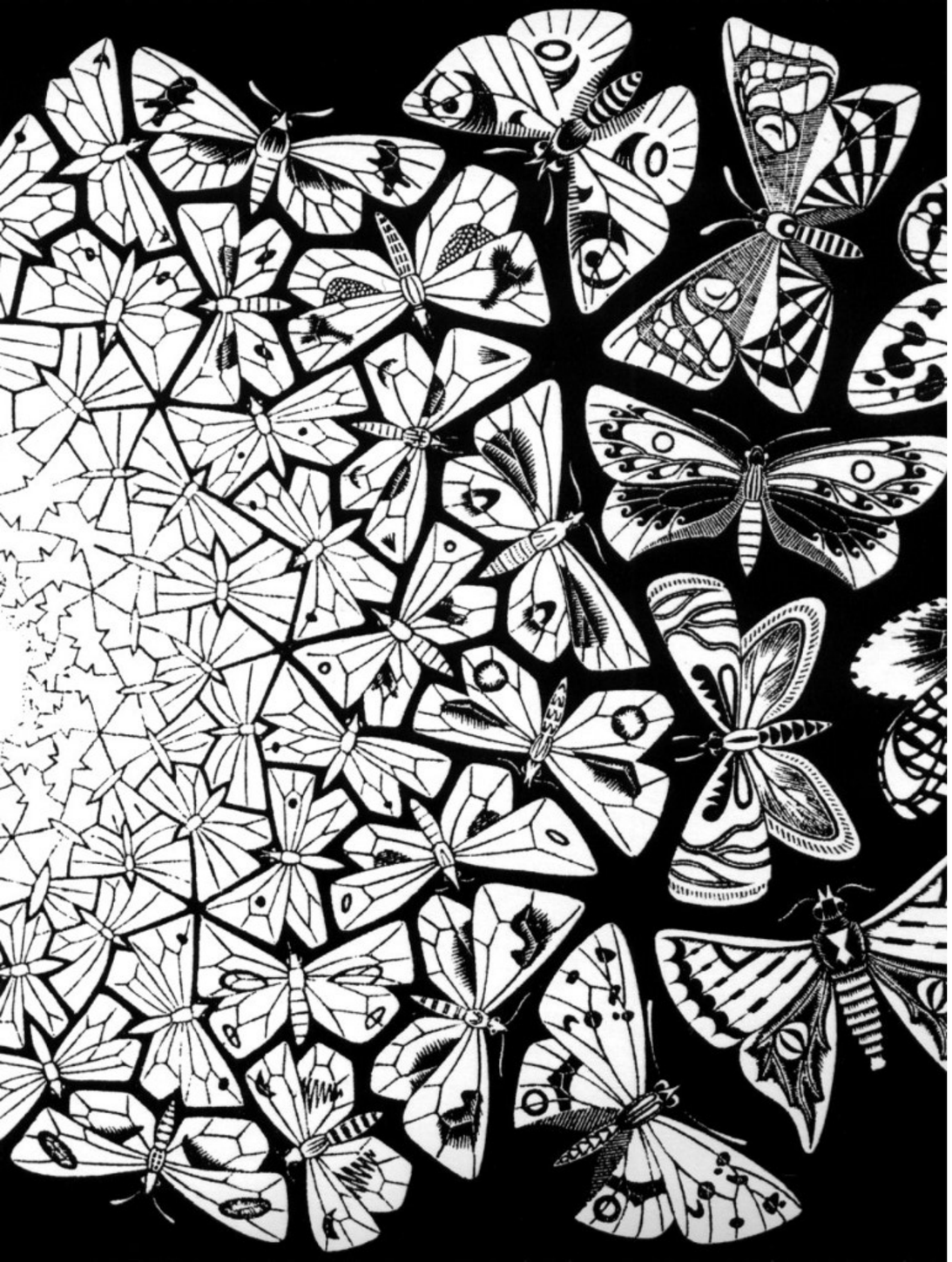
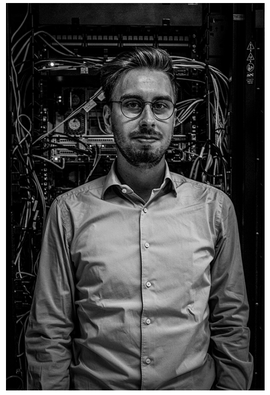


Soft matter physics as a joint between deep learning and renormalisation group

Raffaello Potestio
University of Trento | Physics Department
TIFPA | Trento Institute for Fundamental Physics and Applications

ECT* workshop *Machine learning and the Renormalization Group*
Trento, Italy | May 28, 2024

Acknowledgments



Current team members

Roberto Menichetti
Lorenzo Petrolli
Manuel Micheloni
Margherita Mele
Camilla Spreti

Previous team members

Marta Rigoli
Marco Giulini
Thomas Tarenzi
Raffaele Fiorentini
Giovanni Mattiotti

Statistical and Biological Physics group @UniTn

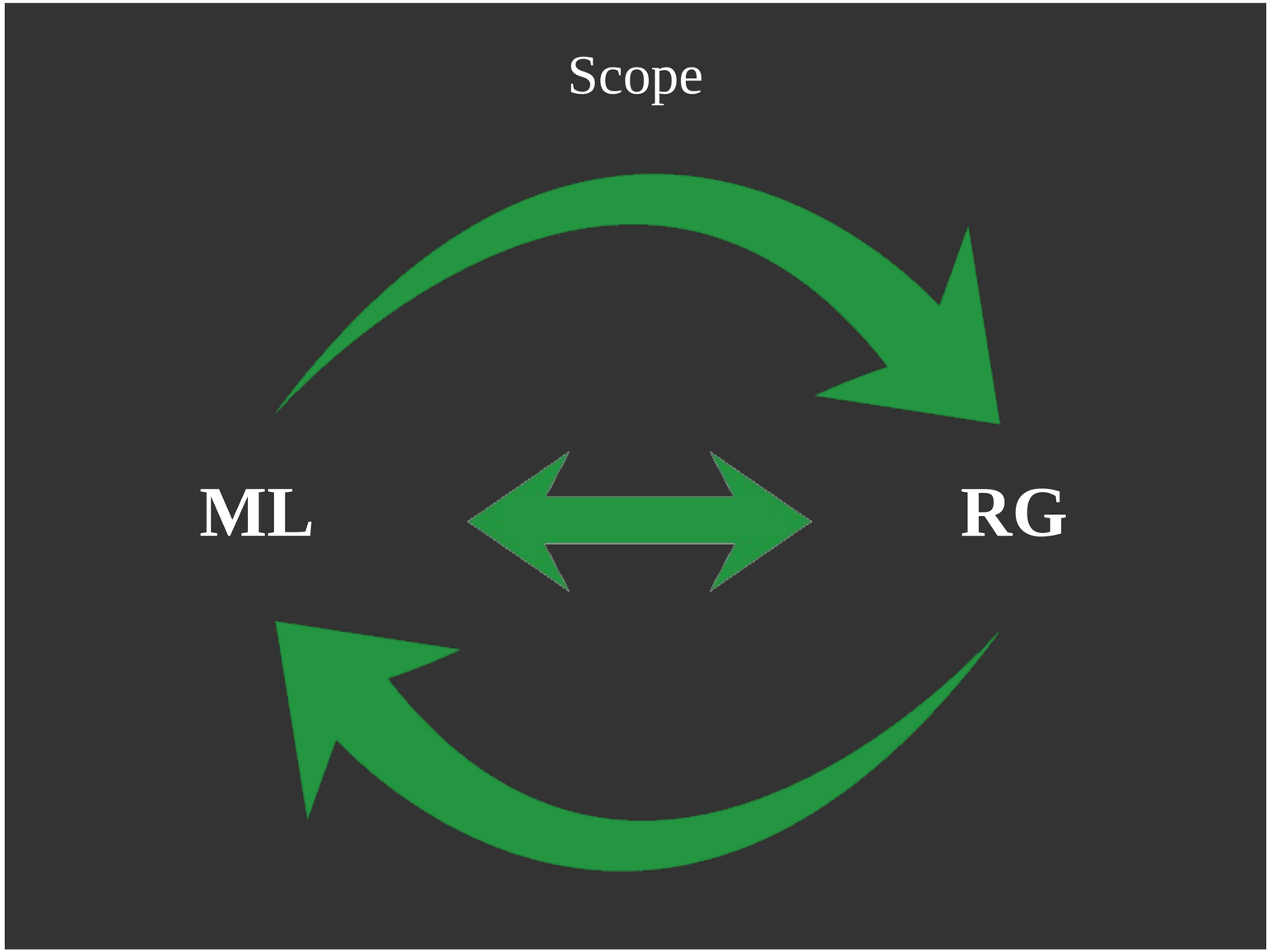
Gianluca Lattanzi
Luca Tubiana
Virginia Agostiniani

External collaborators

M. Scott Shell, UCSB, USA
Roi Holtzman, Weizmann Institute, Israel
Matteo Marsili, ICTP, Italy
Alessandro Ingrosso, ICTP, Italy

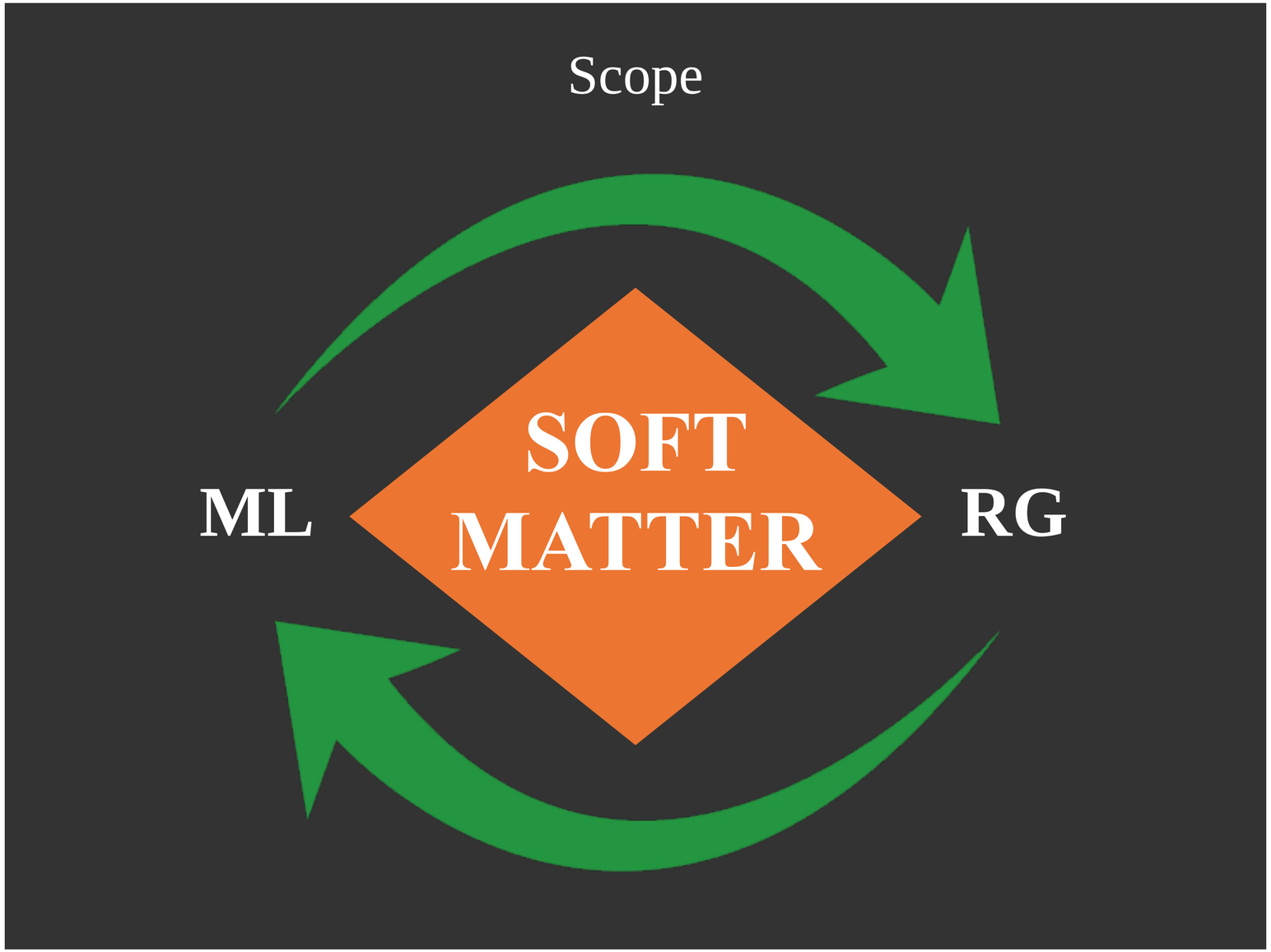
Funding

Framework



J. Urban, 2024 (edited)

Framework



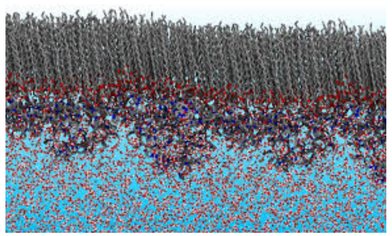
J. Urban, 2024 (edited)

What is soft matter?







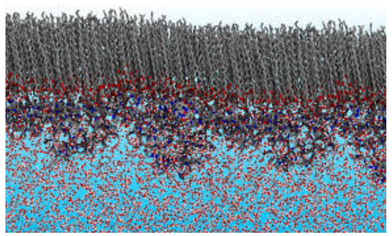



What is soft matter?









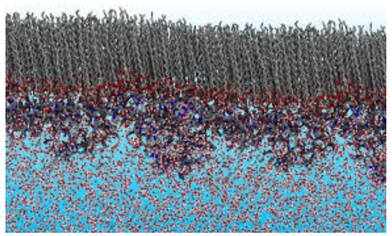

SOFT MATTER SYSTEMS
ARE DEFORMABLE
UPON STRESSES
OF THE ORDER OF MAGNITUDE
OF THERMAL FLUCTUATIONS

What is soft matter?










SOFT MATTER SYSTEMS
ARE DEFORMABLE
UPON STRESSES
OF THE ORDER OF MAGNITUDE
OF THERMAL FLUCTUATIONS

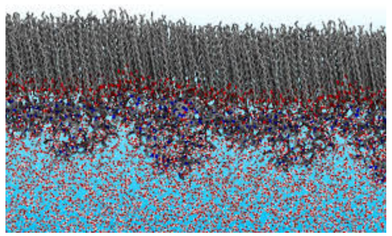
IN SOFT MATTER
PATTERNS EMERGE
THAT CANNOT BE PREDICTED
BY FIRST PRINCIPLES

What is soft matter?










SOFT MATTER SYSTEMS
ARE DEFORMABLE
UPON STRESSES
OF THE ORDER OF MAGNITUDE
OF THERMAL FLUCTUATIONS

IN SOFT MATTER
PATTERNS EMERGE
THAT CANNOT BE PREDICTED
BY FIRST PRINCIPLES

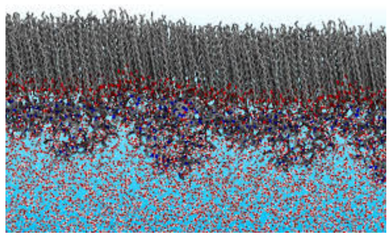
SOFT MATTER
IS MESOSCOPIC:
RELEVANT STRUCTURES
ARE MUCH LARGER
THAN THE CONSTITUENTS
BUT MUCH SMALLER
THAN THE WHOLE

What is soft matter?










SOFT MATTER SYSTEMS
ARE DEFORMABLE
UPON STRESSES
OF THE ORDER OF MAGNITUDE
OF THERMAL FLUCTUATIONS

SM is squishy

IN SOFT MATTER
PATTERNS EMERGE
THAT CANNOT BE PREDICTED
BY FIRST PRINCIPLES

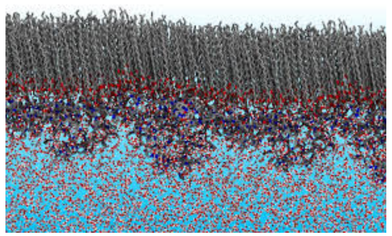
SOFT MATTER
IS MESOSCOPIC:
RELEVANT STRUCTURES
ARE MUCH LARGER
THAN THE CONSTITUENTS
BUT MUCH SMALLER
THAN THE WHOLE

What is soft matter?










SOFT MATTER SYSTEMS
ARE DEFORMABLE
UPON STRESSES
OF THE ORDER OF MAGNITUDE
OF THERMAL FLUCTUATIONS

SM is squishy

IN SOFT MATTER
PATTERNS EMERGE
THAT CANNOT BE PREDICTED
BY FIRST PRINCIPLES

SM is complex

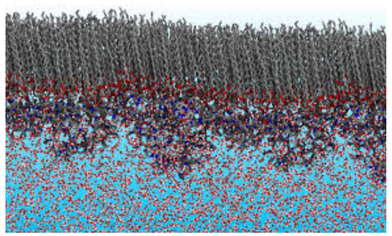
SOFT MATTER
IS MESOSCOPIC:
RELEVANT STRUCTURES
ARE MUCH LARGER
THAN THE CONSTITUENTS
BUT MUCH SMALLER
THAN THE WHOLE

What is soft matter?










SOFT MATTER SYSTEMS
ARE DEFORMABLE
UPON STRESSES
OF THE ORDER OF MAGNITUDE
OF THERMAL FLUCTUATIONS

SM is squishy

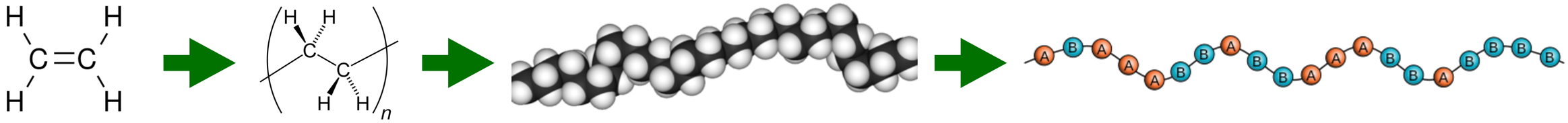
IN SOFT MATTER
PATTERNS EMERGE
THAT CANNOT BE PREDICTED
BY FIRST PRINCIPLES

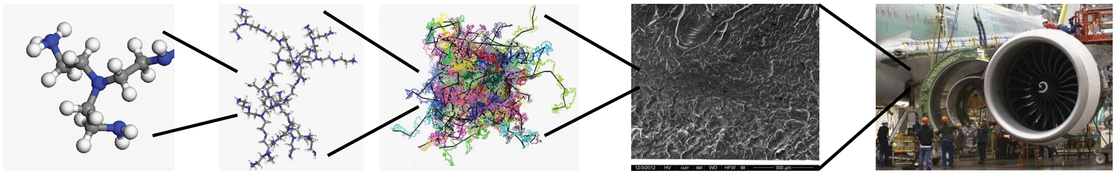
SM is complex

SOFT MATTER
IS MESOSCOPIC:
RELEVANT STRUCTURES
ARE MUCH LARGER
THAN THE CONSTITUENTS
BUT MUCH SMALLER
THAN THE WHOLE

SM is multiscale

Polymers: playground for scaling laws





monomer
1 nm

chain
10-100 nm

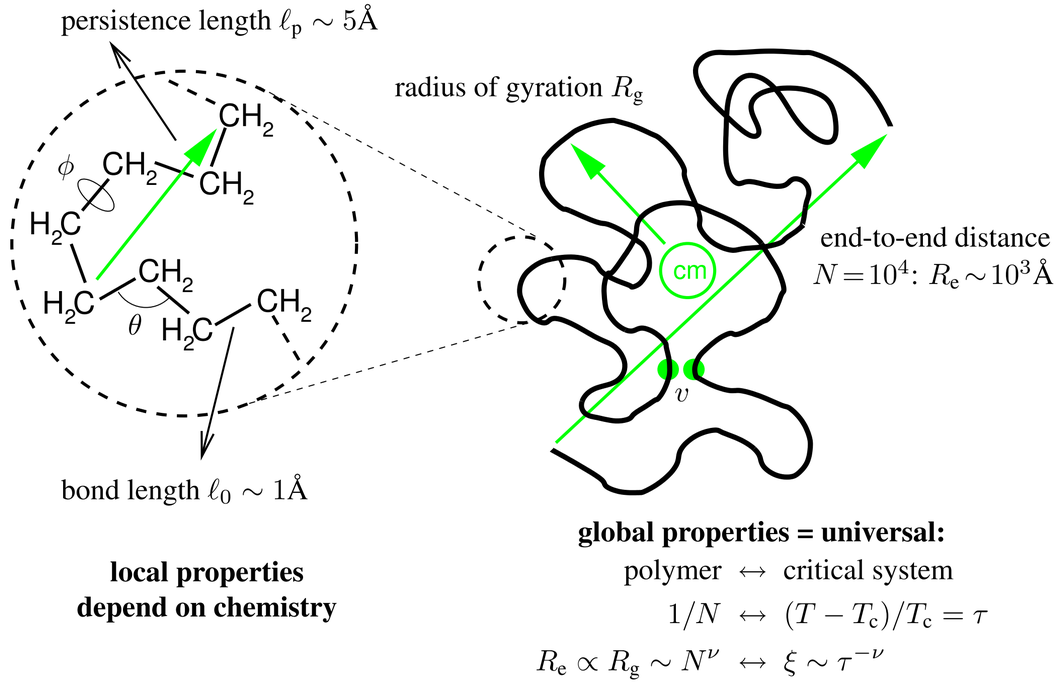
network
1-10 μm

bulk
1 mm

airplane
10 m

Scaling laws apply
Large-scale properties scale with system size through power laws

