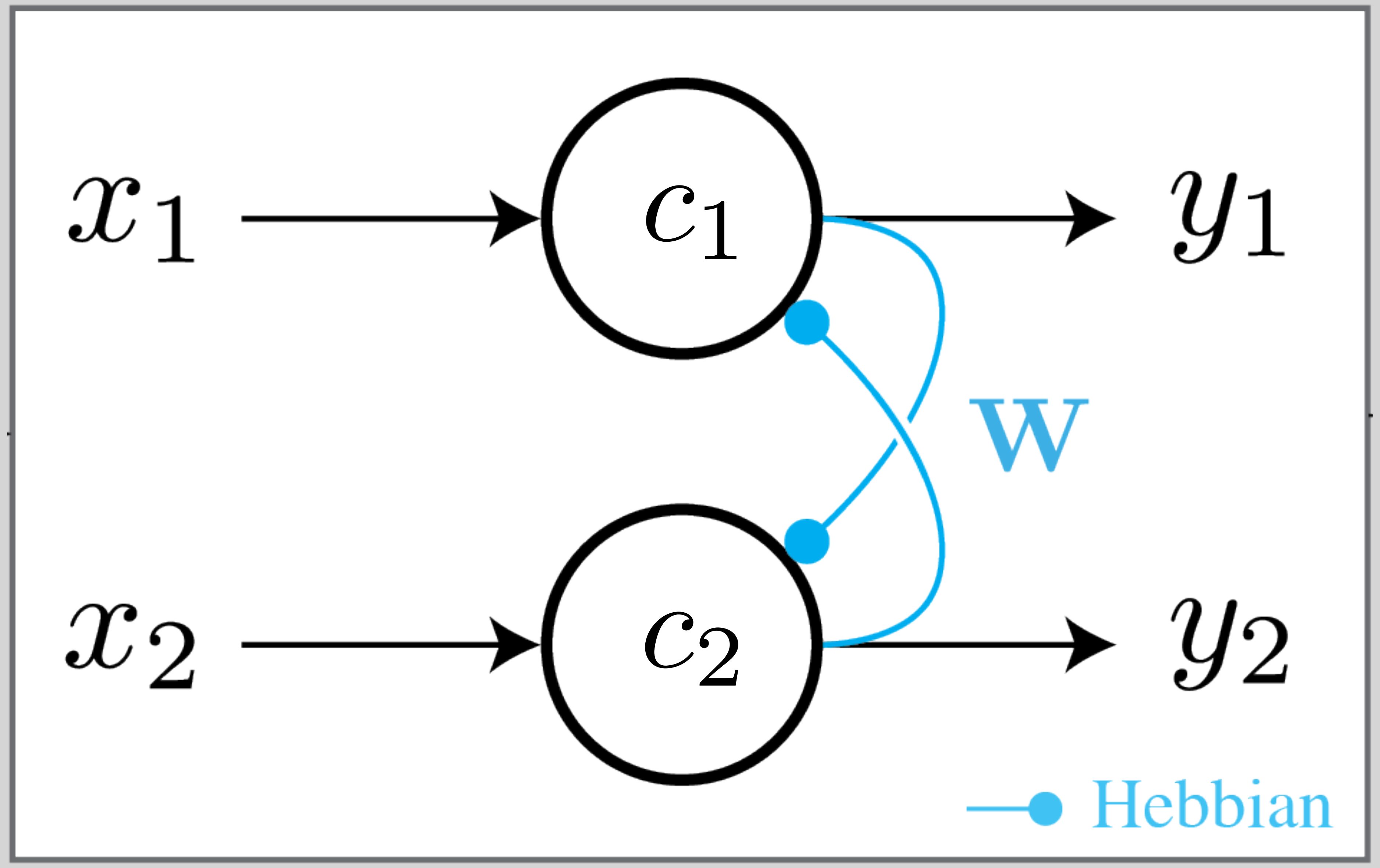
1. Neural dynamics descend $\nabla_y E_{\text{Hopfield}}$:

$$dy/dt = x + Wy - c \circ y$$

Two-timescale optimization

$$\min_W \max_c \left\{ \left\langle \min_y E_{\text{Hopfield}} \right\rangle + \frac{1}{2}\|W\|_F^2 - \|c\|_1 \right\}$$

$$E_{\text{Hopfield}} = -2\langle x, y\rangle - y^\top W y + \sum_i c_i y_i^2$$



Two-timescale gradient-optimization:

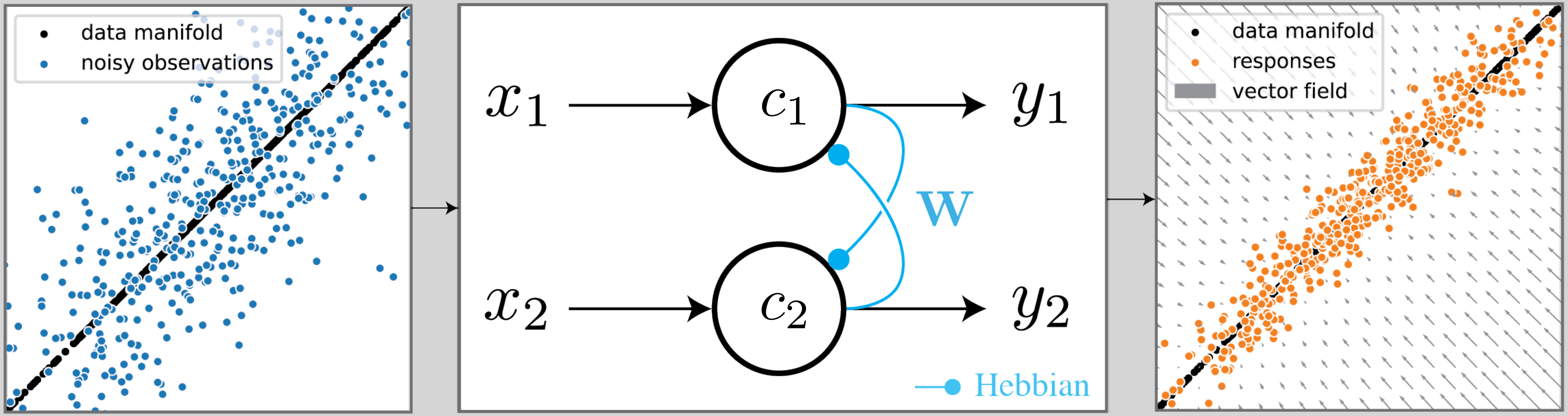
1. Neural dynamics descend $\nabla_y E_{\text{Hopfield}}$:

$$dy/dt = x + Wy - c \circ y$$

2. Hebbian plasticity and local leak updates:

$$\Delta W_{ij} \propto y_i y_j - W_{ij} \qquad \Delta c_i \propto y_i^2 - 1$$

Learning a line attractor



$$dy/dt = x + Wy - c \circ y$$

$$\Delta W_{ij} \propto y_i y_j - W_{ij} \qquad \Delta c_i \propto y_i^2 - 1$$

Kernel IPR and dense AM energy

Krotov & Hopfield 2016, Demircigil *et al.* 2017, Ramsauer *et al.* 2021

Kernel IPR and dense AM energy

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathscr{H}}$$

Krotov & Hopfield 2016, Demircigil *et al.* 2017, Ramsauer *et al.* 2021

Kernel IPR and dense AM energy

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathscr{H}}$$

minimizing the loss
with respect to y

$$\boxed{\mathrm{Loss} = -2\,\langle K(x, y) \rangle - \frac{1}{2} D^{\Phi}_{\mathrm{IPR}} \quad \text{s.t.} \quad \langle K(y, y) \rangle = 1}$$

Krotov & Hopfield 2016, Demircigil *et al.* 2017, Ramsauer *et al.* 2021

Kernel IPR and dense AM energy

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}$$

minimizing the loss
with respect to y

$$\boxed{\text{Loss} = -2\,\langle K(x, y) \rangle - \frac{1}{2} D^{\Phi}_{\text{IPR}} \quad \text{s.t.} \quad \langle K(y, y) \rangle = 1}$$

is equivalent to the
nested optimization

$$\boxed{\min_{A} \max_{g} \left\{ \left\langle \min_{y} E_{\text{dense}} \right\rangle + \frac{1}{2} \|A\|^2_{\text{HS}} - g \right\}}$$