Problems are multi-scale
Interesting physics on a range of different scales, which interplay and affect one another



local properties depend on chemistry

global properties = universal:

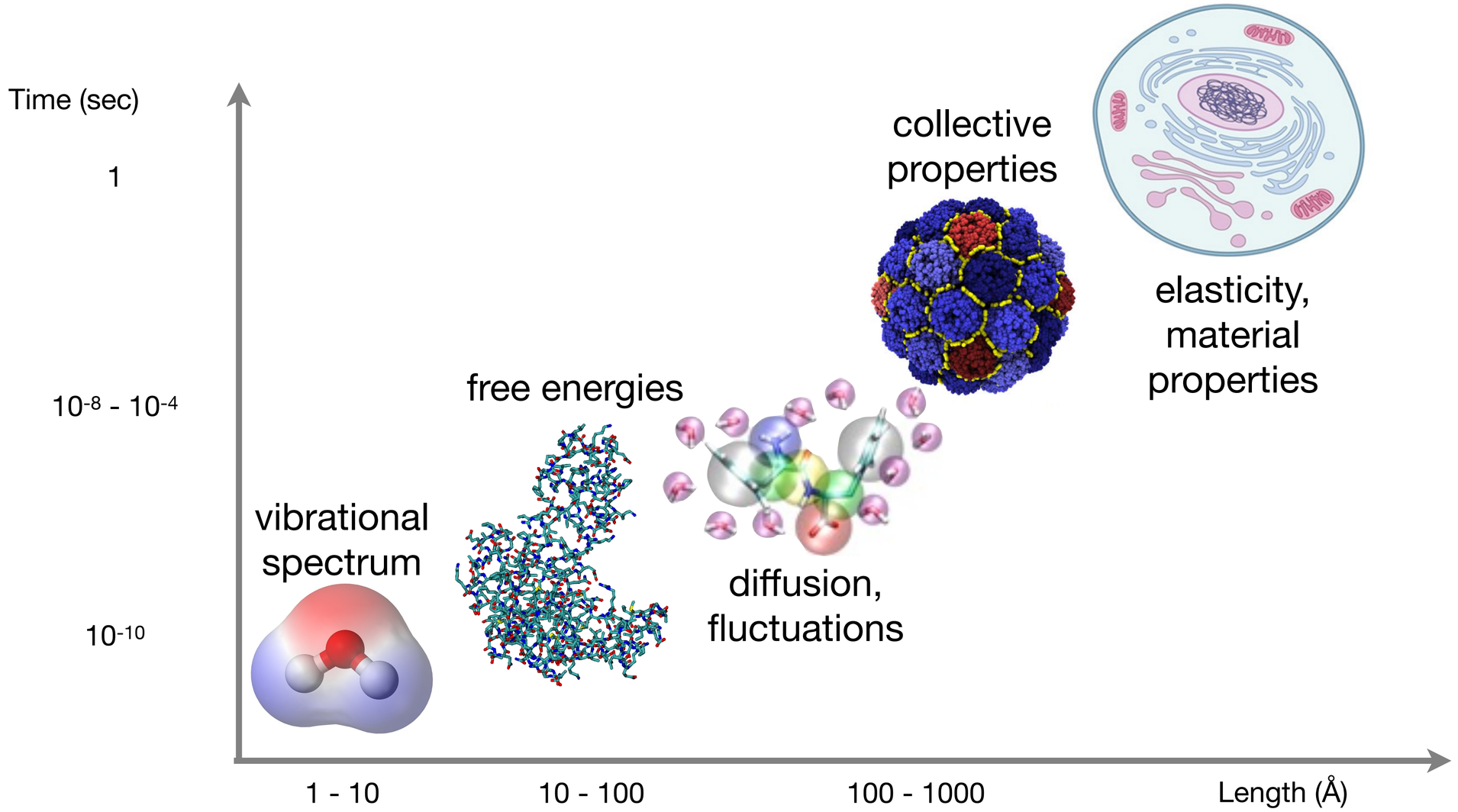
polymer $\leftrightarrow$ critical system

$1/N \leftrightarrow (T - T_c)/T_c = \tau$

$R_e \propto R_g \sim N^\nu \leftrightarrow \xi \sim \tau^{-\nu}$

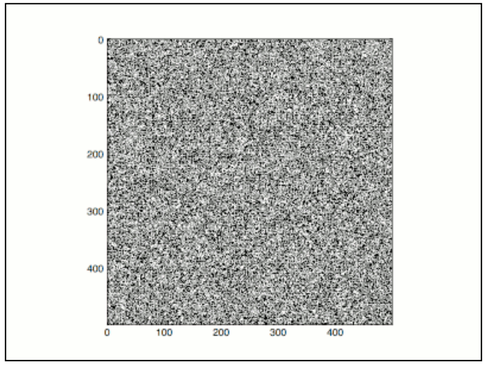
P.-J. De Gennes, *Scaling concepts in polymer physics*, 1979

The multi-scale challenge in SM physics



The Ising model

Stat-mech prototype for coarse-graining



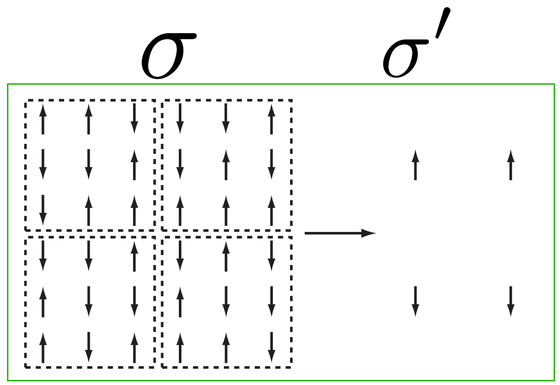
Up/down spins, first-neighbour coupling
Zero temperature phase transition in 1D
Finite temperature phase transition in 2D

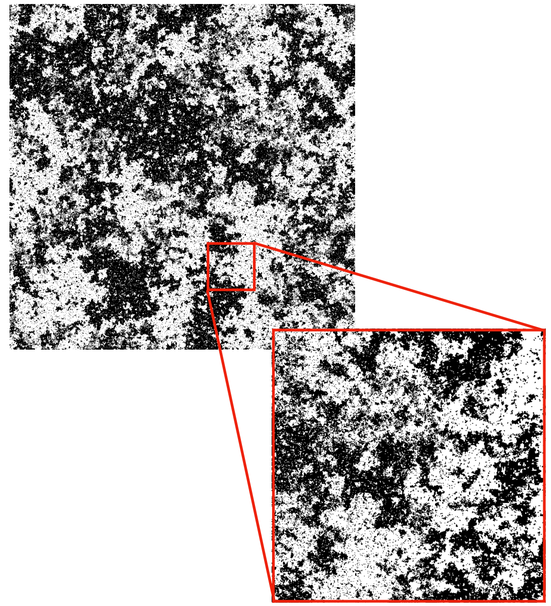
$$H = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i$$

Scale invariance
Correlation length diverges
No typical length scale, power law correlations

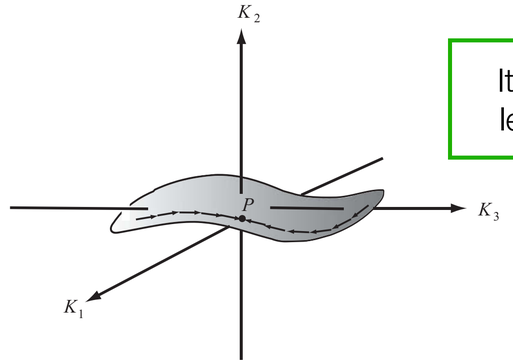
Coarse-graining
Spins can be lumped together to make a coarser model

Self-similarity
The coarse-grained system is described by
the same Hamiltonian of the fine-grained system





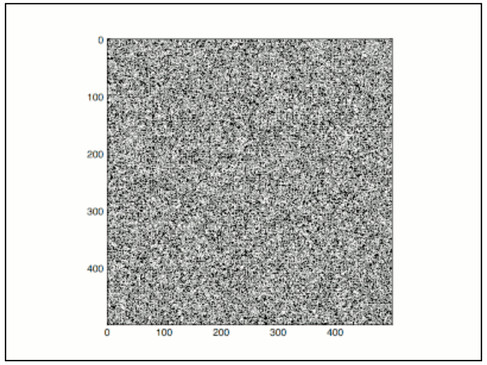
$$\mathrm{Tr}_{\sigma'} \mathrm{e}^{-\beta \mathcal{H}'_0(\{\sigma'\})} = \mathrm{Tr}_{\sigma} \mathrm{e}^{-\beta \mathcal{H}_0(\{\sigma\})}$$



Iterative application
leads to fixed point

The Ising model

Stat-mech prototype for coarse-graining



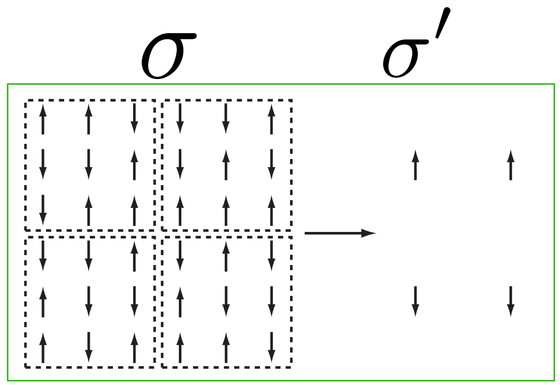
Up/down spins, first-neighbour coupling
Zero temperature phase transition in 1D
Finite temperature phase transition in 2D

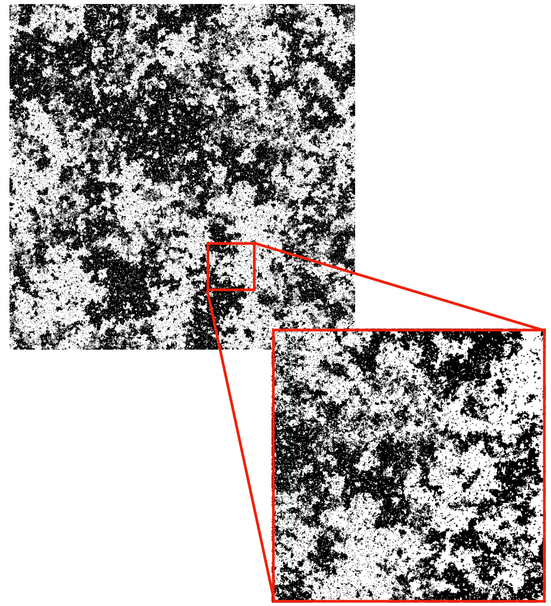
$$H = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i$$

Scale invariance
Correlation length diverges
No typical length scale, power law correlations

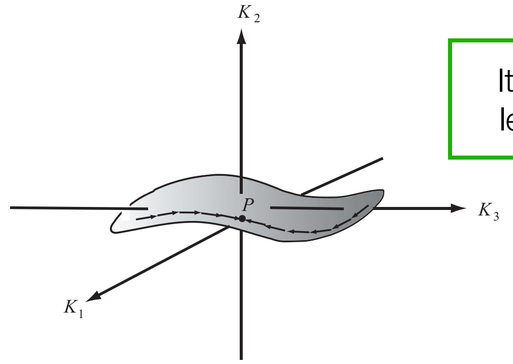
Coarse-graining
Spins can be lumped together to make a coarser model

Self-similarity
The coarse-grained system is described by
the same Hamiltonian of the fine-grained system



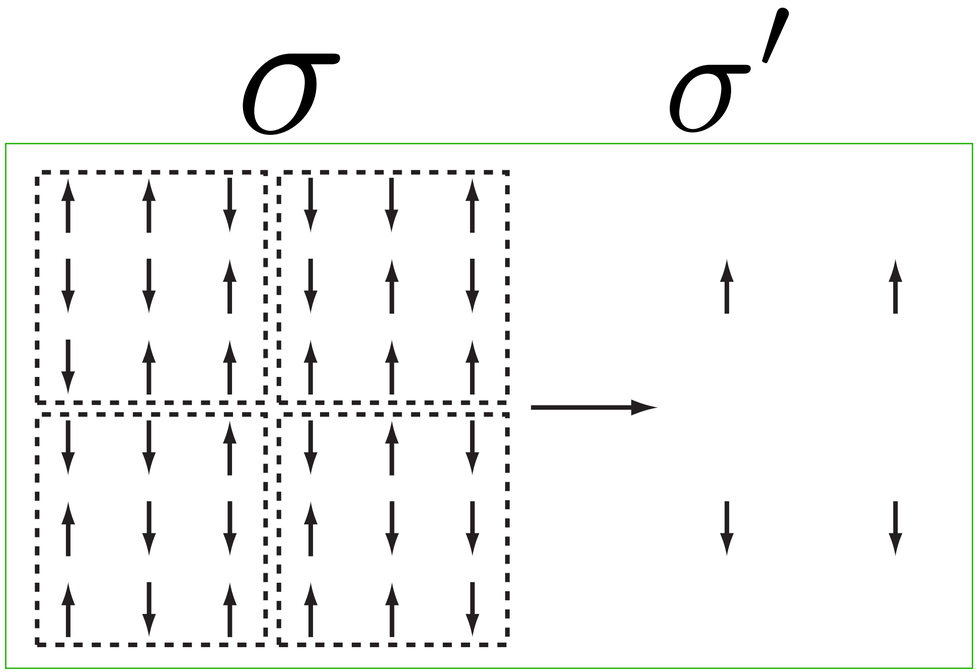


$$\mathrm{Tr}_{\sigma'} \mathrm{e}^{-\beta \mathcal{H}'_0(\{\sigma'\})} = \mathrm{Tr}_{\sigma} \mathrm{e}^{-\beta \mathcal{H}_0(\{\sigma\})}$$



Iterative application
leads to fixed point

Kadanoff's spin-block transformation



$$Q(N,T) = \sum_{\sigma_1} \cdots \sum_{\sigma_N} \mathrm{e}^{-\beta \mathcal{H}_0(\sigma_1,\ldots,\sigma_N)} \equiv \mathrm{Tr}_\sigma \mathrm{e}^{-\beta \mathcal{H}_0(\sigma_1,\ldots,\sigma_N)}. \tag{16.9.1}$$

The transformation function $T(\sigma';\sigma_1,...,\sigma_9)$ that yields the single spin σ' for each 3×3 block of 9 spin variables can be expressed mathematically as follows:

$$T(\sigma';\sigma_1,...,\sigma_9) = \begin{cases} 1 & \sigma' \sum_{i=1}^{9} \sigma_i > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{16.9.2}$$

$$\sum_{\sigma'=\pm 1} T(\sigma';\sigma_1,...,\sigma_9) = 1, \tag{16.9.3}$$

which means simply that only one of the two values of σ' can satisfy the block spin transformation rule. The new spin variables $\{\sigma'\}$ can now be used to define a new partition function. To see how this is done, let the Hamiltonian of the new lattice be defined according to

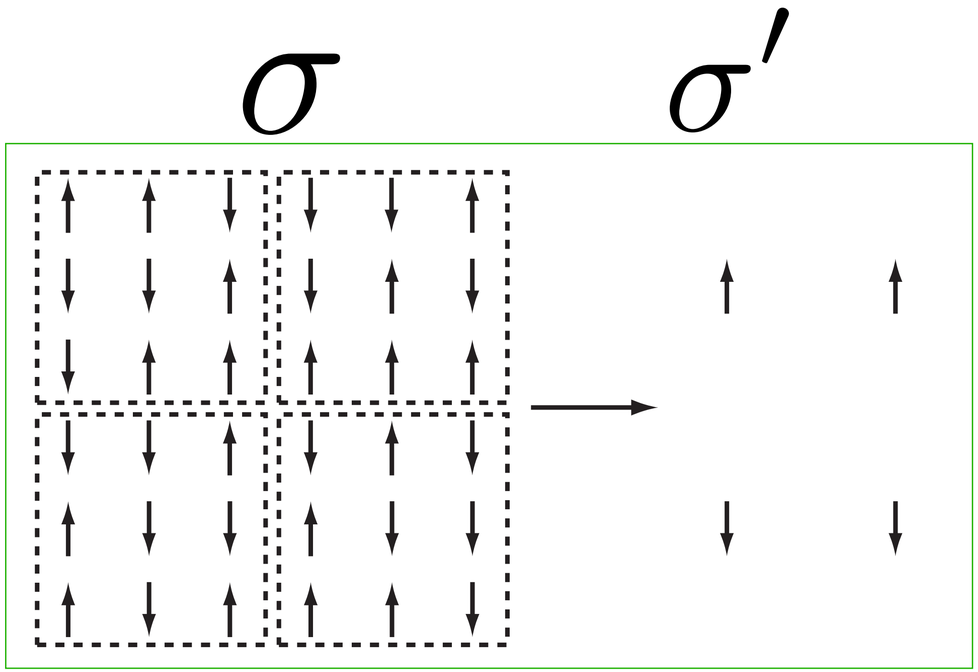
$$\mathrm{e}^{-\beta \mathcal{H}_0'(\{\sigma'\})} = \mathrm{Tr}_\sigma \left[\prod_{\text{blocks}} T(\sigma';\sigma_1,...,\sigma_9) \right] \mathrm{e}^{-\beta \mathcal{H}_0(\{\sigma\})}, \tag{16.9.4}$$

which follows from eqn. (16.9.3). Summing both sides of eqn. (16.9.4) over the relevant spin variables yields

$$\mathrm{Tr}_{\sigma'} \mathrm{e}^{-\beta \mathcal{H}_0'(\{\sigma'\})} = \mathrm{Tr}_\sigma \mathrm{e}^{-\beta \mathcal{H}_0(\{\sigma\})}. \tag{16.9.5}$$

Tuckerman, *Statistical Mechanics: Theory and Simulation*, 2010

Kadanoff's spin-block transformation



$$Q(N,T) = \sum_{\sigma_1} \cdots \sum_{\sigma_N} \mathrm{e}^{-\beta \mathcal{H}_0(\sigma_1,\ldots,\sigma_N)} \equiv \mathrm{Tr}_\sigma \mathrm{e}^{-\beta \mathcal{H}_0(\sigma_1,\ldots,\sigma_N)}. \tag{16.9.1}$$

The transformation function $T(\sigma';\sigma_1,...,\sigma_9)$ that yields the single spin σ' for each 3×3 block of 9 spin variables can be expressed mathematically as follows:

$$T(\sigma';\sigma_1,...,\sigma_9) = \begin{cases} 1 & \sigma' \sum_{i=1}^{9} \sigma_i > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{16.9.2}$$

$$\sum_{\sigma'=\pm 1} T(\sigma';\sigma_1,...,\sigma_9) = 1, \tag{16.9.3}$$

which means simply that only one of the two values of σ' can satisfy the block spin transformation rule. The new spin variables $\{\sigma'\}$ can now be used to define a new partition function. To see how this is done, let the Hamiltonian of the new lattice be defined according to

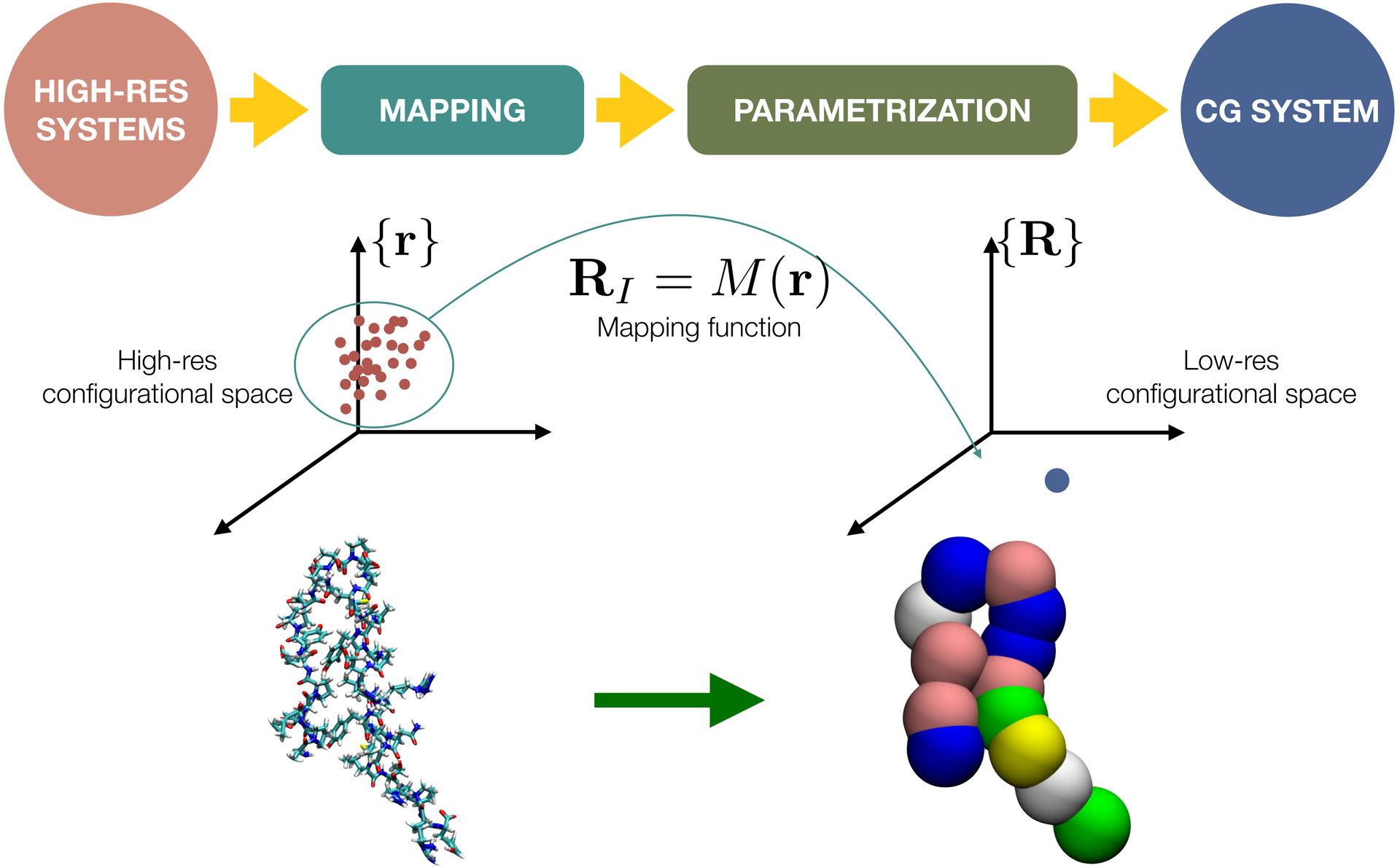
$$\mathrm{e}^{-\beta \mathcal{H}'_0(\{\sigma'\})} = \mathrm{Tr}_\sigma \left[\prod_{\text{blocks}} T(\sigma';\sigma_1,...,\sigma_9) \right] \mathrm{e}^{-\beta \mathcal{H}_0(\{\sigma\})} \tag{16.9.4}$$

over the relevant

(16.9.5)

Tuckerman, *Statistical Mechanics: Theory and Simulation*, 2010

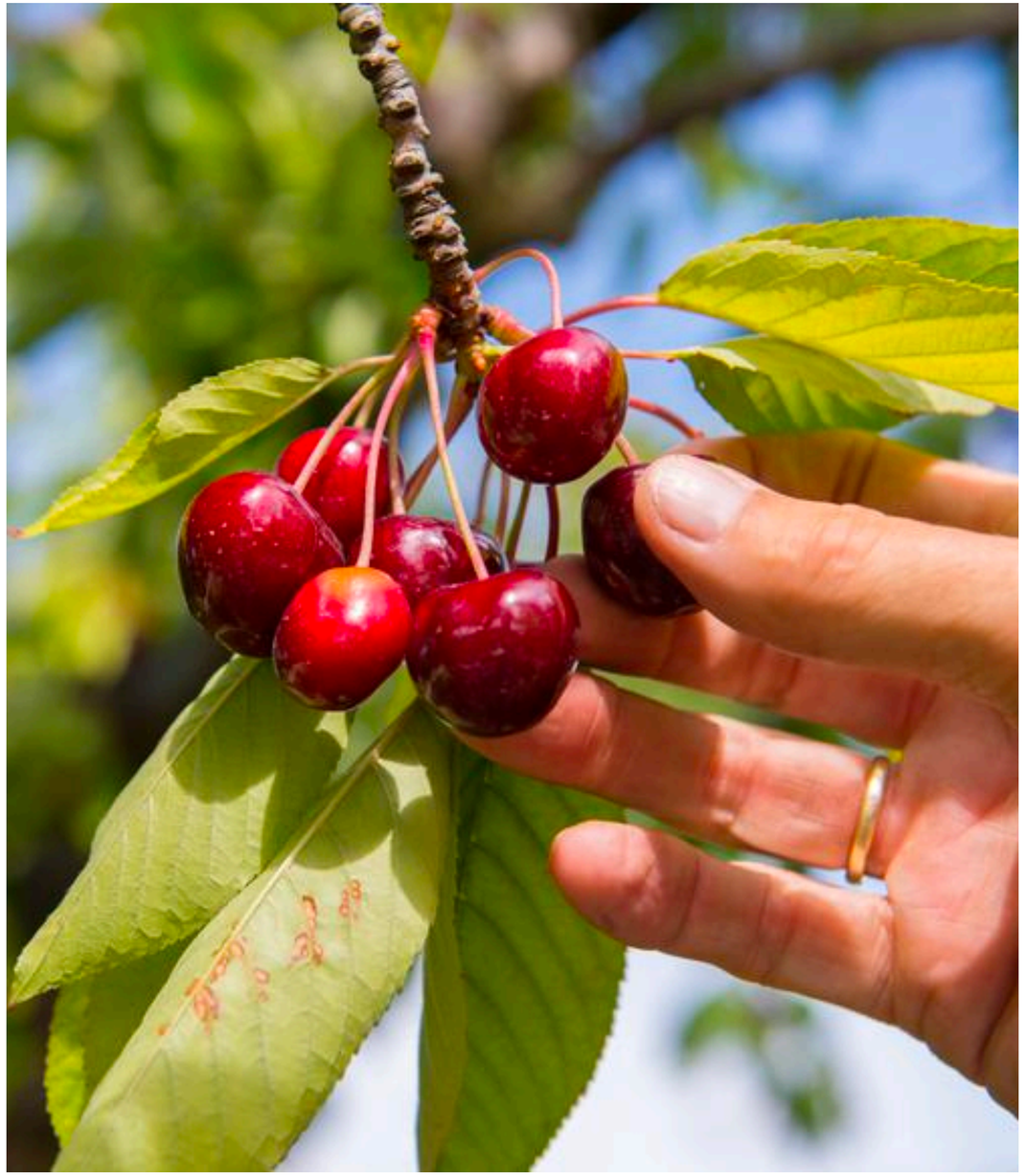
Bottom-up (systematic) CG'ing



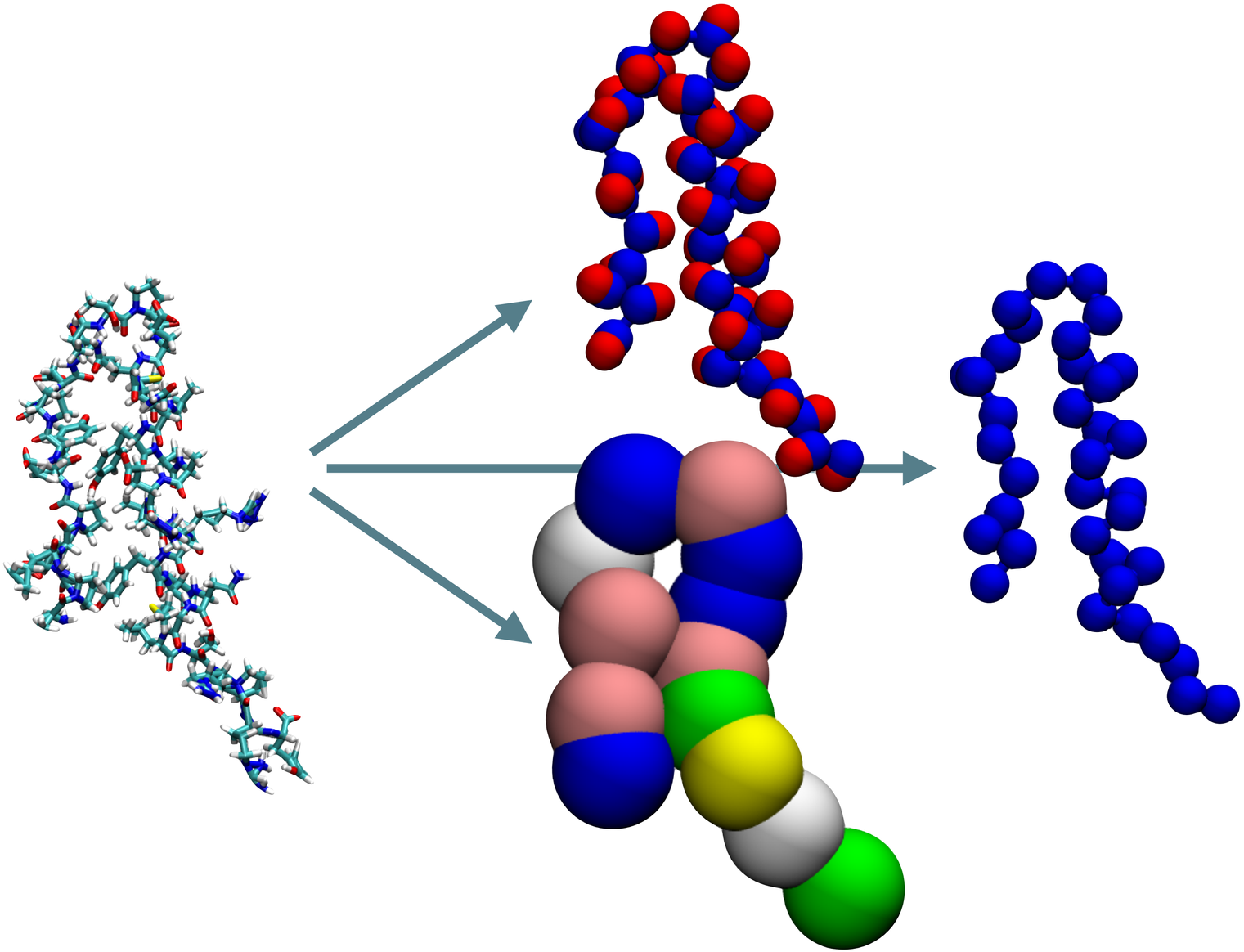
The optimal resolution

or

For a given number of atoms to be retained, which specific subset should I choose?



One system, many models



Decimation mapping

How do we define the resolution level of a protein?

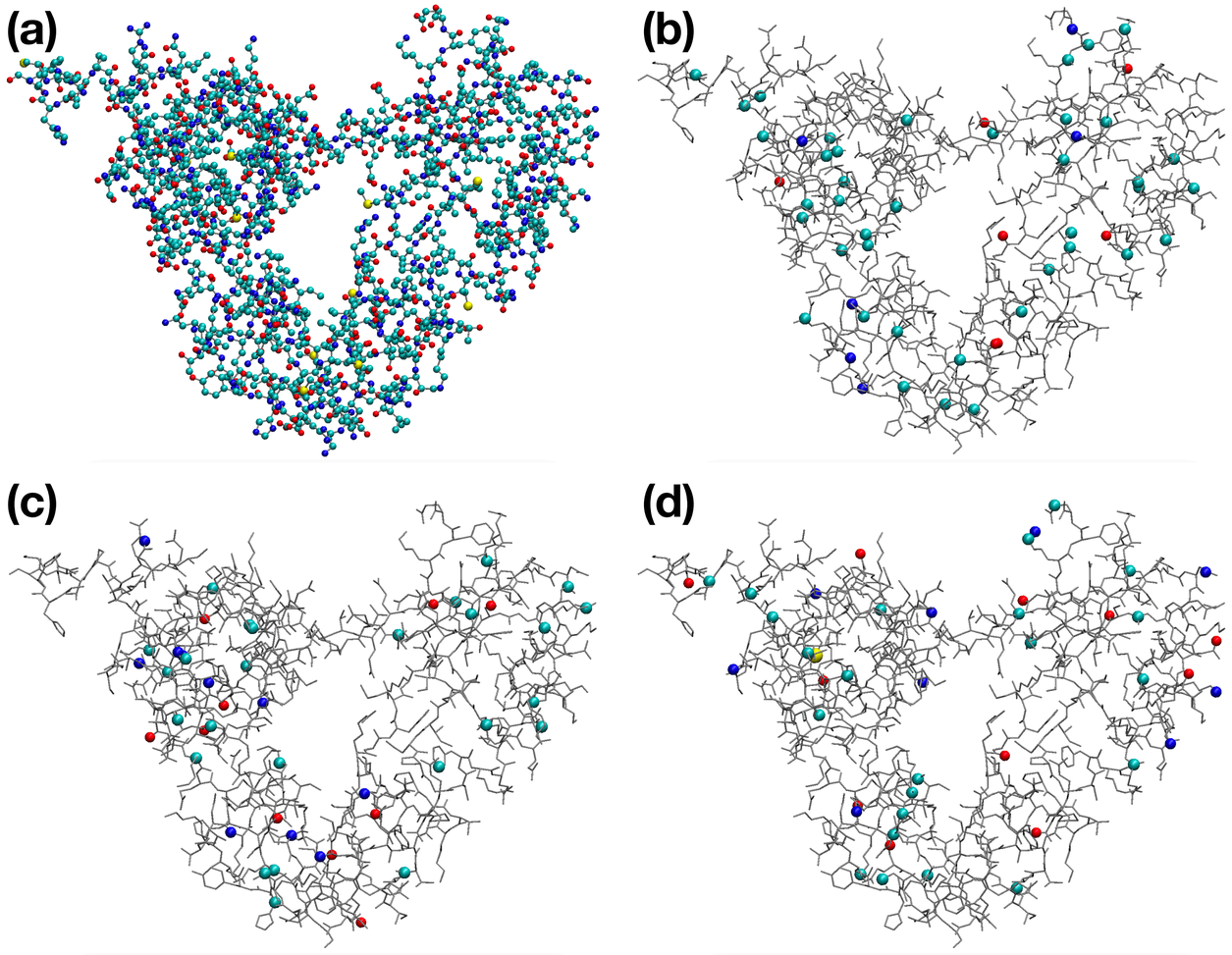
How do we select a given atom subset?

Decimation mapping

Originally introduced in the context of RG and critical phenomena

Retain a subset of the (heavy) atoms

Describe the system in terms of this subset



Giulini, Rigoli, Matiotti, Menichetti, Tarenzi, Fiorentini, RP, *Front. Mol. Biosci.* 2021
Menichetti, Giulini, RP, *EPJB* 2021

Mapping entropy optimisation workflow
MEOW

Giulini, Menichetti, Shell, RP, *JCTC* 2020
Holtzman, Giulini, RP, *Phys Rev E* 2022
Giulini, Fiorentini, Tubiana, RP, Menichetti, *arXiv* 2024

Mapping entropy optimisation workflow

MEOW

$$S_{map} = k_B \int d\mathbf{r}\ p_r(\mathbf{r}) ln\left[\frac{p_r(\mathbf{r})}{\bar{p}_r(\mathbf{r})}\right]$$

$$\bar{p}(\mathbf{r}) = \frac{p_R(\mathbf{M}(\mathbf{r}))}{\Sigma(\mathbf{M}(\mathbf{r}))}$$

Kullback-Leibler divergence between the exact probability density and its coarse-grained counterpart

Giulini, Menichetti, Shell, RP, *JCTC* 2020
Holtzman, Giulini, RP, *Phys Rev E* 2022
Giulini, Fiorentini, Tubiana, RP, Menichetti, *arXiv* 2024

Mapping entropy optimisation workflow

MEOW

$$S_{map} = k_B \int d\mathbf{r}\ p_r(\mathbf{r}) ln\left[\frac{p_r(\mathbf{r})}{\bar{p}_r(\mathbf{r})}\right]$$

$$\bar{p}(\mathbf{r}) = \frac{p_R(\mathbf{M}(\mathbf{r}))}{\Sigma(\mathbf{M}(\mathbf{r}))}$$

Probability to sample an AA configuration $\mathbf{r}$

Kullback-Leibler divergence between the exact probability density and its coarse-grained counterpart

Giulini, Menichetti, Shell, RP, *JCTC* 2020
Holtzman, Giulini, RP, *Phys Rev E* 2022
Giulini, Fiorentini, Tubiana, RP, Menichetti, *arXiv* 2024

Mapping entropy optimisation workflow

MEOW

$$S_{map} = k_B \int d\mathbf{r}\ p_r(\mathbf{r}) ln\left[\frac{p_r(\mathbf{r})}{\bar{p}_r(\mathbf{r})}\right]$$

$$\bar{p}(\mathbf{r}) = \frac{p_R(\mathbf{M}(\mathbf{r}))}{\Sigma(\mathbf{M}(\mathbf{r}))}$$

Kullback-Leibler divergence between the exact probability density and its coarse-grained counterpart

Probability to sample an AA configuration $\mathbf{r}$

Probability to sample a CG configuration $\mathbf{R}$ such that $\mathbf{M(r)} = \mathbf{R}$

Giulini, Menichetti, Shell, RP, *JCTC* 2020
Holtzman, Giulini, RP, *Phys Rev E* 2022
Giulini, Fiorentini, Tubiana, RP, Menichetti, *arXiv* 2024

Mapping entropy optimisation workflow

MEOW

$$S_{map} = k_B \int d\mathbf{r}\; p_r(\mathbf{r}) ln\left[\frac{p_r(\mathbf{r})}{\bar{p}_r(\mathbf{r})}\right]$$

$$\bar{p}(\mathbf{r}) = \frac{p_R(\mathbf{M}(\mathbf{r}))}{\Sigma(\mathbf{M}(\mathbf{r}))}$$

Kullback-Leibler divergence between the exact probability density and its coarse-grained counterpart

Probability to sample an AA configuration $\mathbf{r}$

Number of AA configurations $\mathbf{r}$ such that $\mathbf{M(r)} = \mathbf{R}$

Probability to sample a CG configuration $\mathbf{R}$ such that $\mathbf{M(r)} = \mathbf{R}$

Giulini, Menichetti, Shell, RP, *JCTC* 2020
Holtzman, Giulini, RP, *Phys Rev E* 2022
Giulini, Fiorentini, Tubiana, RP, Menichetti, *arXiv* 2024

Mapping entropy optimisation workflow

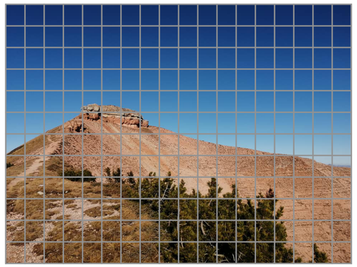
MEOW

$$S_{map} = k_B \int d\mathbf{r}\ p_r(\mathbf{r}) ln\left[\frac{p_r(\mathbf{r})}{\bar{p}_r(\mathbf{r})}\right]$$

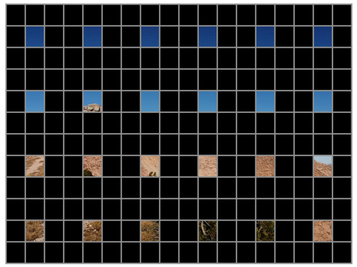
$$\bar{p}(\mathbf{r}) = \frac{p_R(\mathbf{M}(\mathbf{r}))}{\Sigma(\mathbf{M}(\mathbf{r}))}$$

Kullback-Leibler divergence between the exact probability density and its coarse-grained counterpart

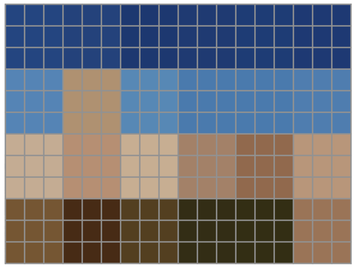
A)



Original high resolution representation

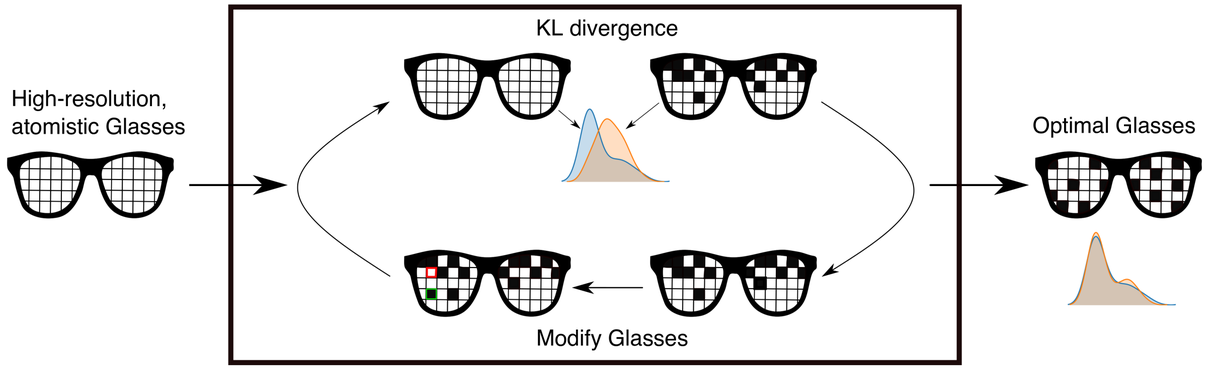


Filtering



Reconstructed high resolution representation

B)