Krotov & Hopfield 2016, Demircigil *et al.* 2017, Ramsauer *et al.* 2021

Kernel IPR and dense AM energy

$$K(x,y) = \langle \Phi(x), \Phi(y) \rangle_{\mathscr{H}}$$

minimizing the loss
with respect to y

$$\text{Loss} = -2\,\langle K(x,y) \rangle - \frac{1}{2} D^{\Phi}_{\text{IPR}} \quad \text{s.t.} \quad \langle K(y,y) \rangle = 1$$

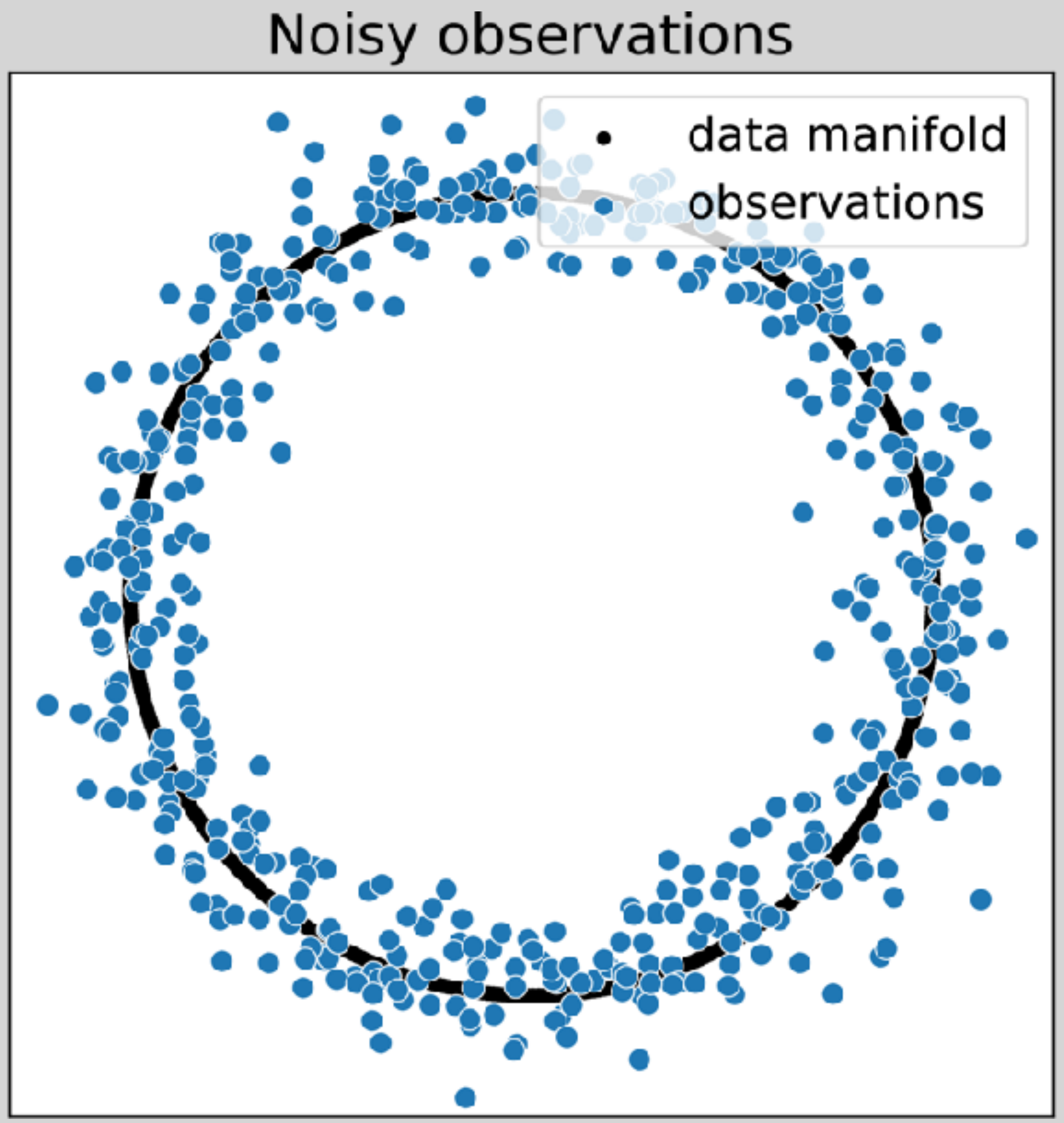
is equivalent to the
nested optimization

$$\min_A \max_g \left\{ \left\langle \min_y E_{\text{dense}} \right\rangle + \frac{1}{2} \|A\|^2_{\text{HS}} - g \right\}$$

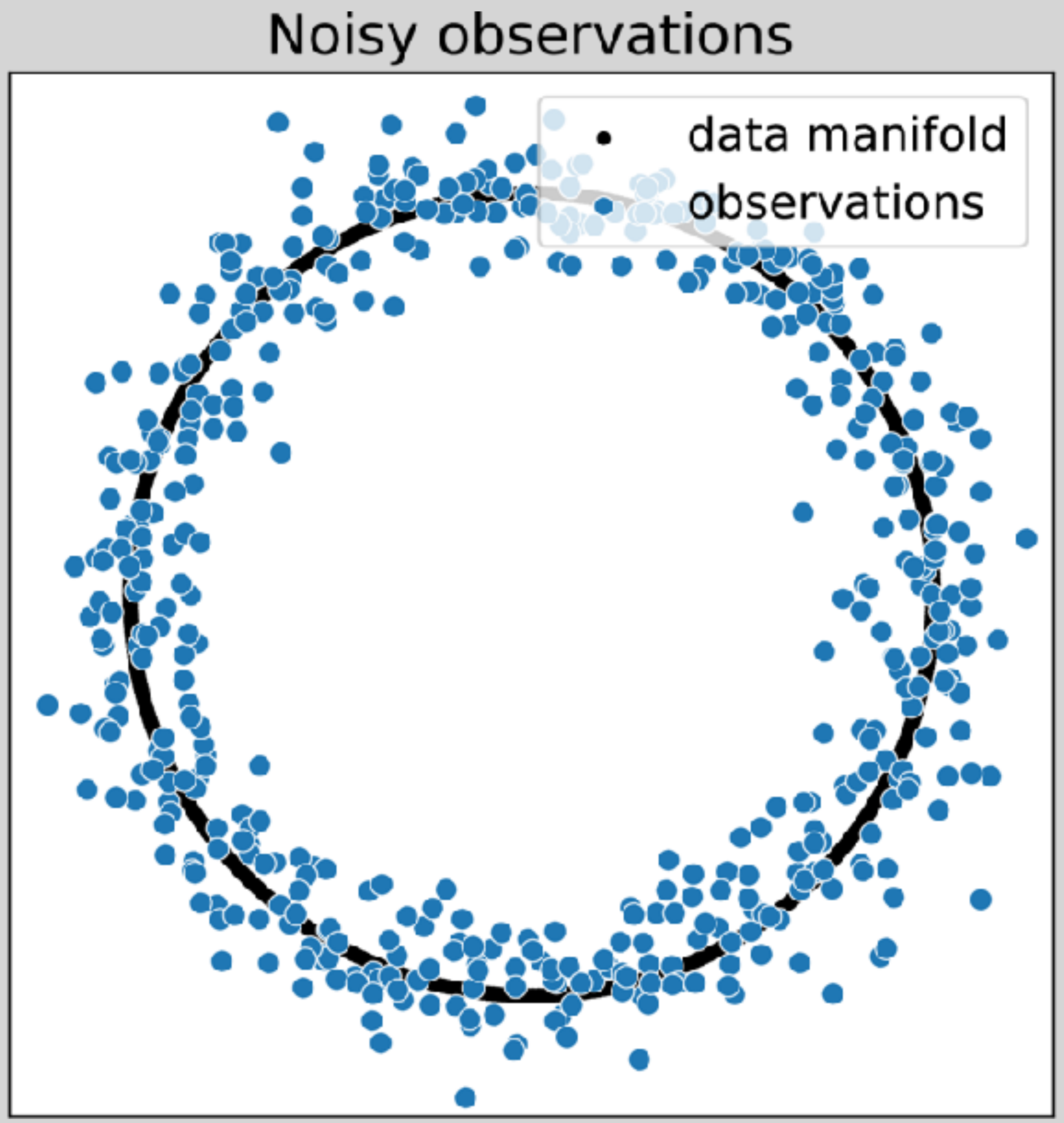
$$E_{\text{dense}} = -2K(x,y) - \langle \Phi(y), A\Phi(y) \rangle_{\mathscr{H}} + gK(y,y)$$