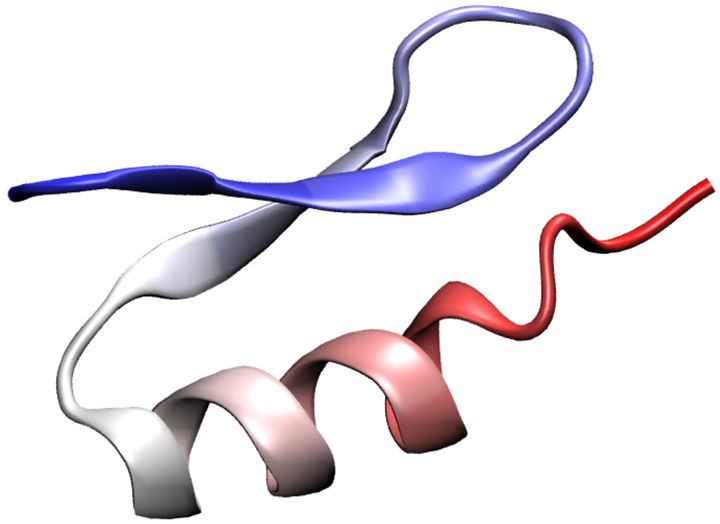


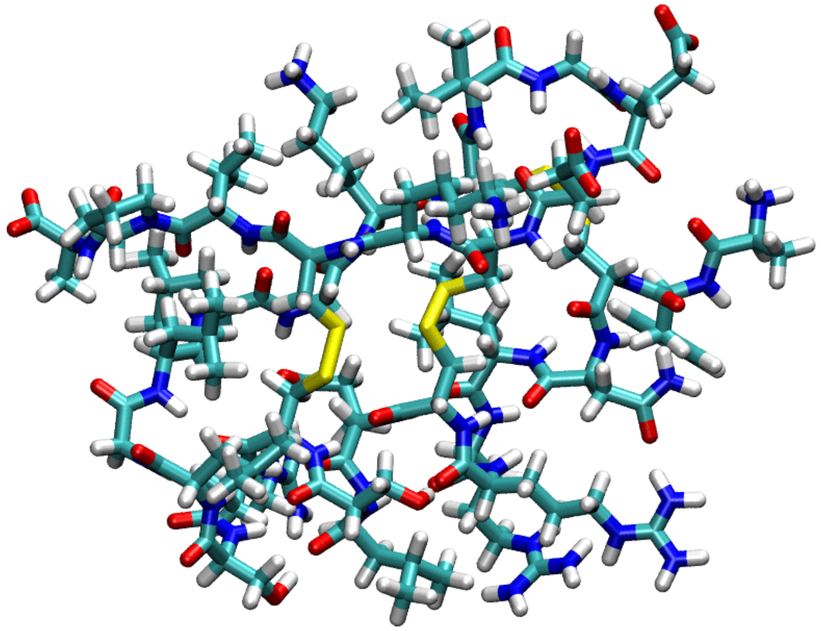
Giulini, Menichetti, Shell, RP, *JCTC* 2020
Holtzman, Giulini, RP, *Phys Rev E* 2022
Giulini, Fiorentini, Tubiana, RP, Menichetti, *arXiv* 2024

Application to the tamapin toxin

Application to tamapin

Toxin from Indian red scorpion
Blocker of SK2 potassium channel
High pharmaceutical interest



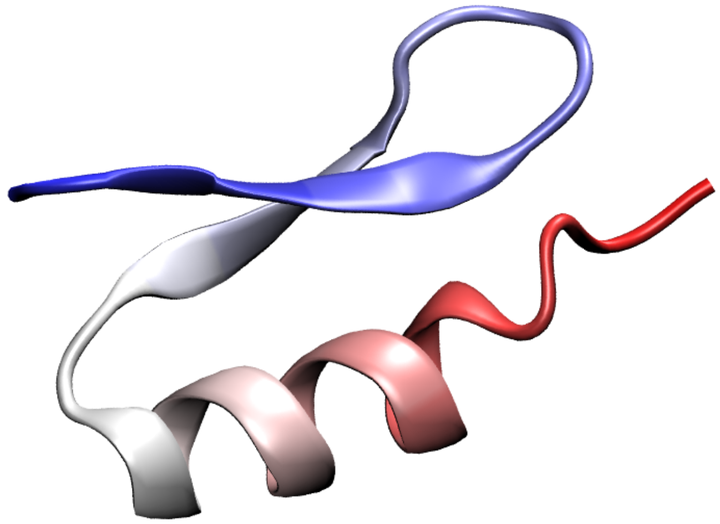


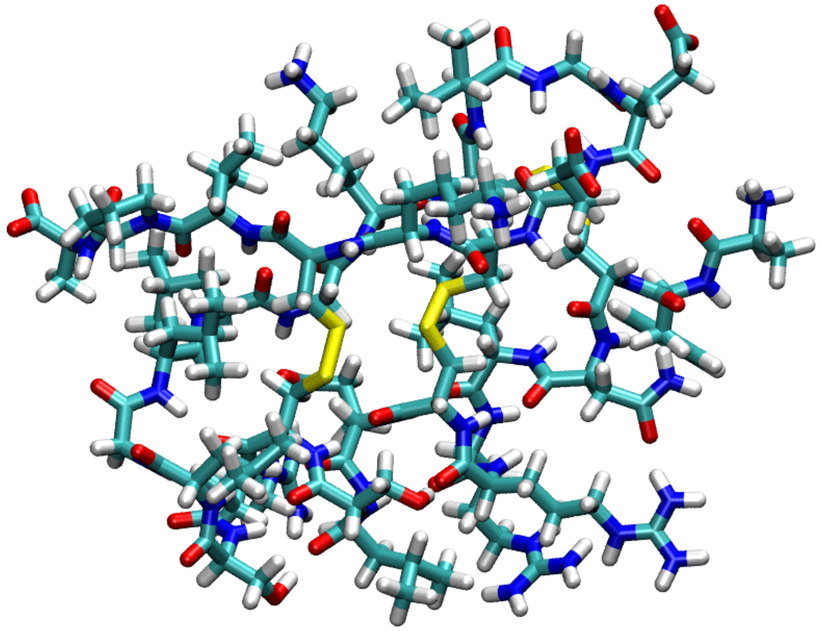
Giulini, Menichetti, Shell, RP, *JCTC* 2020

Application to the tamapin toxin

Application to tamapin

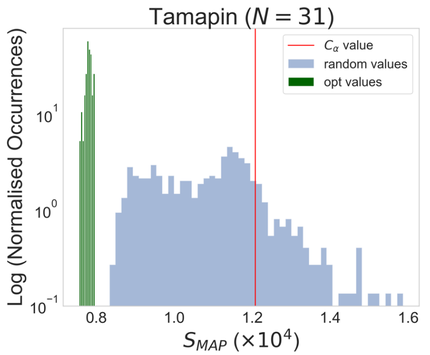
Toxin from Indian red scorpion
Blocker of SK2 potassium channel
High pharmaceutical interest

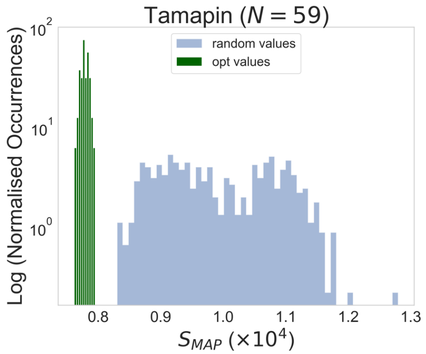


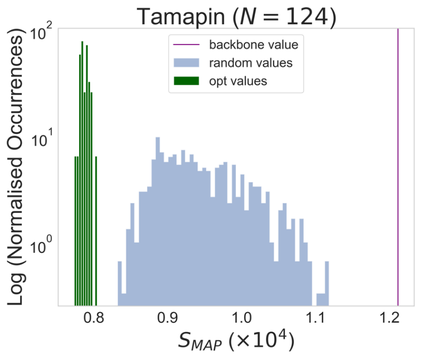


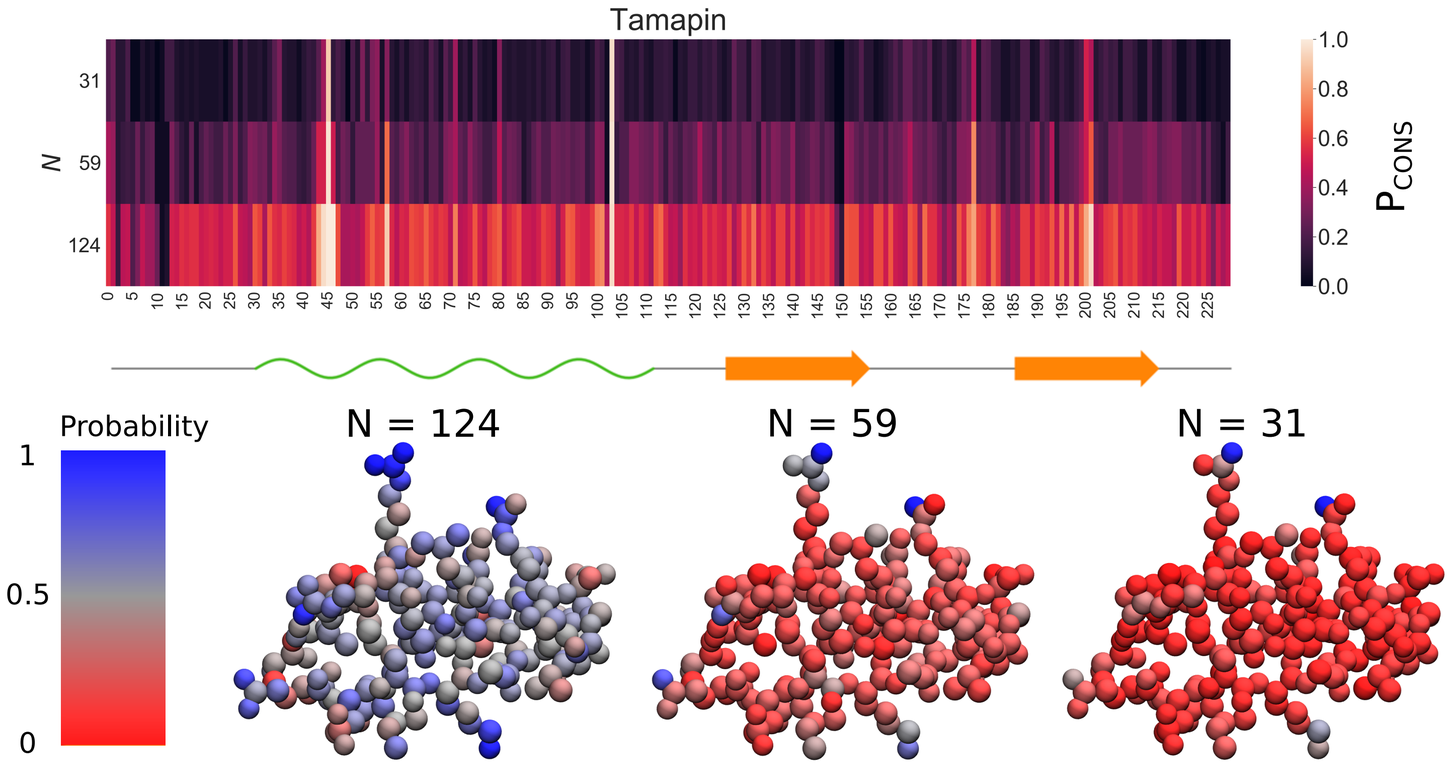
Giulini, Menichetti, Shell, RP, *JCTC* 2020

Application to the tamapin toxin









ARG6 and ARG13: main actors involved in the tamapin-SK2 channel interaction through electrostatics and hydrogen bonding

Mutation of an arginine dramatically decreases selectivity

Giulini, Menichetti, Shell, RP, *JCTC* 2020

Andreotti, di Luccio, Sampieri, Waard, Sabatier, Peptides 2005
Quintero-Hernndez, Jimnez-Vargas, Gurrola, Valdivia, Possani Toxicon 2013
Ramirez-Cordero, Toledano, Cano-Snchez, Hernndez-Lopez, Flores-Solis, Saucedo-Yez, Chavez-Uribe, Brieba, and Rio-Portilla, Chemical Research in Toxicology 2014

Take-home messages #1

The construction of low-resolution models in soft matter
is akin to Kadanoff block-spin transformation in statistical physics

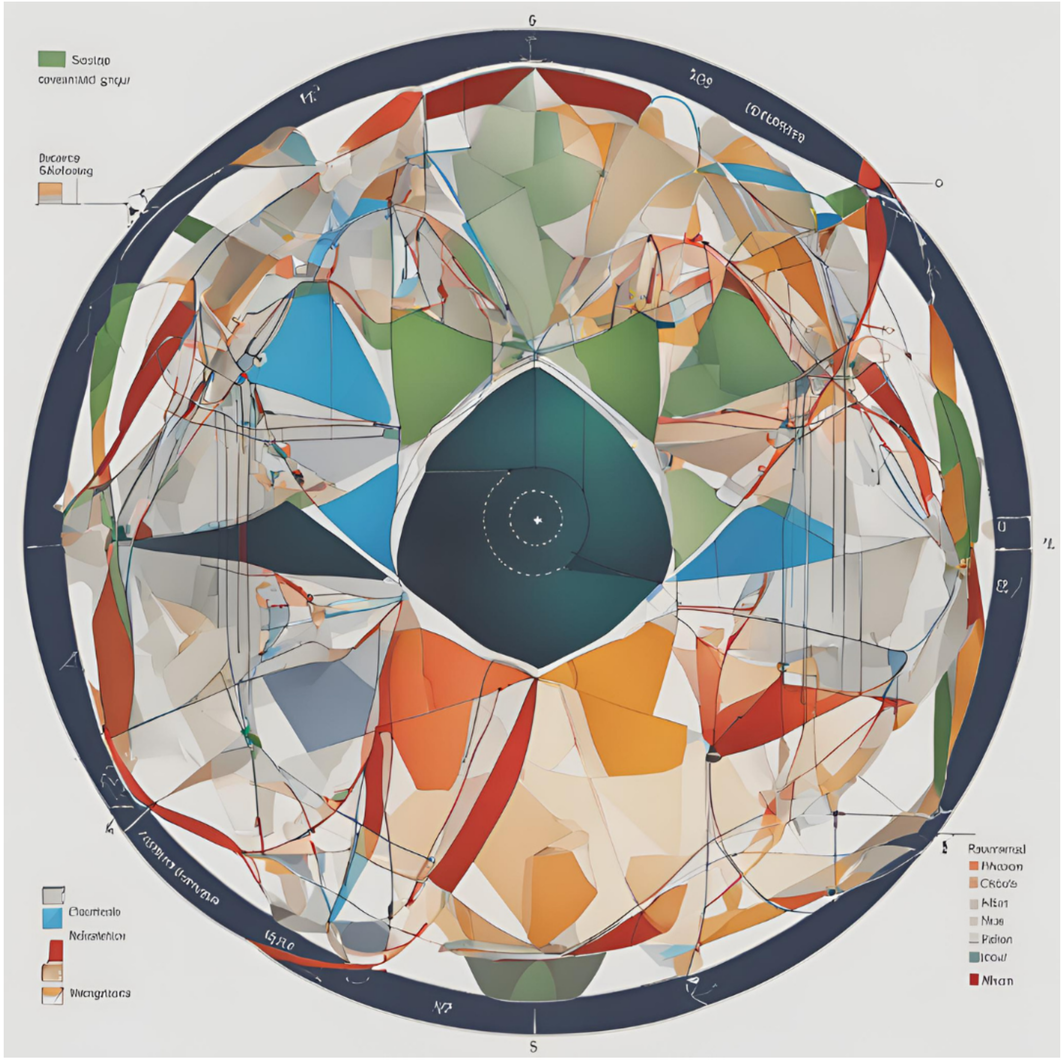
The choice of the mapping is not as neutral in SM as it is *e.g.* in FT

Optimally-chosen mappings can let important information emerge

Solving the one-step RG flow

or

Given the high-resolution model of our system, how can we parametrise the interactions of the coarse-grained one?



Framework of bottom-up CG'ing

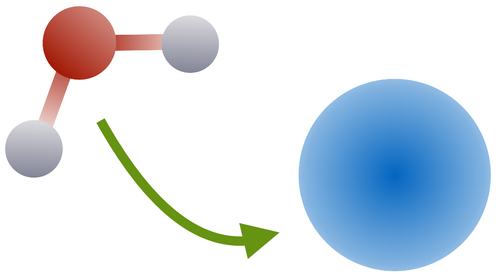
$$h(\mathbf{r},\mathbf{p}) = \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i} + u(\mathbf{r})$$

$$p(\mathbf{r},\mathbf{p}) \propto \exp\left(-\beta \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i} - \beta u(\mathbf{r})\right)$$

$$p_r(\mathbf{r}) \propto \exp\left(-\beta u(\mathbf{r})\right)$$

$$p_p(\mathbf{p}) \propto \exp\left(-\beta \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i}\right)$$

All-atom (high-resolution) representation



$$\mathbf{R} = \mathbf{M}(\mathbf{r}) = \{\mathbf{M}(\mathbf{r}_1), \mathbf{M}(\mathbf{r}_2), \cdots \mathbf{M}(\mathbf{r}_n)\}$$

$$\mathbf{R}_I = \mathbf{M}_I(\mathbf{r}) = \sum_{i=1}^{n} c_{Ii}\mathbf{r}_i$$

Mapping function - usually linear

$$H(\mathbf{R},\mathbf{P}) = \sum_{I=1}^{N} \frac{\mathbf{P}_I^2}{2M_I} + U(\mathbf{R})$$

$$P(\mathbf{R},\mathbf{P}) \propto \exp\left(-\beta \sum_{I=1}^{N} \frac{\mathbf{P}_I^2}{2M_I} + U(\mathbf{R})\right)$$

$$P_R(\mathbf{R}) \propto \exp\left(-\beta U(\mathbf{R})\right)$$

$$P_P(\mathbf{P}) \propto \exp\left(-\beta \sum_{I=1}^{N} \frac{\mathbf{P}_I^2}{2M_I}\right)$$

Coarse-grained (low-resolution) representation

Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014
Noid, *J. Phys. Chem. B* 2023

Consistency conditions

$$p_R(\mathbf{R}) = \langle \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R}) \rangle = \int d\mathbf{r} \, \frac{e^{-\beta u(\mathbf{r})}}{Z_{AA}} \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R})$$

$$\delta(\mathbf{M}(\mathbf{r}) - \mathbf{R}) \equiv \prod_{I=1}^{N} \delta(\mathbf{M}_I(\mathbf{r}) - \mathbf{R}_I)$$

$$P_R(\mathbf{R}) = p_R(\mathbf{R})$$

Consistency conditions

$$p_R(\mathbf{R}) = \langle \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R}) \rangle = \int d\mathbf{r} \; \frac{e^{-\beta u(\mathbf{r})}}{Z_{AA}} \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R})$$

$$\delta(\mathbf{M}(\mathbf{r}) - \mathbf{R}) \equiv \prod_{I=1}^{N} \delta(\mathbf{M}_I(\mathbf{r}) - \mathbf{R}_I)$$

$$P_R(\mathbf{R}) = p_R(\mathbf{R})$$

MULTI-BODY POTENTIAL OF MEAN FORCE

$$U(\mathbf{R}) = -\frac{1}{\beta} \ln \int d\mathbf{r} \; e^{-\beta u(\mathbf{r})} \; \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R}) + C$$

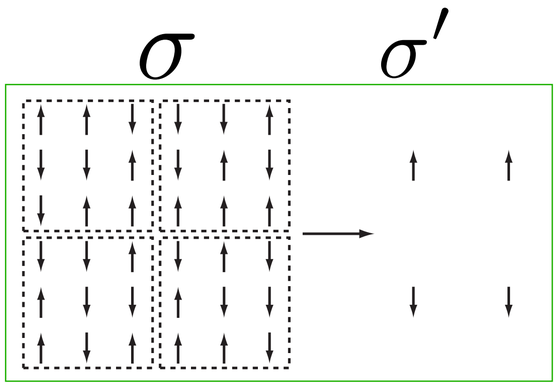
RG-like equation for the "renormalised" interactions

Consistency conditions

MULTI-BODY POTENTIAL OF MEAN FORCE

$$U(\mathbf{R}) = -\frac{1}{\beta} \ln \int d\mathbf{r}\; e^{-\beta u(\mathbf{r})}\; \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R}) + C$$

RG-like equation for the "renormalised" interactions



$$Q(N,T) = \sum_{\sigma_1} \cdots \sum_{\sigma_N} \mathrm{e}^{-\beta\mathcal{H}_0(\sigma_1,\ldots,\sigma_N)} \equiv \mathrm{Tr}_\sigma \mathrm{e}^{-\beta\mathcal{H}_0(\sigma_1,\ldots,\sigma_N)}. \tag{16.9.1}$$

The transformation function $T(\sigma';\sigma_1,...,\sigma_9)$ that yields the single spin σ' for each 3×3 block of 9 spin variables can be expressed mathematically as follows:

$$T(\sigma';\sigma_1,...,\sigma_9) = \begin{cases} 1 & \sigma' \sum_{i=1}^{9} \sigma_i > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{16.9.2}$$

$$\sum_{\sigma'=\pm 1} T(\sigma';\sigma_1,...,\sigma_9) = 1, \tag{16.9.3}$$

which means simply that only one of the two values of σ' can satisfy the block spin transformation rule. The new spin variables $\{\sigma'\}$ can now be used to define a new partition function. To see how this is done, let the Hamiltonian of the new lattice be defined according to

$$\mathrm{e}^{-\beta\mathcal{H}'_0(\{\sigma'\})} = \mathrm{Tr}_\sigma \left[\prod_{\text{blocks}} T(\sigma';\sigma_1,...,\sigma_9) \right] \mathrm{e}^{-\beta\mathcal{H}_0(\{\sigma\})}, \tag{16.9.4}$$

which follows from eqn. (16.9.3). Summing both sides of eqn. (16.9.4) over the relevant spin variables yields

$$\mathrm{Tr}_{\sigma'} \mathrm{e}^{-\beta\mathcal{H}'_0(\{\sigma'\})} = \mathrm{Tr}_\sigma \mathrm{e}^{-\beta\mathcal{H}_0(\{\sigma\})}. \tag{16.9.5}$$