Krotov & Hopfield 2016, Demircigil *et al.* 2017, Ramsauer *et al.* 2021

Learning a ring attractor

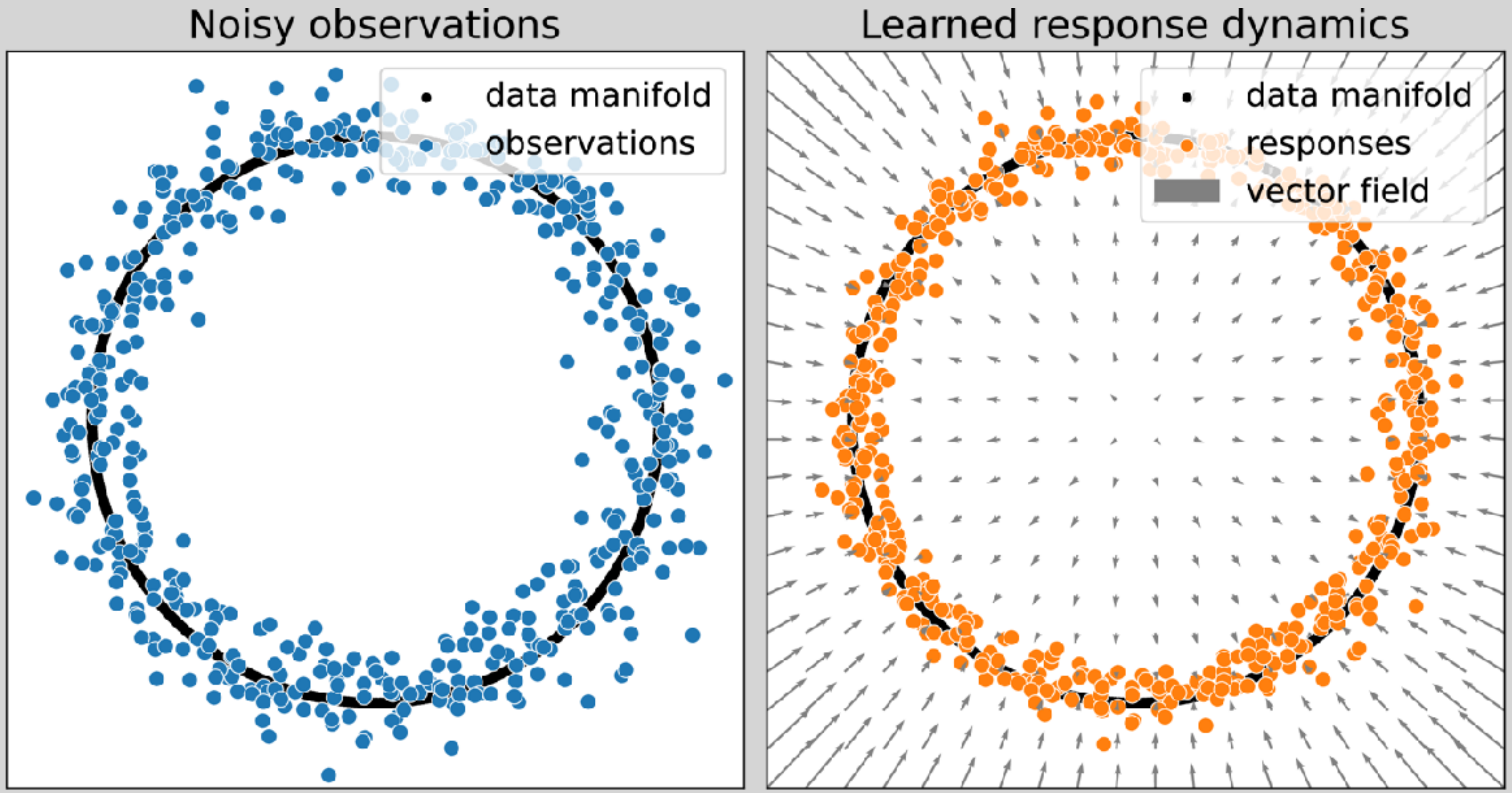


Learning a ring attractor



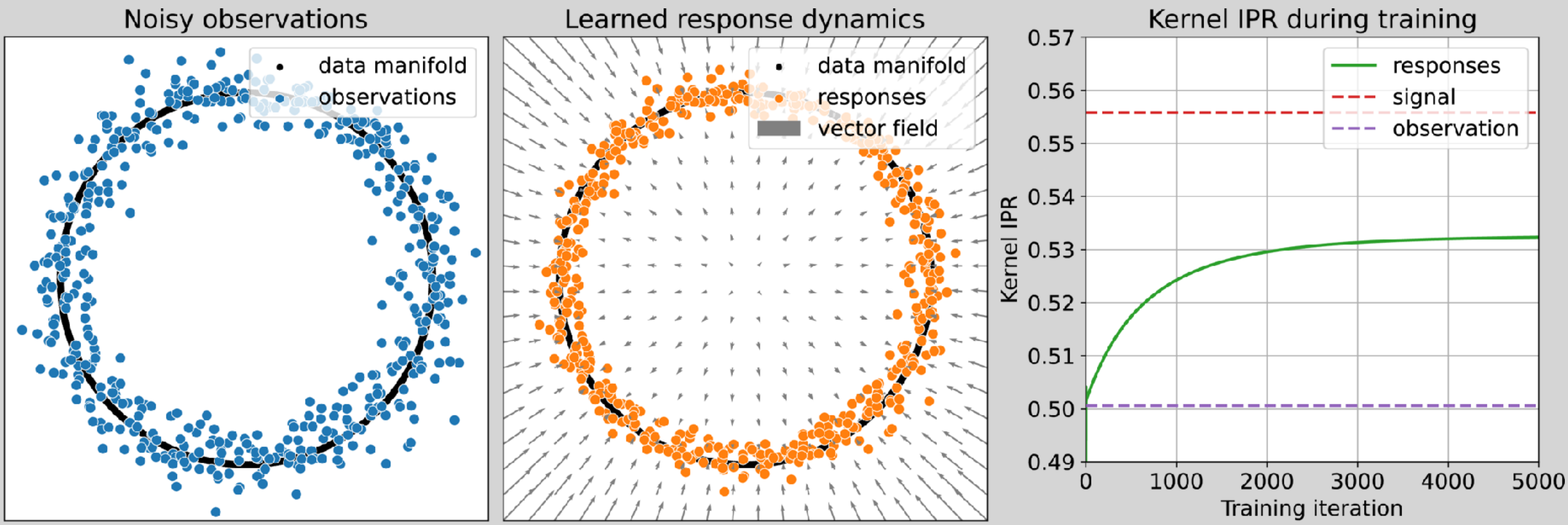
Quadratic kernel: $\Phi(x) = (2x_1^2 - 1, 2x_2^2 - 1, 2x_1x_2)$

Learning a ring attractor



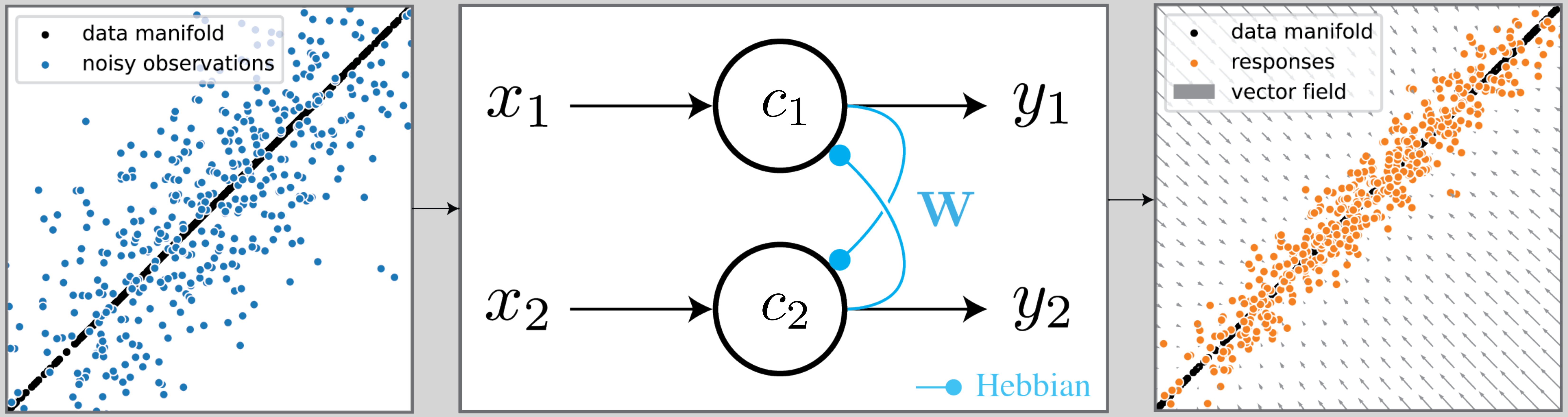
Quadratic kernel: $\Phi(x) = (2x_1^2 - 1, 2x_2^2 - 1, 2x_1x_2)$