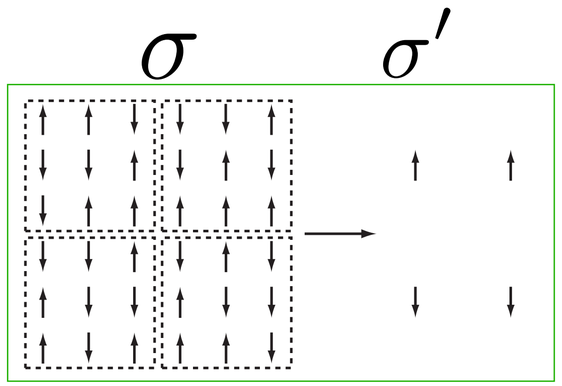
Tuckerman, *Statistical Mechanics: Theory and Simulation*, 2010

Consistency conditions

MULTI-BODY POTENTIAL OF MEAN FORCE

$$U(\mathbf{R}) = -\frac{1}{\beta} \ln \int d\mathbf{r}\; e^{-\beta u(\mathbf{r})}\; \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R}) + C$$

RG-like equation for the "renormalised" interactions



$$e^{-\beta \mathcal{H}_0'(\{\sigma'\})} = \mathrm{Tr}_\sigma \left[\prod_{\text{blocks}} T(\sigma'; \sigma_1, ..., \sigma_9) \right] e^{-\beta \mathcal{H}_0(\{\sigma\})}$$

$$\left(0 \qquad \text{otherwise} \right.$$

$..., \sigma_9) = 1,$ (16.9.3)

wo values of σ' can satisfy the block spin
s $\{\sigma'\}$ can now be used to define a new
let the Hamiltonian of the new lattice be

$\sigma'; \sigma_1, ..., \sigma_9) \Big] e^{-\beta \mathcal{H}_0(\{\sigma\})},$ (16.9.4)

which follows from eqn. (16.9.3). Summing both sides of eqn. (16.9.4) over the relevant spin variables yields

$$\mathrm{Tr}_{\sigma'} e^{-\beta \mathcal{H}_0'(\{\sigma'\})} = \mathrm{Tr}_\sigma e^{-\beta \mathcal{H}_0(\{\sigma\})}. \quad (16.9.5)$$

Tuckerman, *Statistical Mechanics: Theory and Simulation*, 2010

Take-home messages #2

The parametrisation of effective "potentials" in SM is a one-step RG

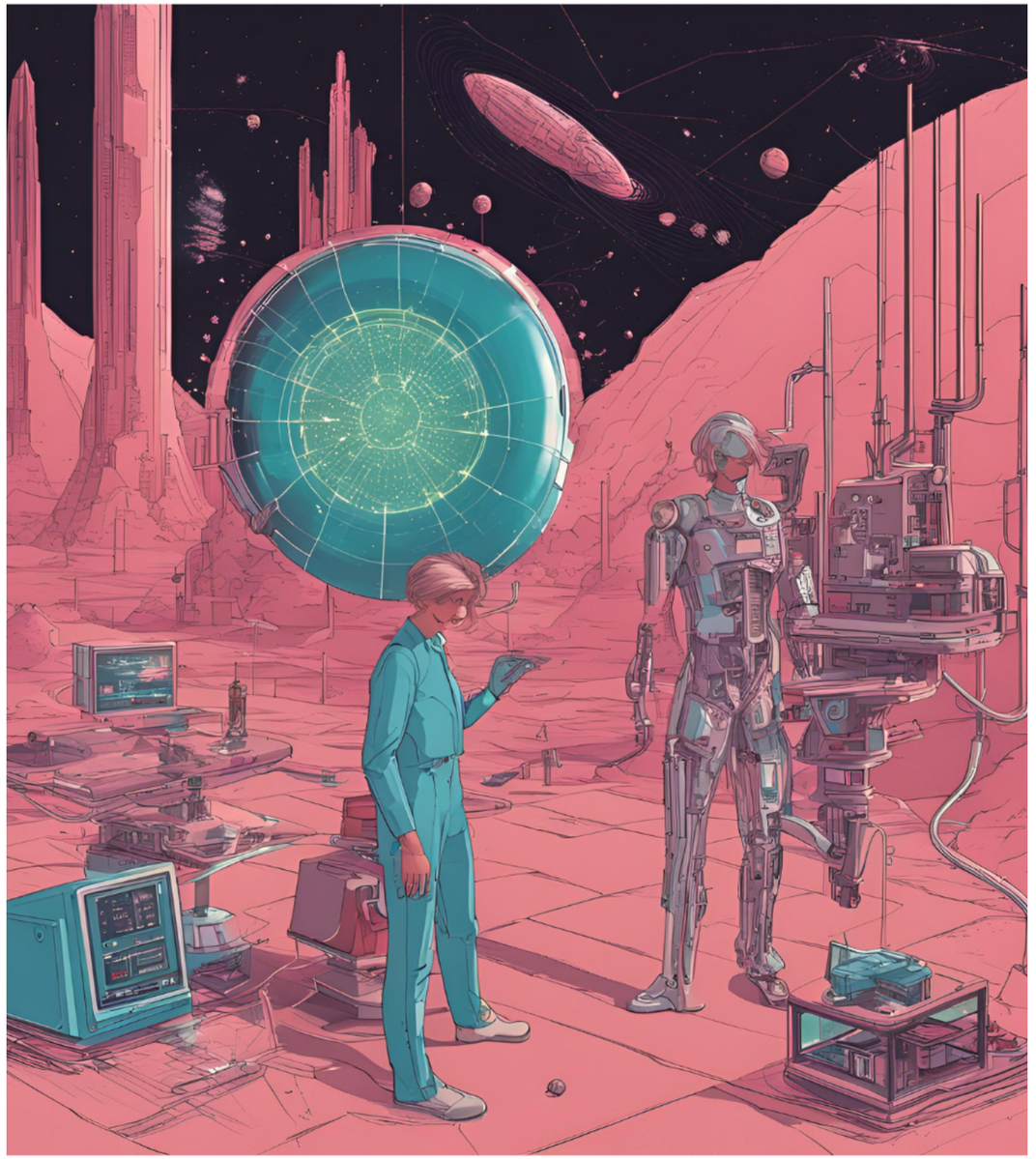
The most appropriate functional form of the effective interactions
for CG models is greatly system-specific

Various parametrisation methods exist,
which aim *e.g.* to match probability distributions or average forces

Machine learning in soft matter

or

Is there a link between SM, RG, and ML? Can we leverage ML in the study of soft matter systems?



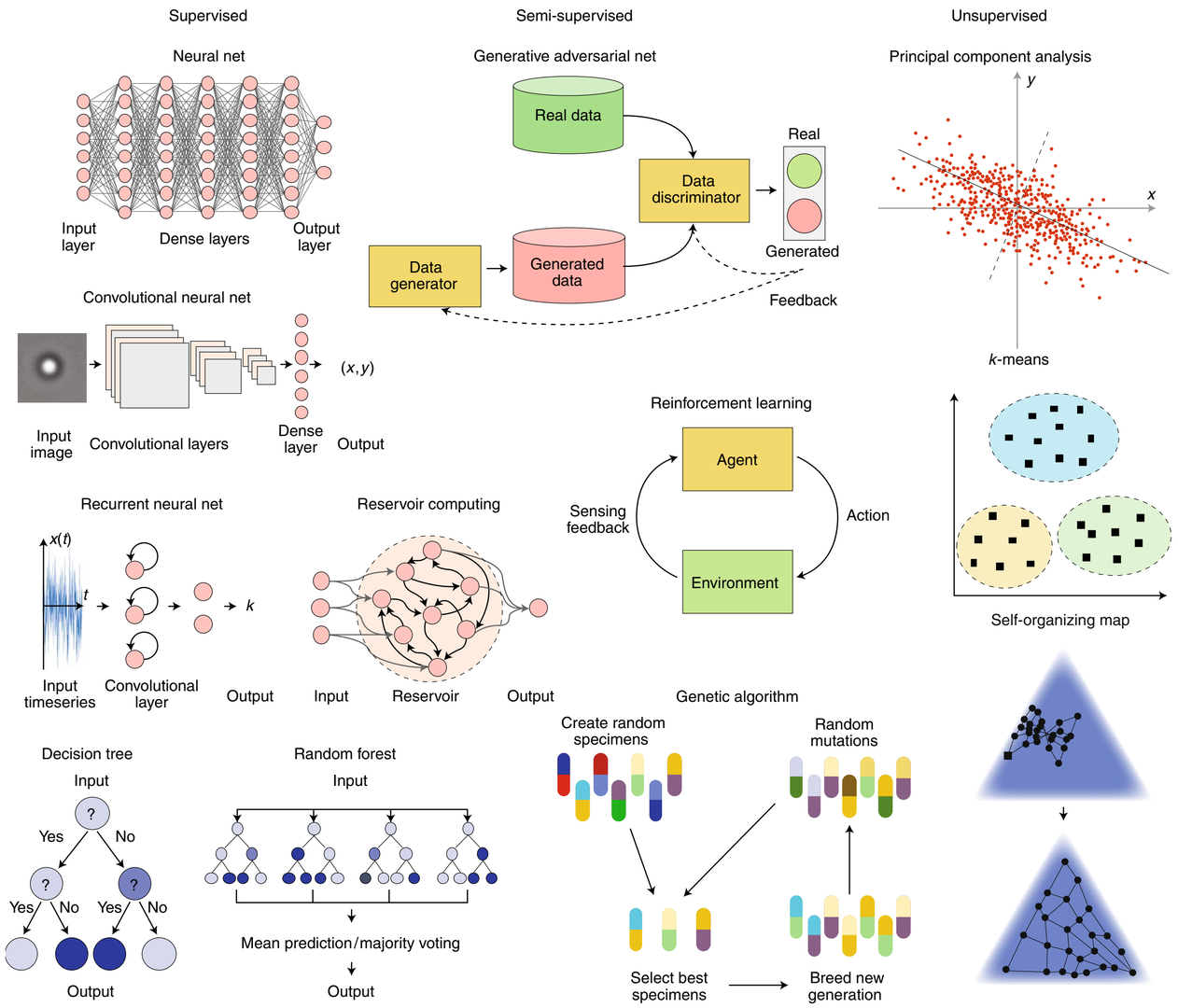
Know your machine learning

There is a world of ML methods out there

Different techniques are more appropriate for different tasks

Some ML methods are actually old-school statistical analysis

The more sophisticated the method,
the less we understand of how it works



Chicos, Gustavsson, Mehlig, Volpe, *Nature Machine Learning* 2020

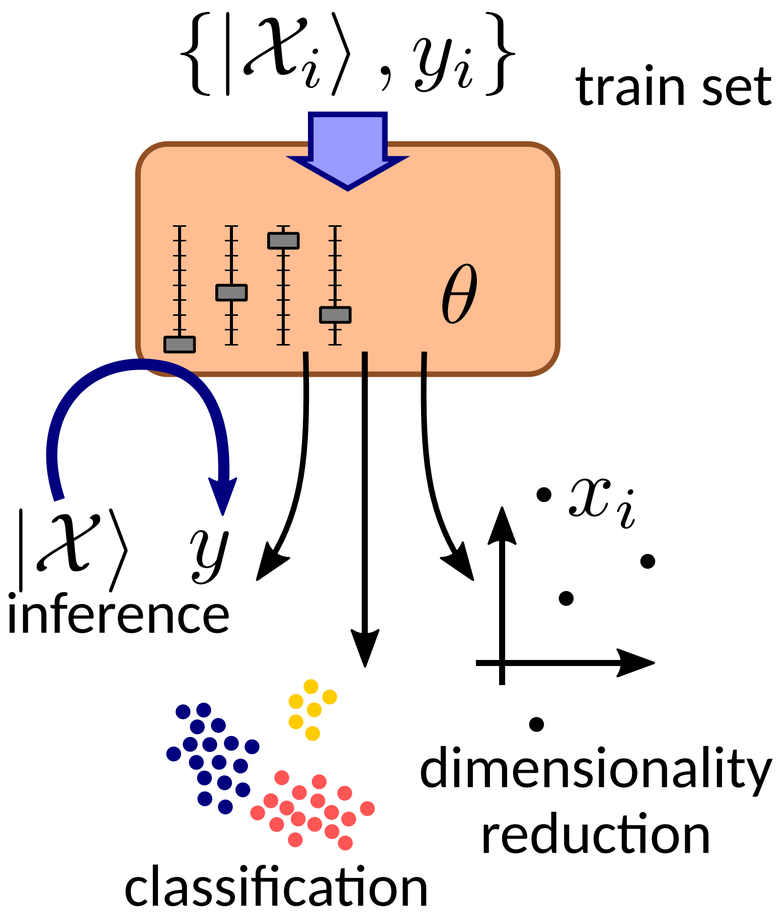
What can we use deep learning for?

Evaluation of quantities

Classification

Generation of structures

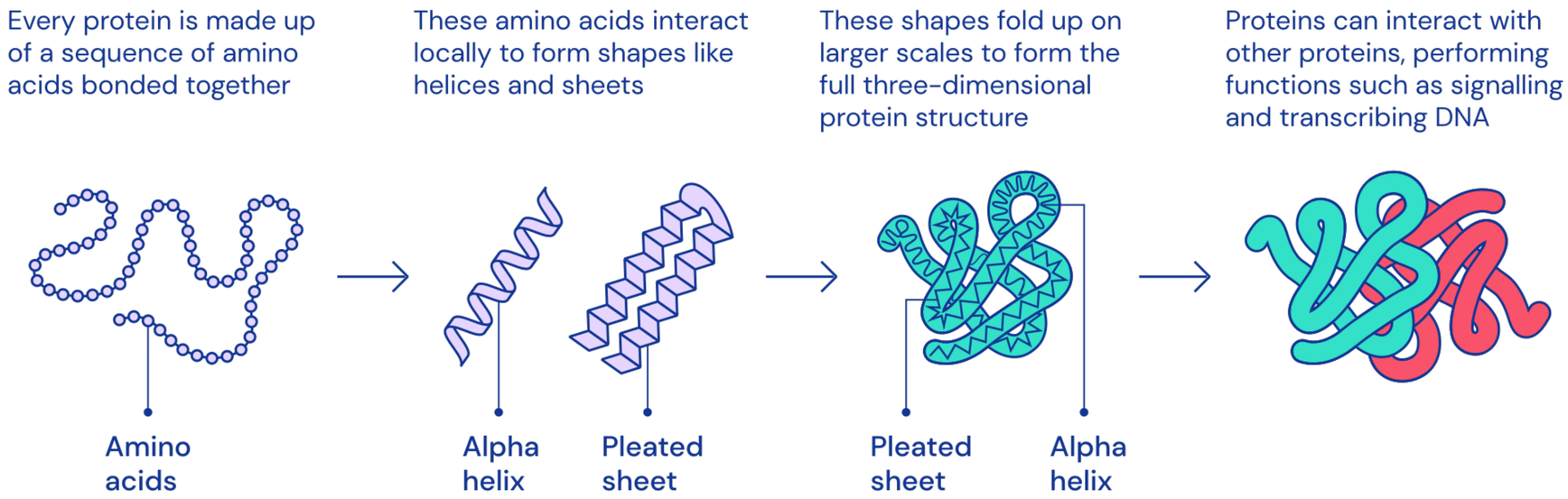
Analysis of the input



Ceriotti, *JCP* 2019

Protein structure prediction

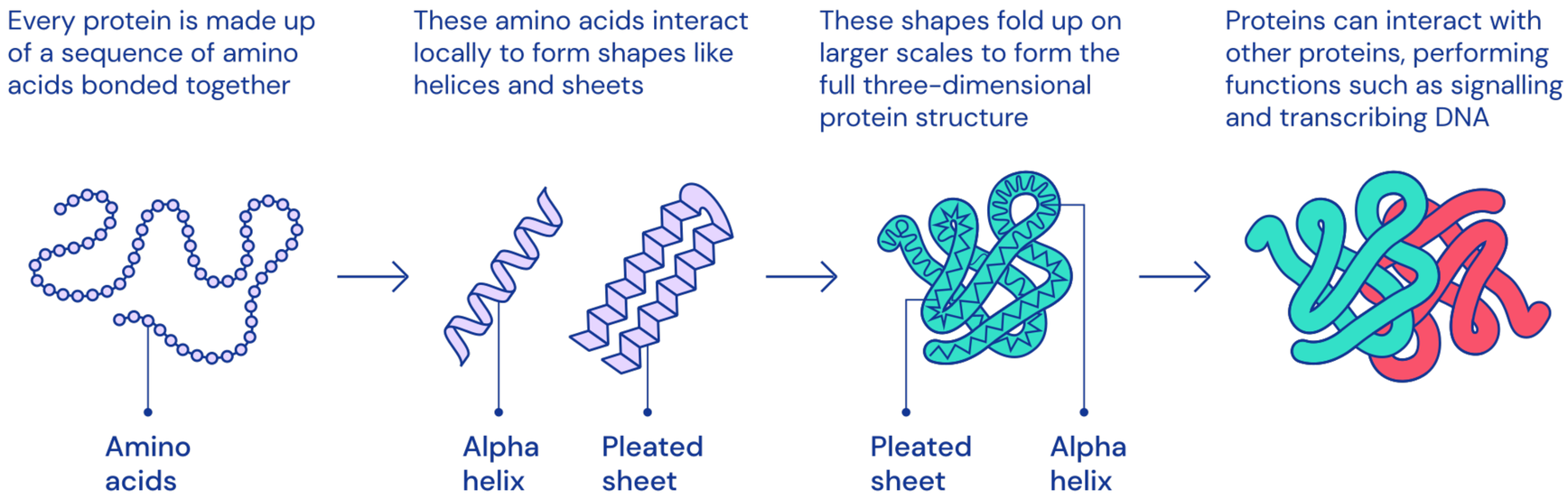
Protein folding is a hell of a problem



https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

Protein structure prediction

Protein folding is a hell of a problem



https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

Protein structure prediction

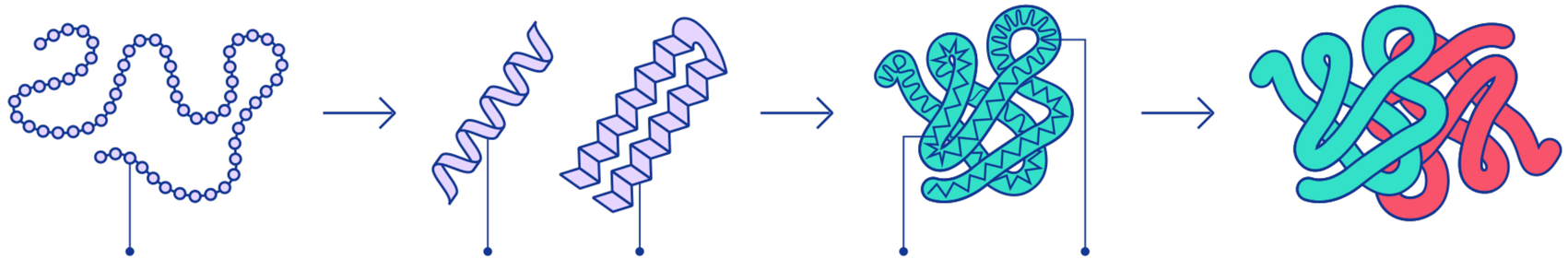
Protein folding is a hell of a problem

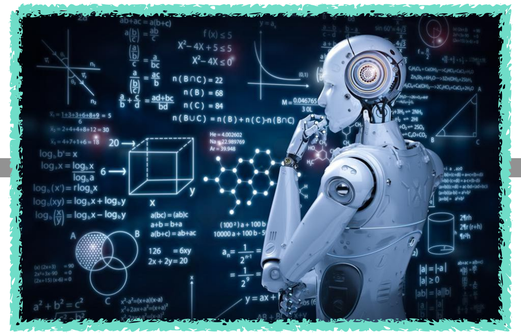
Every protein is made up of a sequence of amino acids bonded together

These amino acids interact locally to form shapes like helices and sheets

These shapes fold up on larger scales to form the full three-dimensional protein structure

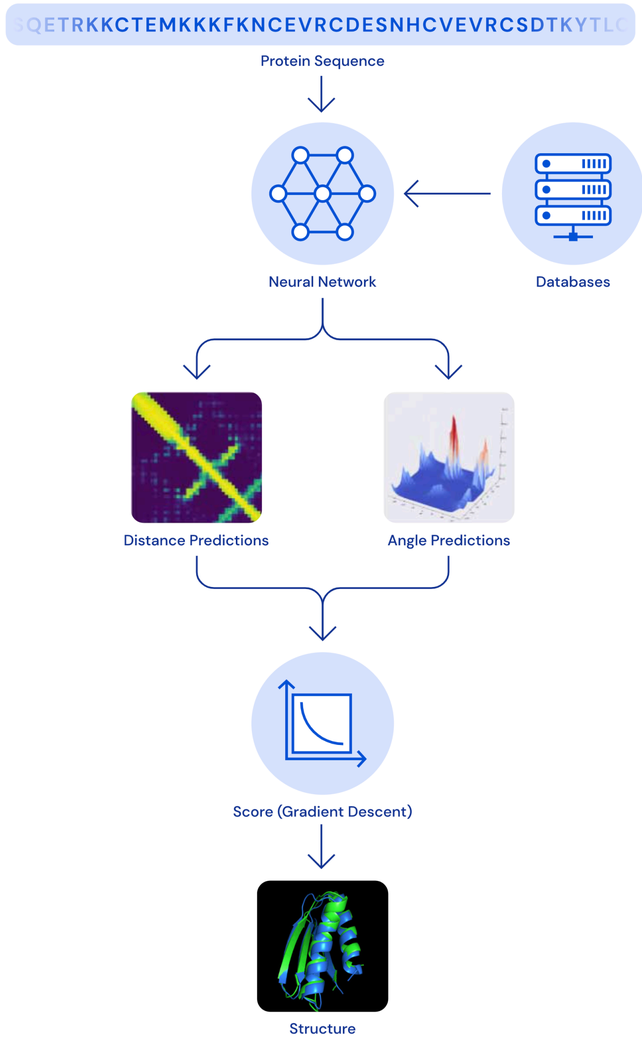
Proteins can interact with other proteins, performing functions such as signalling and transcribing DNA

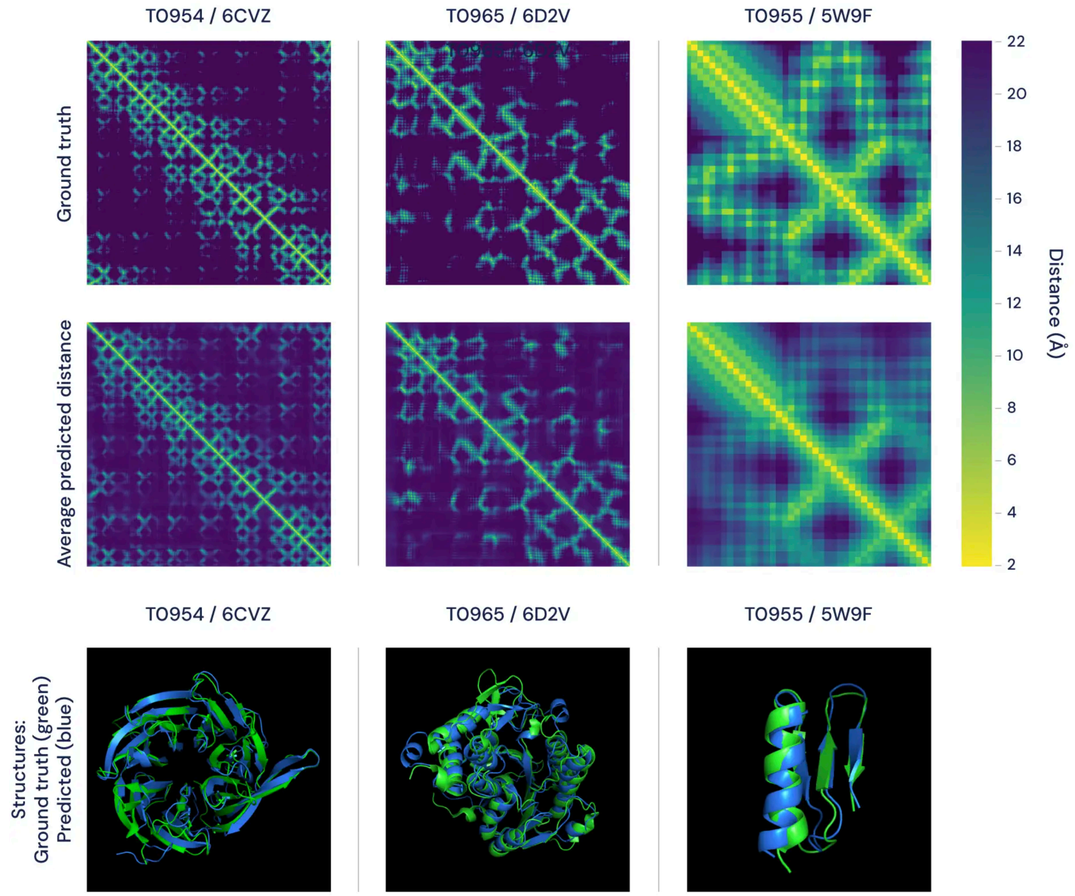


Sequence →  → Structure

https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

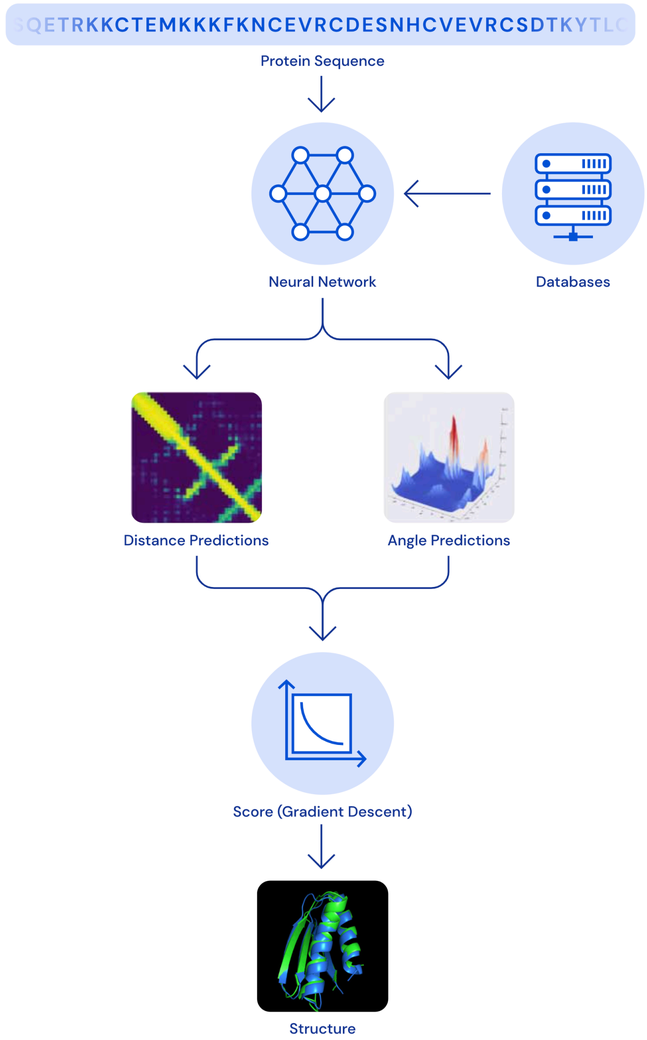
Protein structure prediction

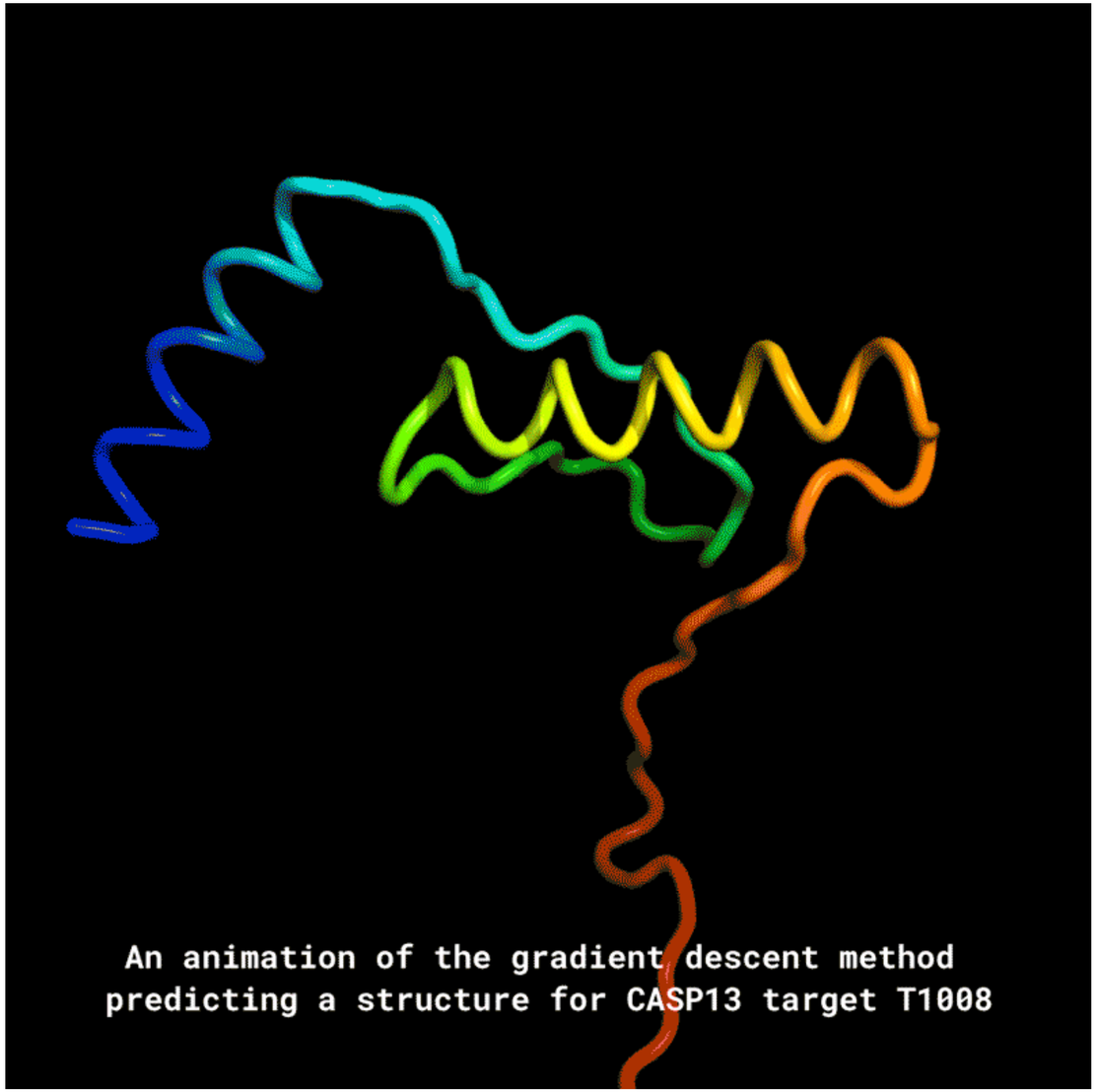




https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

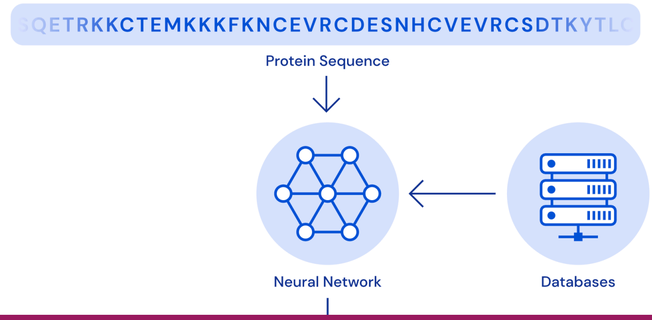
Protein structure prediction

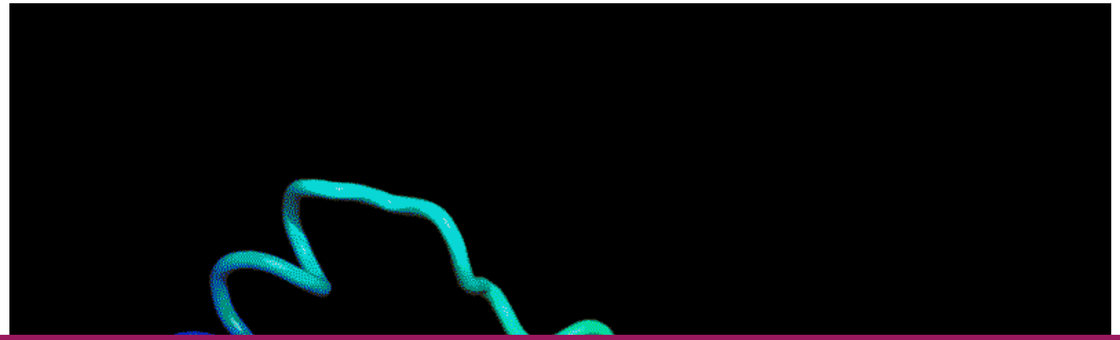




https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

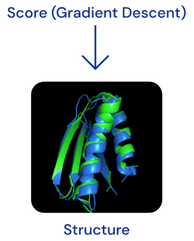
Protein structure prediction

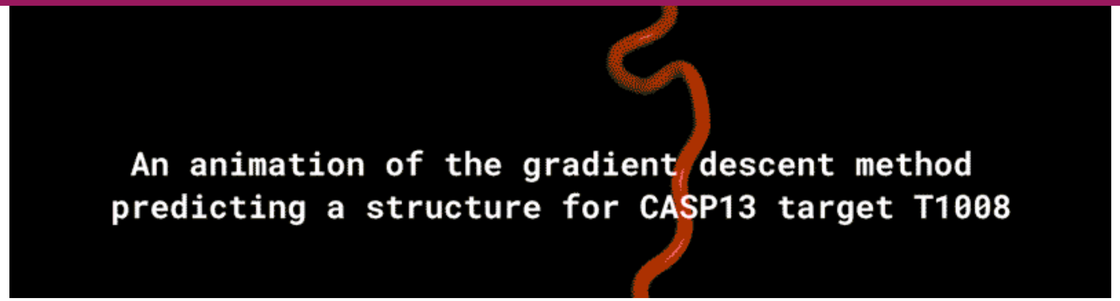




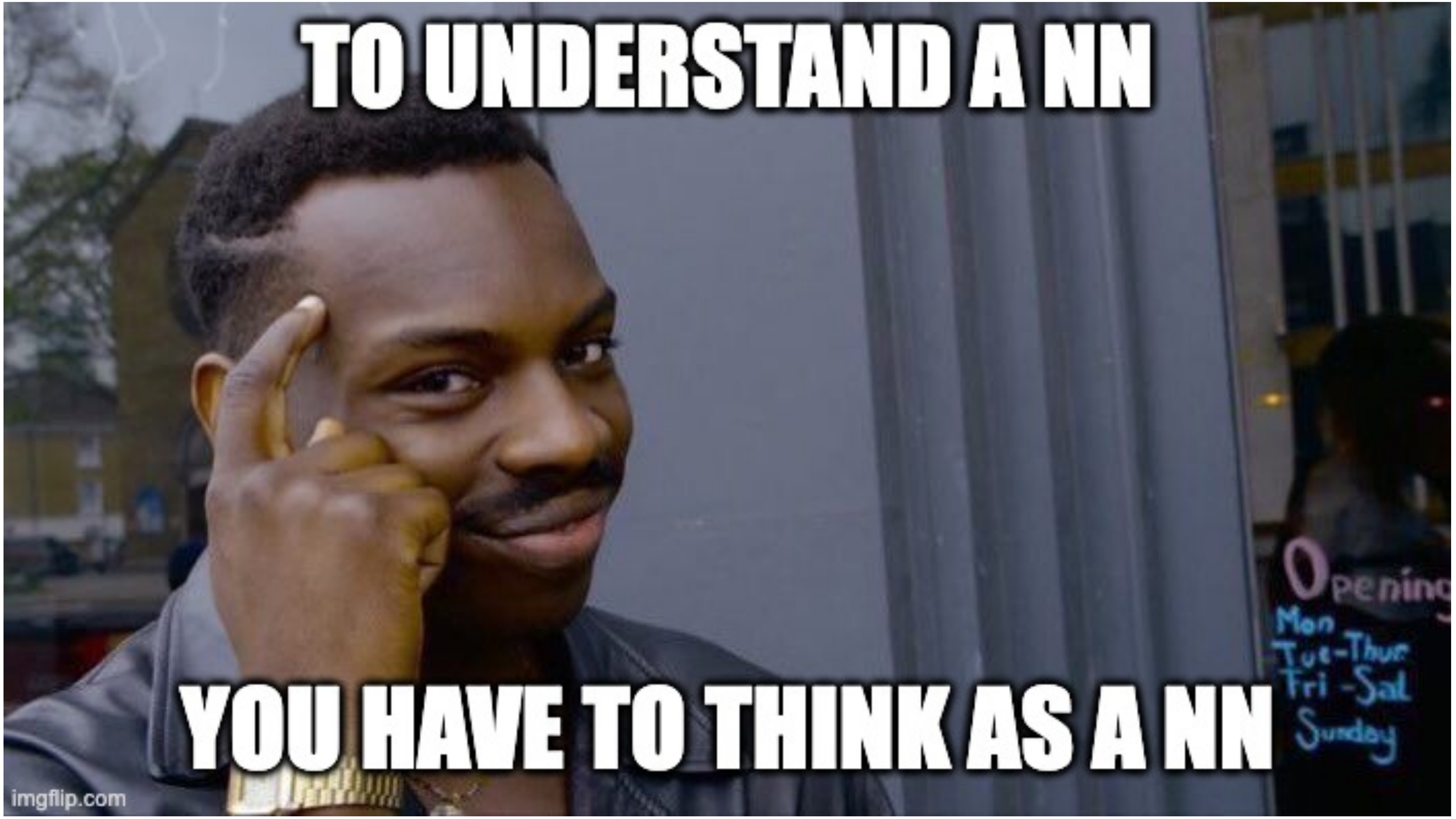
The downside of the story: neural networks work in mysterious ways

Can we understand how NN's work?
Can we understand something about the system from the way the NN's "sees" it?