Learning a ring attractor



Quadratic kernel: $\Phi(x) = (2x_1^2 - 1, 2x_2^2 - 1, 2x_1x_2)$

Summary

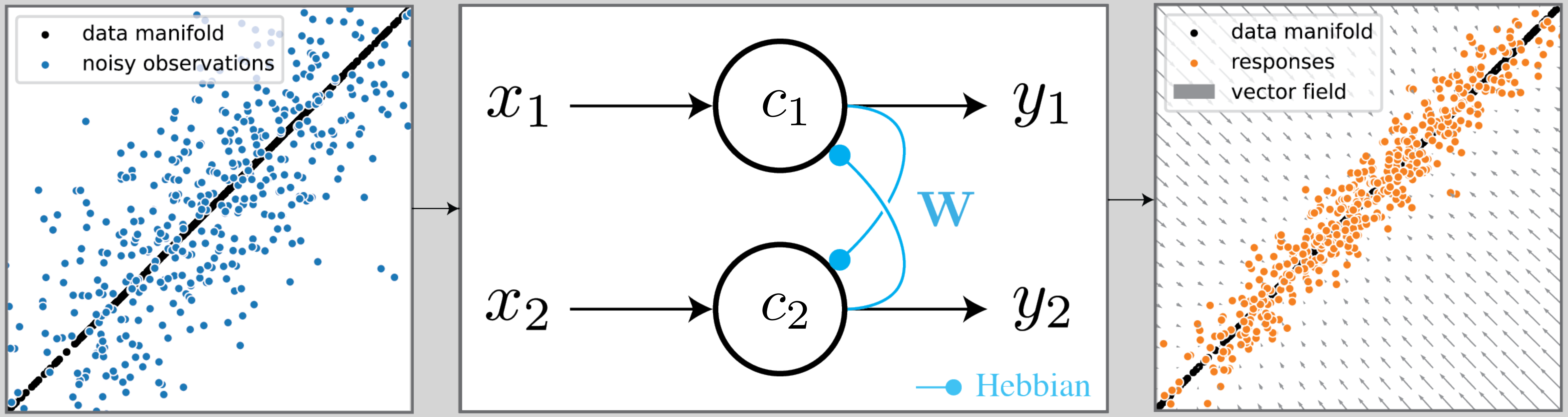


RICE NEUROENGINEERING INITIATIVE

Baylor College of Medicine

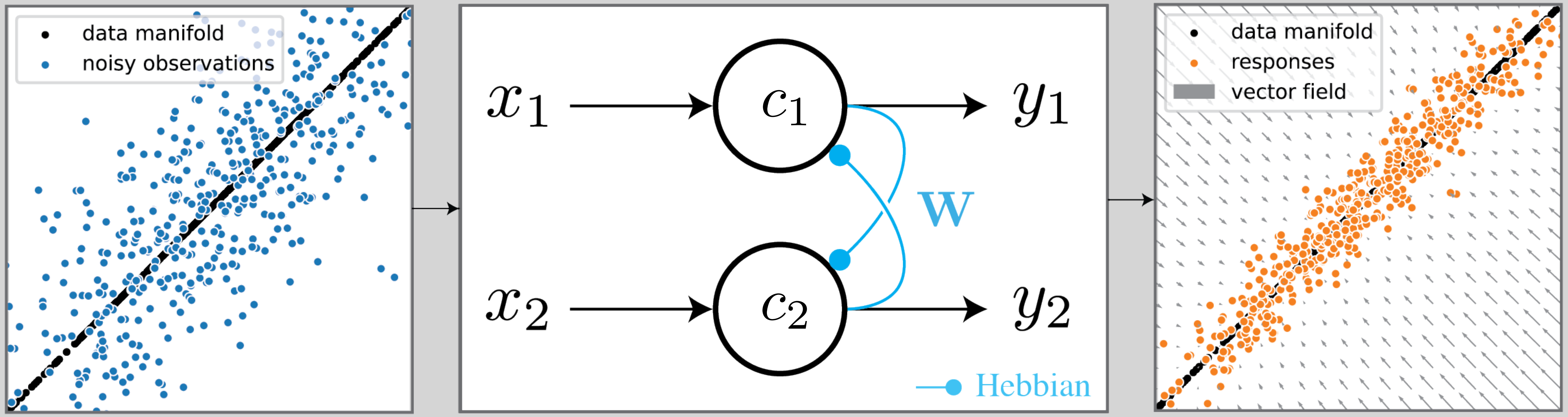
Summary

- AM energy function is related to objectives for <u>reducing dimensionality</u>



Summary

- AM energy function is related to objectives for reducing dimensionality
- Relation to blind denoising and work connecting AM models to denoising?
  (Hoover *et al.* 2023; Raya & Ambrogioni 2023; Pham *et al.* 2024; Smart *et al.* 2025)