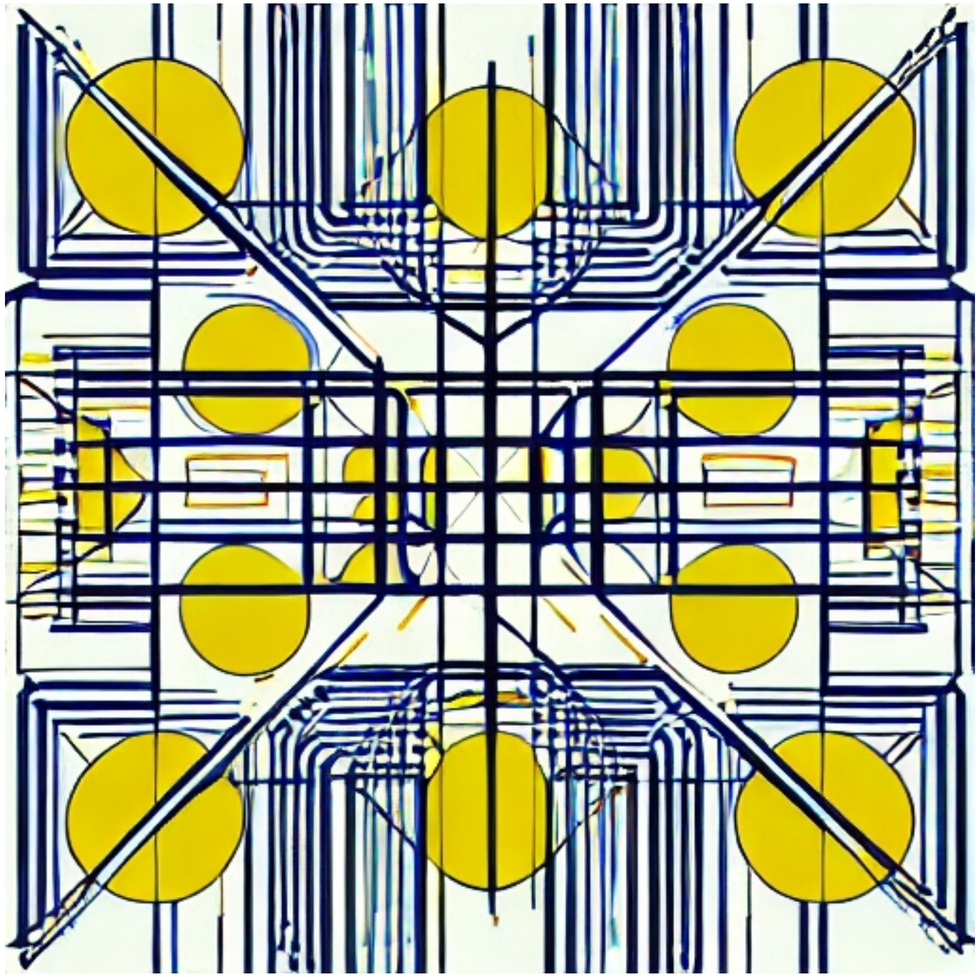




https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery


TO UNDERSTAND A NN
YOU HAVE TO THINK AS A NN
imgflip.com

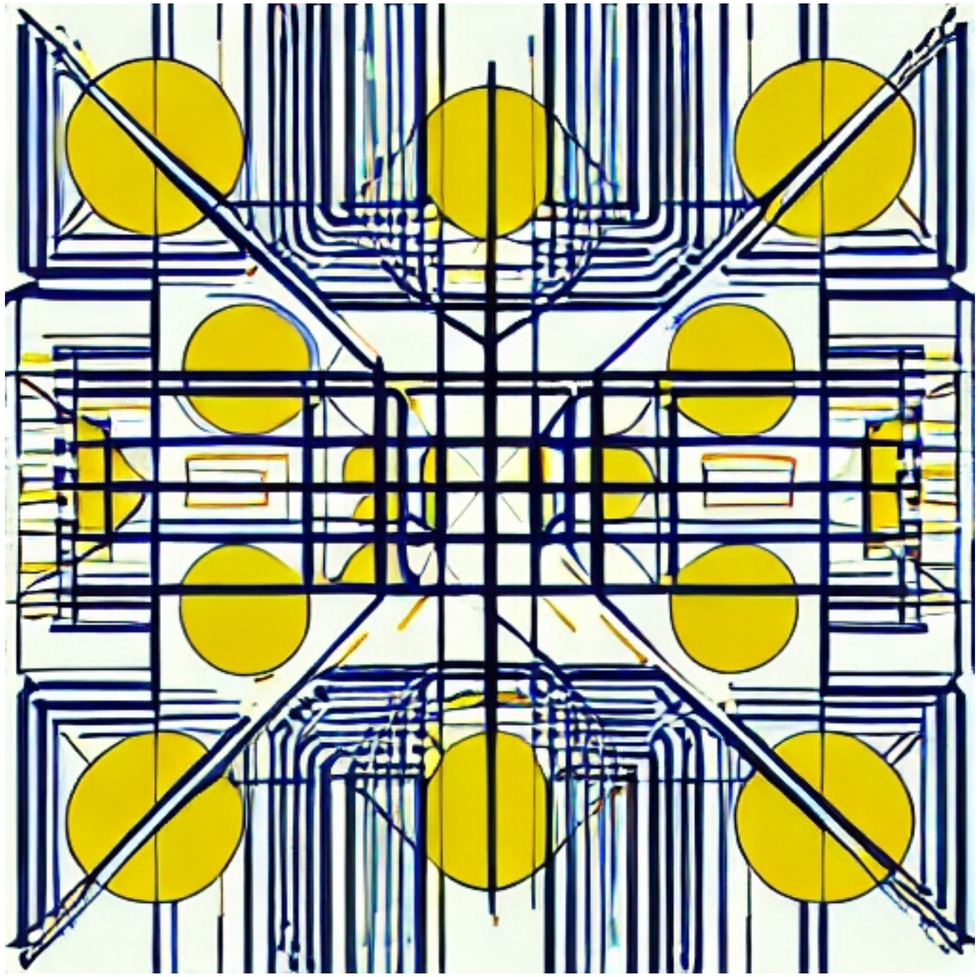
Learning from NN's learning capacity



Neural Networks identify pattern
that we are not able to see

Mele, Ingrosso, Menichetti, RP, *in preparation*

Learning from NN's learning capacity

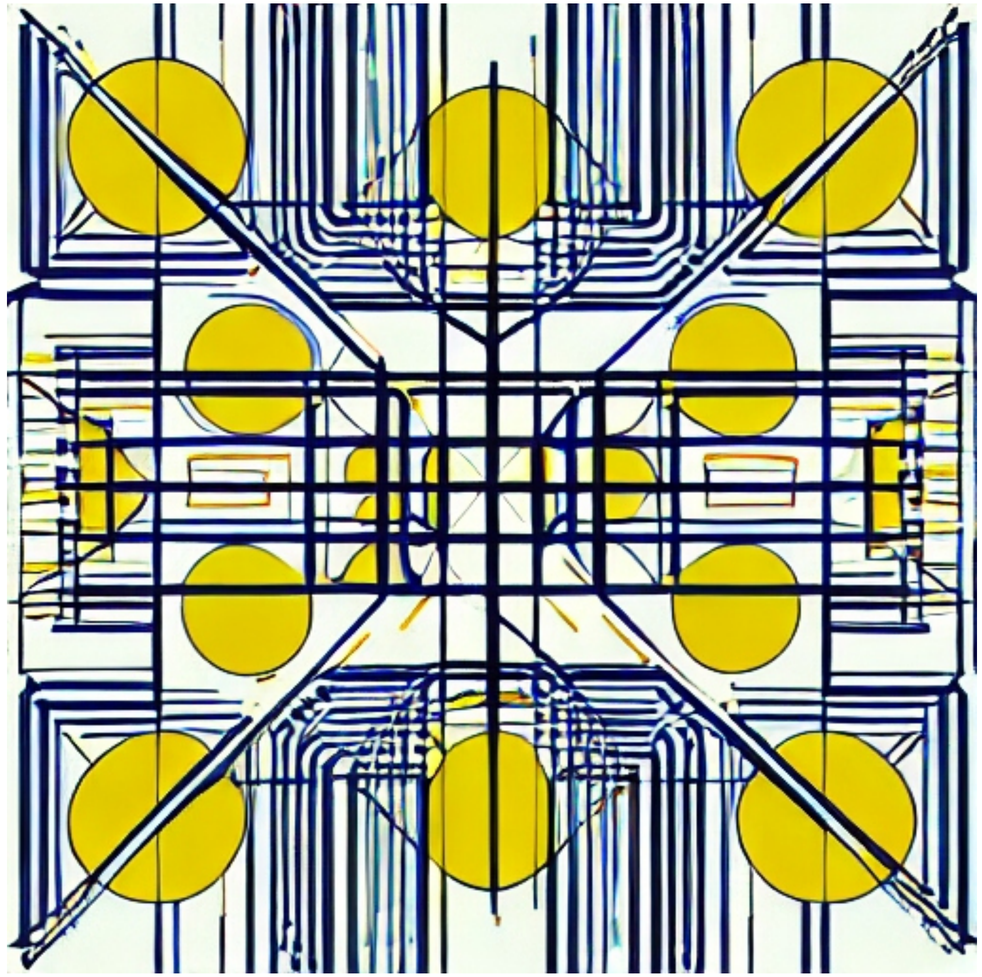


Neural Networks identify pattern
that we are not able to see

Can we identify the key data traits
networks learn from?

Mele, Ingrosso, Menichetti, RP, *in preparation*

Learning from NN's learning capacity



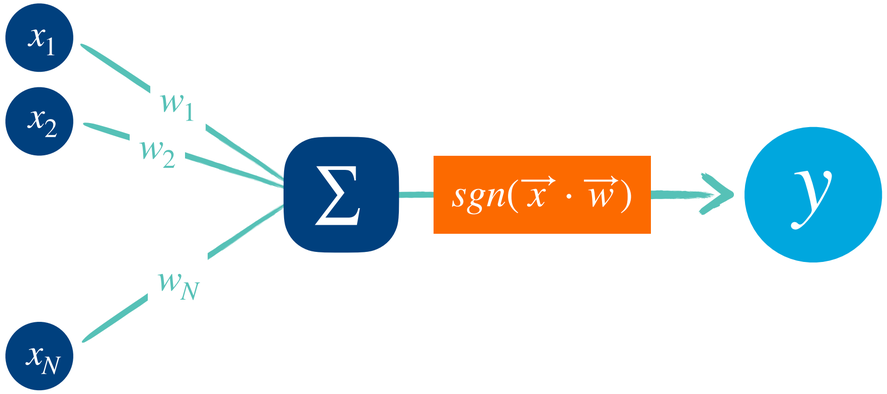
Neural Networks identify pattern
that we are not able to see

Can we identify the key data traits
networks learn from?

Describe the network configuration space while varying the input data structure

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Inputs $\{x_i\}_1^P$

$$\begin{bmatrix} 1,1,-1,\ldots,1,-1,-1 \\ [-1,1,1,\ldots,-1,1,-1] \\ \vdots \\ [-1,-1,1,\ldots,-1,1,1] \end{bmatrix}$$

P

N

Weight vector w

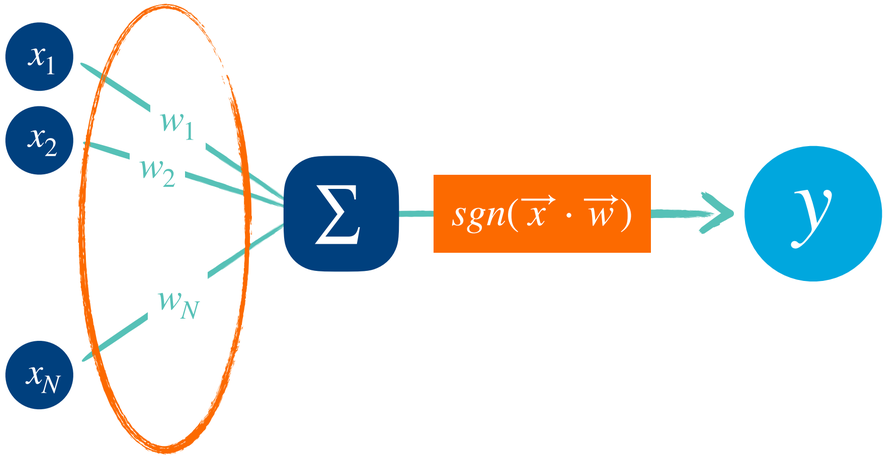
$$[-1,1,-1\ \ldots,-1,-1,1]$$

Labels $\{y_i\}$

$$y_i = \mathrm{sgn}\left(w\, x_i\right)$$

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Inputs $\{x_i\}_1^P$

$[1,1,-1,\ldots,1,-1,-1]$
$[-1,1,1,\ldots,-1,1,-1]$
$\vdots$
$[-1,-1,1,\ldots,-1,1,1]$

P

N

Weight vector w

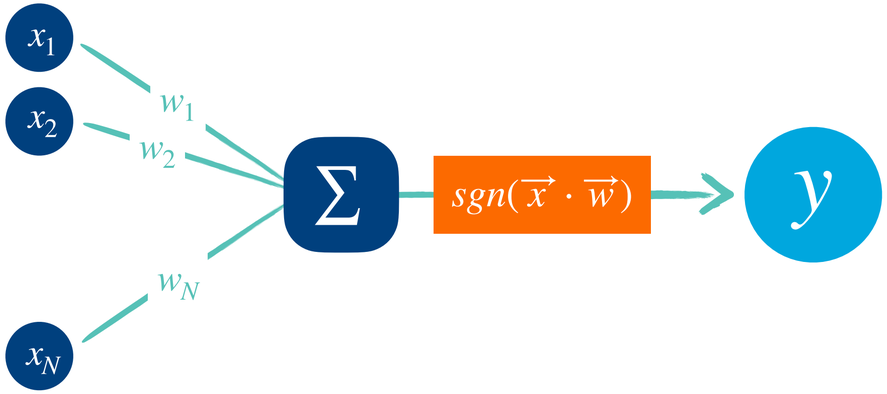
$[-1,1,-1\ \ldots,-1,-1,1]$

Labels $\{y_i\}$

$y_i = \mathrm{sgn}\left(w\, x_i\right)$

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Energy

$$E_w = \sum_i^P \Theta\left(-y_i \cdot \mathrm{sgn}\left(w\, x_i\right)\right)$$

Entropy

$$S(\overline{E}) = \log\left(\sum_{\{w\}} \delta(E_w - \overline{E})\right)$$

Inputs $\{x_i\}_1^P$

$$\begin{matrix} [1,1,-1,\ldots,1,-1,-1] \\ [-1,1,1,\ldots,-1,1,-1] \\ \vdots \\ [-1,-1,1,\ldots,-1,1,1] \end{matrix}$$

P

N

Weight vector w

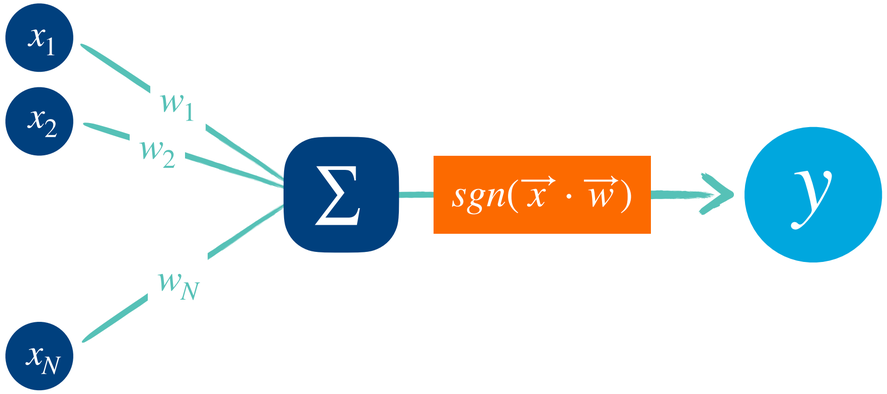
$$[-1,1,-1\ \ldots,-1,-1,1]$$

Labels $\{y_i\}$

$$y_i = \mathrm{sgn}\left(w\, x_i\right)$$

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Energy

$$E_w = \sum_i^P \Theta\left(-y_i \cdot \mathrm{sgn}\left(w\, x_i\right)\right)$$

Number of Errors

Entropy

$$S(\overline{E}) = \log\left(\sum_{\{w\}} \delta(E_w - \overline{E})\right)$$

Inputs $\{x_i\}_1^P$

$$\begin{bmatrix} 1,1,-1,\ldots,1,-1,-1 \\ [-1,1,1,\ \ldots,-1,1,-1] \\ \vdots \\ [-1,-1,1,\ \ldots,-1,1,1] \end{bmatrix}$$

P

N

Weight vector w

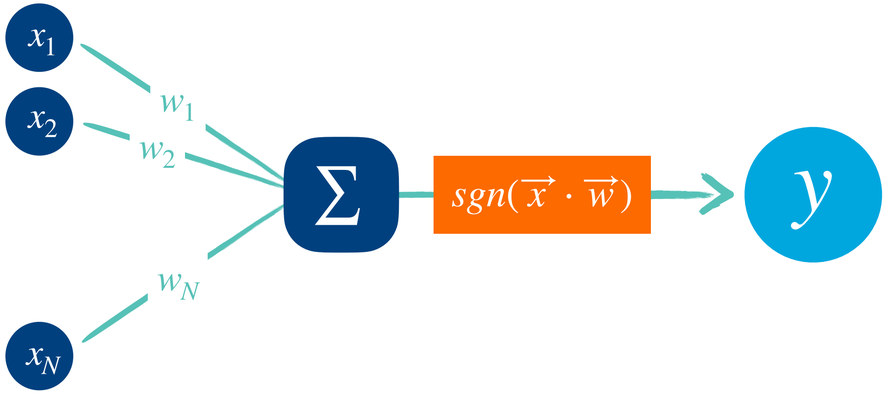
$$[-1,1,-1\ \ldots,-1,-1,1]$$

Labels $\{y_i\}$

$$y_i = \mathrm{sgn}\left(w\, x_i\right)$$

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Energy

$$E_w = \sum_i^P \Theta\left(-y_i \cdot \mathrm{sgn}\left(w\, x_i\right)\right)$$

Number of Errors

Entropy

$$S(\overline{E}) = \log\left(\sum_{\{w\}} \delta(E_w - \overline{E})\right)$$

Inputs $\{x_i\}_1^P$

$$\begin{bmatrix} 1,1,-1,\ldots,1,-1,-1 \\ -1,1,1,\ldots,-1,1,-1 \\ \vdots \\ -1,-1,1,\ldots,-1,1,1 \end{bmatrix}$$

P

N

Weight vector w

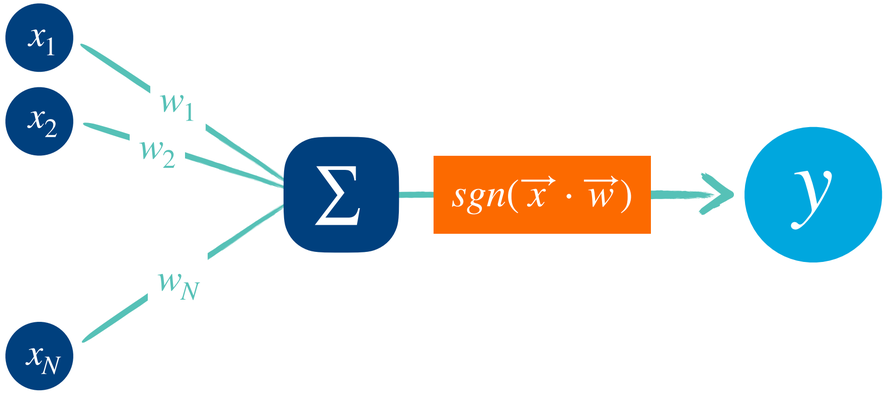
$$[-1,1,-1\ \ldots,-1,-1,1]$$

Labels $\{y_i\}$

$$y_i = \mathrm{sgn}\left(w\, x_i\right)$$

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Energy

$$E_w = \sum_i^P \Theta\left(-y_i \cdot \mathrm{sgn}\left(w\, x_i\right)\right)$$

Number of Errors

Entropy

$$S(\overline{E}) = \log\left(\sum_{\{w\}} \delta(E_w - \overline{E})\right)$$

N big (>30) → HARD

Inputs $\{x_i\}_1^P$

$$\begin{bmatrix} 1,1,-1,\ldots,1,-1,-1 \\ -1,1,1,\ldots,-1,1,-1 \\ \vdots \\ -1,-1,1,\ldots,-1,1,1 \end{bmatrix}$$

P × N

Weight vector w

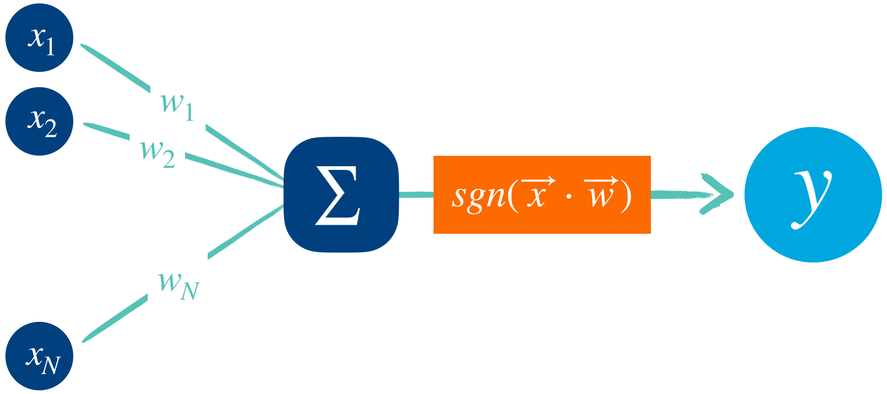
$$[-1,1,-1\ \ldots,-1,-1,1]$$

Labels $\{y_i\}$

$$y_i = \mathrm{sgn}\left(w\, x_i\right)$$

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Energy

$$E_w = \sum_i^P \Theta\left(-y_i \cdot \mathrm{sgn}\left(w\, x_i\right)\right)$$

Number of Errors

Entropy

$$S(\overline{E}) = \log\left(\sum_{\{w\}} \delta(E_w - \overline{E})\right)$$

N big (>30) → HARD

Inputs $\{x_i\}_1^P$

$$\begin{bmatrix} 1,1,-1,\ldots,1,-1,-1 \\ [-1,1,1,\ldots,-1,1,-1] \\ \vdots \\ [-1,-1,1,\ldots,-1,1,1] \end{bmatrix}$$

P

N

Weight vector w

$$[-1,1,-1\ \ldots,-1,-1,1]$$

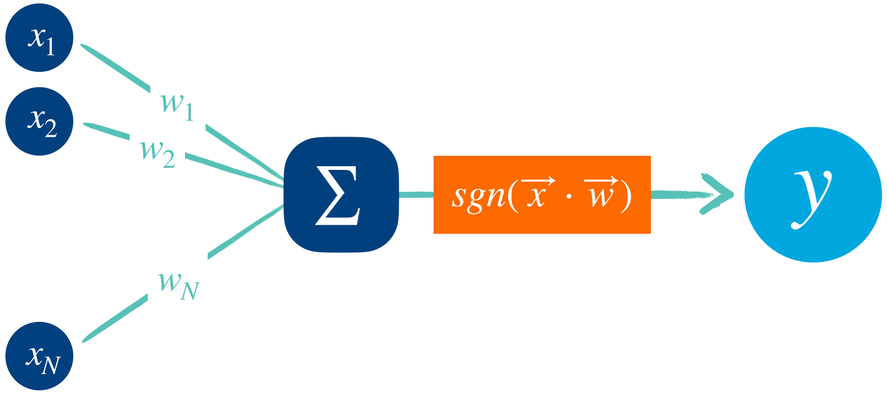
Labels $\{y_i\}$

$$y_i = \mathrm{sgn}\left(w\, x_i\right)$$

- Huge configurational space $2^{30} = 1073741824$

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Energy

$$E_w = \sum_i^P \Theta\left(-y_i \cdot \mathrm{sgn}\left(w\, x_i\right)\right)$$

Number of Errors

Entropy

$$S(\overline{E}) = \log\left(\sum_{\{w\}} \delta(E_w - \overline{E})\right)$$

N big (>30) → HARD

Inputs $\{x_i\}_1^P$

$$\begin{bmatrix} 1,1,-1,\ldots,1,-1,-1 \\ -1,1,1,\ldots,-1,1,-1 \\ \vdots \\ -1,-1,1,\ldots,-1,1,1 \end{bmatrix}$$

P (rows) × N (columns)