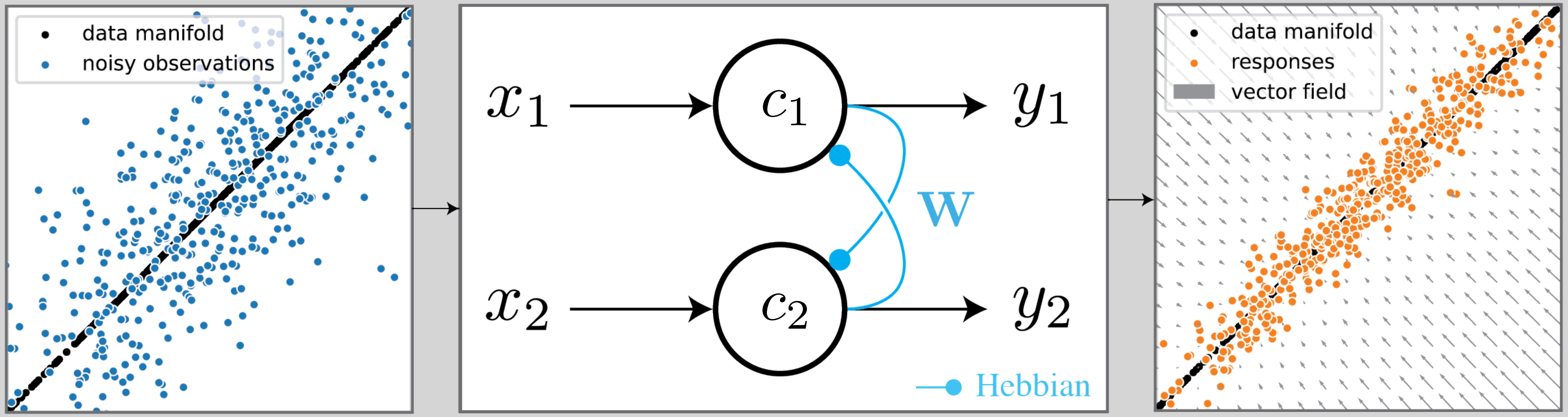


RICE NEUROENGINEERING INITIATIVE

Baylor College of Medicine

Summary

- AM energy function is related to objectives for reducing dimensionality
- Relation to blind denoising and work connecting AM models to denoising?
  (Hoover *et al.* 2023; Raya & Ambrogioni 2023; Pham *et al.* 2024; Smart *et al.* 2025)
- Biological connections? Noise reduction and memory storage can be implemented using similar circuit dynamics, learning rules and architectures.



RICE NEUROENGINEERING INITIATIVE

Baylor College of Medicine