Weight vector w

$$[-1,1,-1\ \ldots,-1,-1,1]$$

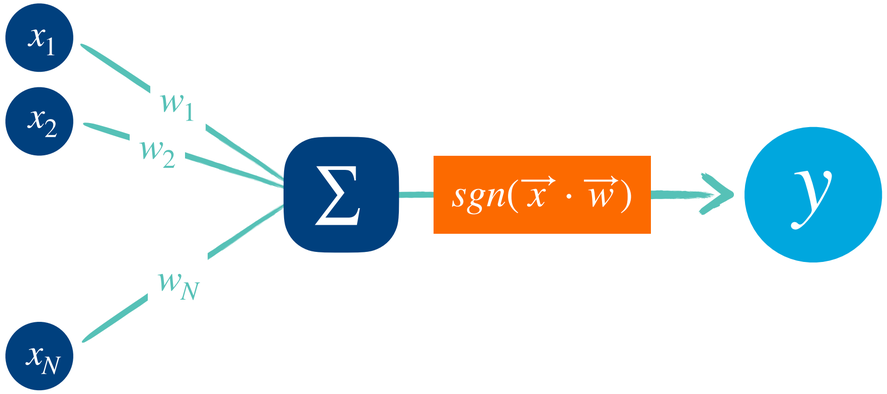
Labels $\{y_i\}$

$$y_i = \mathrm{sgn}\left(w\, x_i\right)$$

- Huge configurational space $2^{30} = 1073741824$
- Sampling algorithms get trapped in local minima

Mele, Ingrosso, Menichetti, RP, *in preparation*

The simplest NN: the perceptron



Energy

$$E_w = \sum_i^P \Theta\left(-y_i \cdot \mathrm{sgn}\left(w\, x_i\right)\right)$$

Number of Errors

Entropy

$$S(\overline{E}) = \log\left(\sum_{\{w\}} \delta(E_w - \overline{E})\right)$$

N big (>30) → HARD

Inputs $\{x_i\}_1^P$

$$\begin{bmatrix} 1,1,-1,\ldots,1,-1,-1 \\ -1,1,1,\ldots,-1,1,-1 \\ \vdots \\ -1,-1,1,\ldots,-1,1,1 \end{bmatrix}$$

P ↕ N ↔

Weight vector w

$$[-1,1,-1\ \ldots,-1,-1,1]$$

Labels $\{y_i\}$

$$y_i = \mathrm{sgn}\left(w\, x_i\right)$$

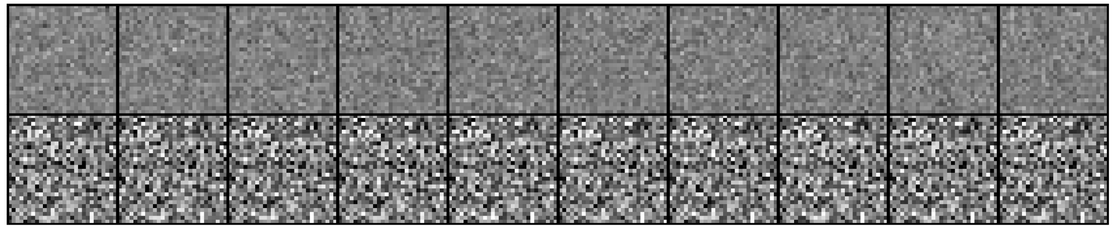
- Huge configurational space $2^{30} = 1073741824$
- Sampling algorithms get trapped in local minima

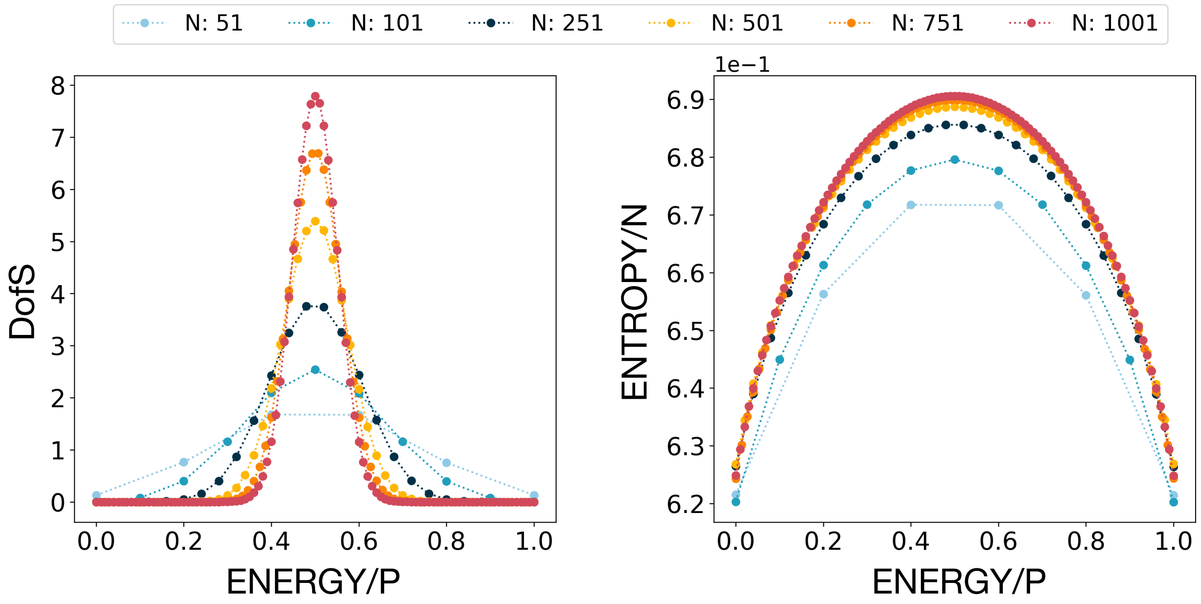
→ **Self-consistent entropy estimation**

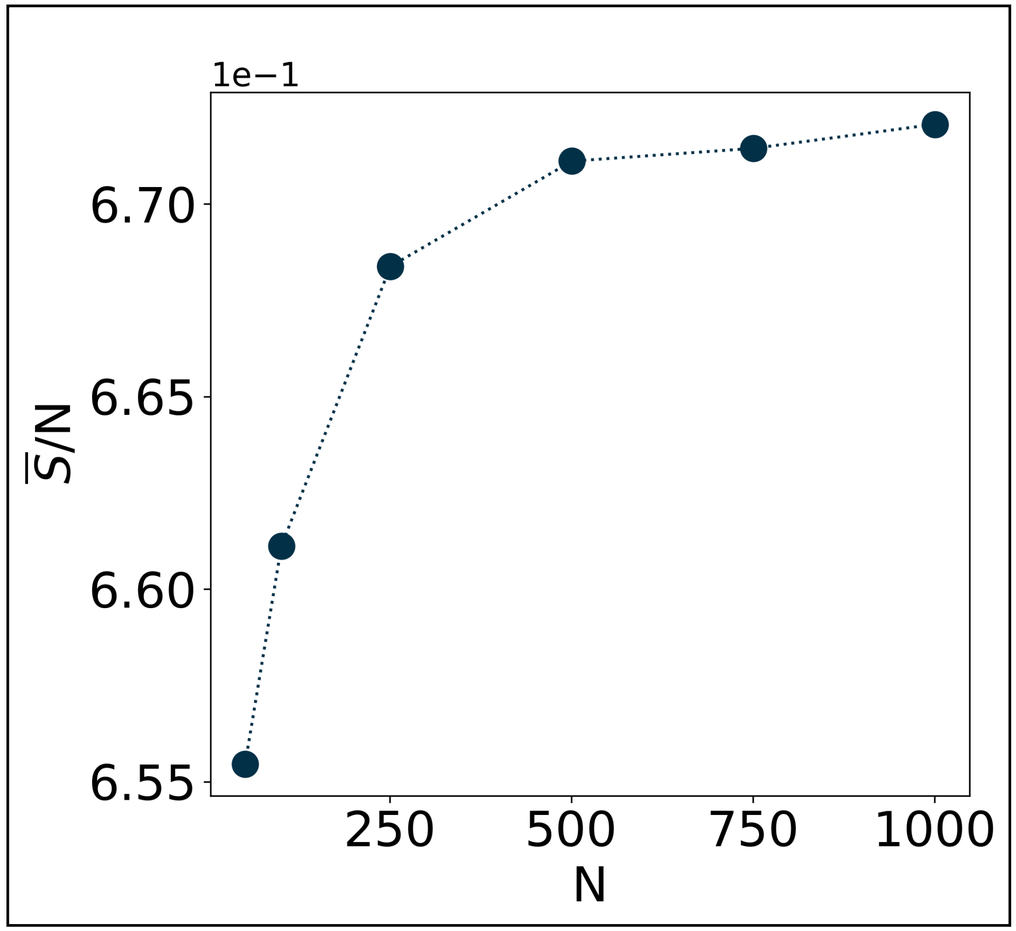
→ **Uniform exploration of the energy spectrum**

Mele, Ingrosso, Menichetti, RP, *in preparation*

Random data



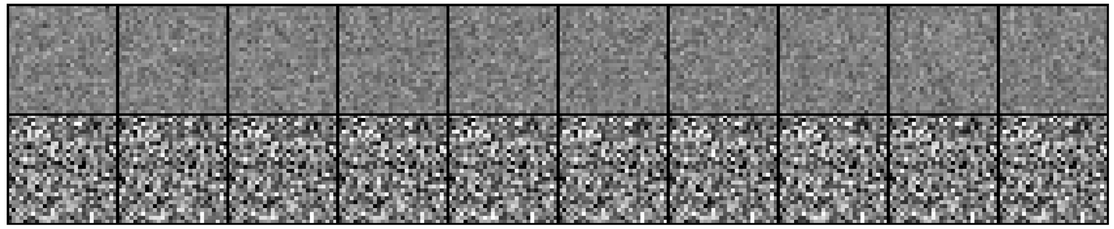


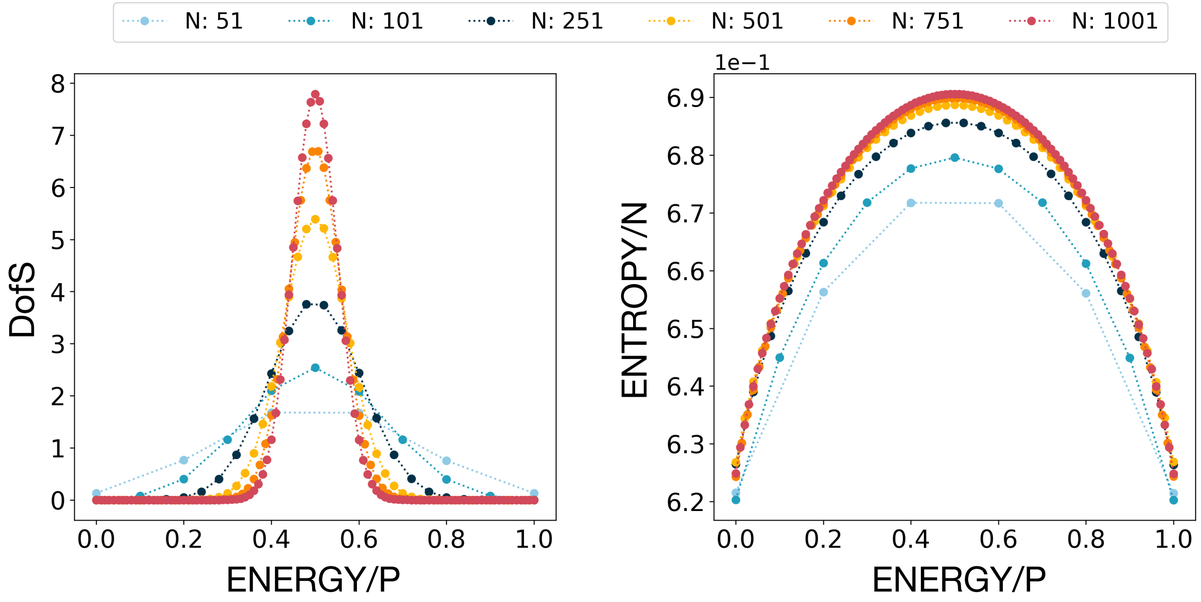


Mele, Ingrosso, Menichetti, RP, *in preparation*

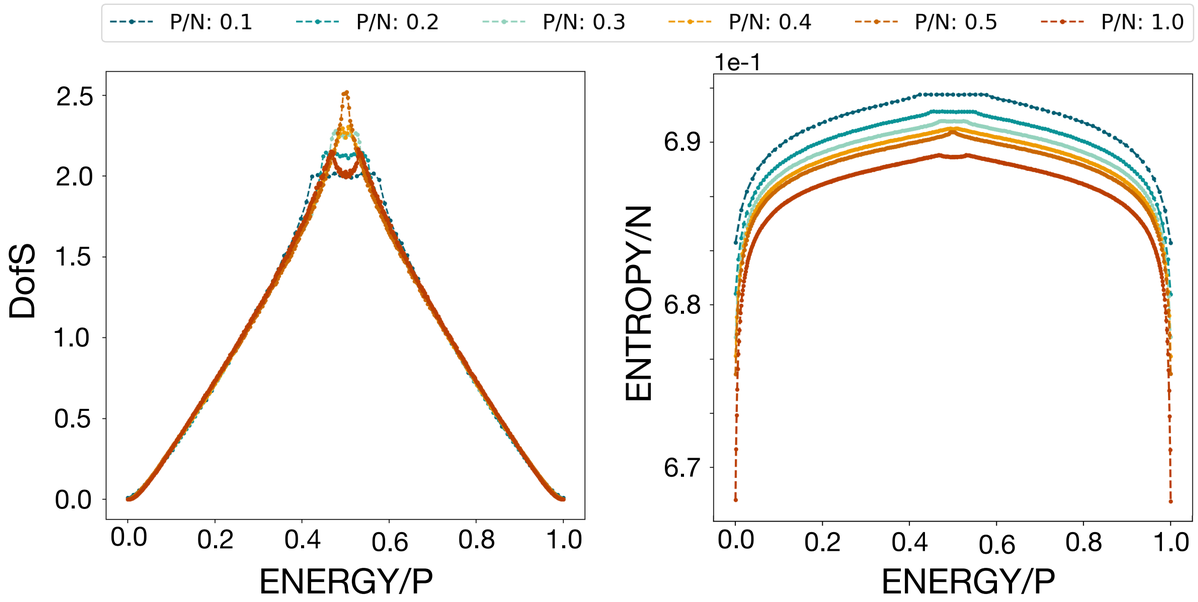
Random vs. real(istic) data

Random



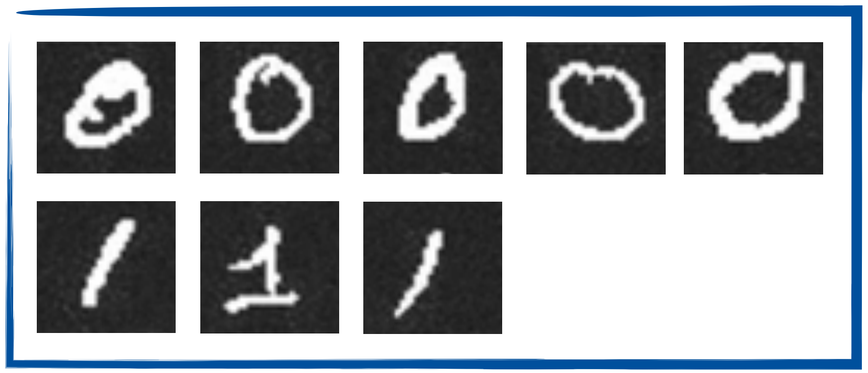


MNIST



Mele, Ingrosso, Menichetti, RP, *in preparation*

Class unbalancing

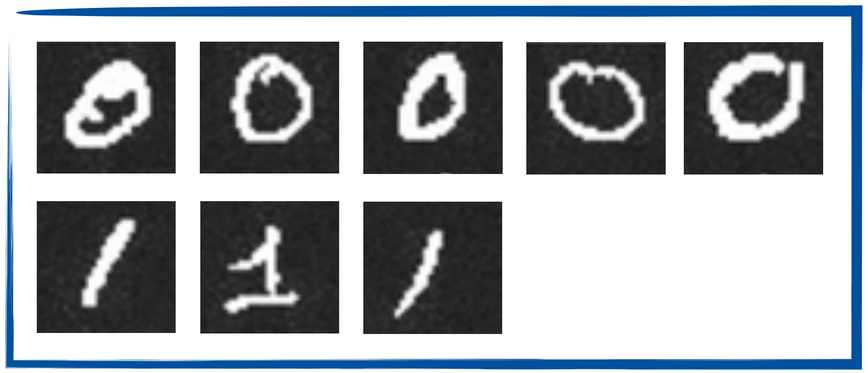


$$P_0 \neq P_1$$

$$\frac{P_1}{P} \neq 0.5$$

Mele, Ingrosso, Menichetti, RP, *in preparation*

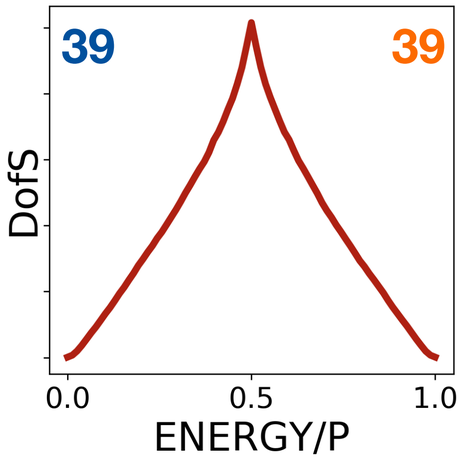
Class unbalancing



$$P_0 \neq P_1$$

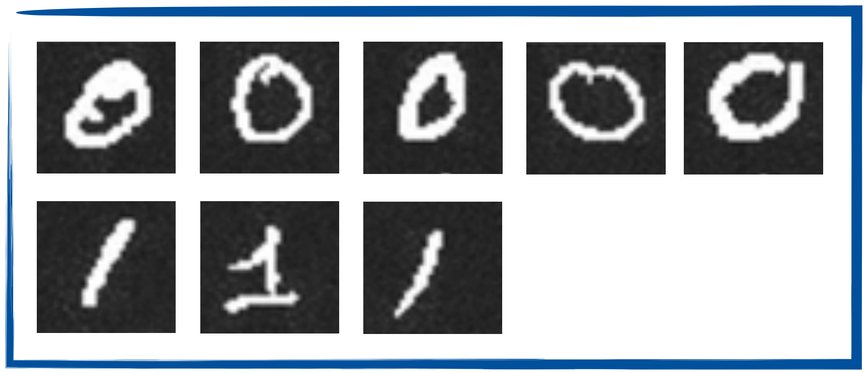
$$\frac{P_1}{P} \neq 0.5$$

$$P_1 = P_0$$



Mele, Ingrosso, Menichetti, RP, *in preparation*

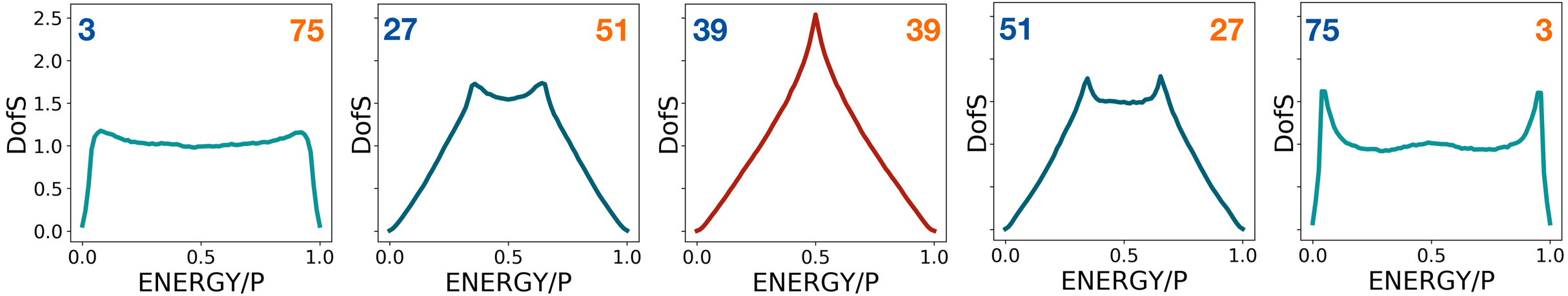
Class unbalancing



$$P_0 \neq P_1$$

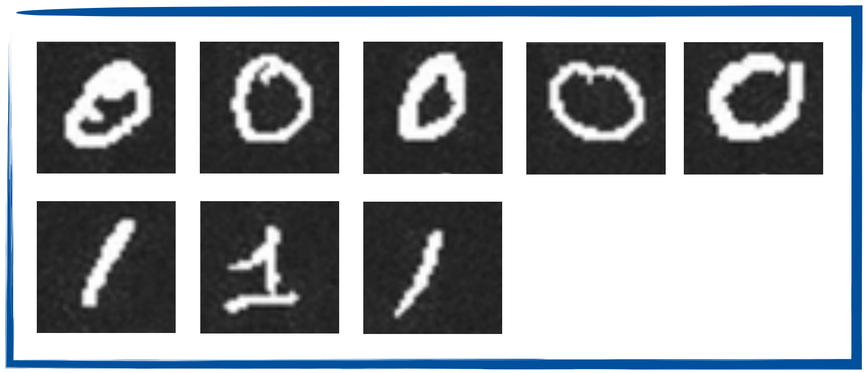
$$\frac{P_1}{P} \neq 0.5$$

$P_1 \ll P_0$ $P_1 = P_0$ $P_1 \gg P_0$



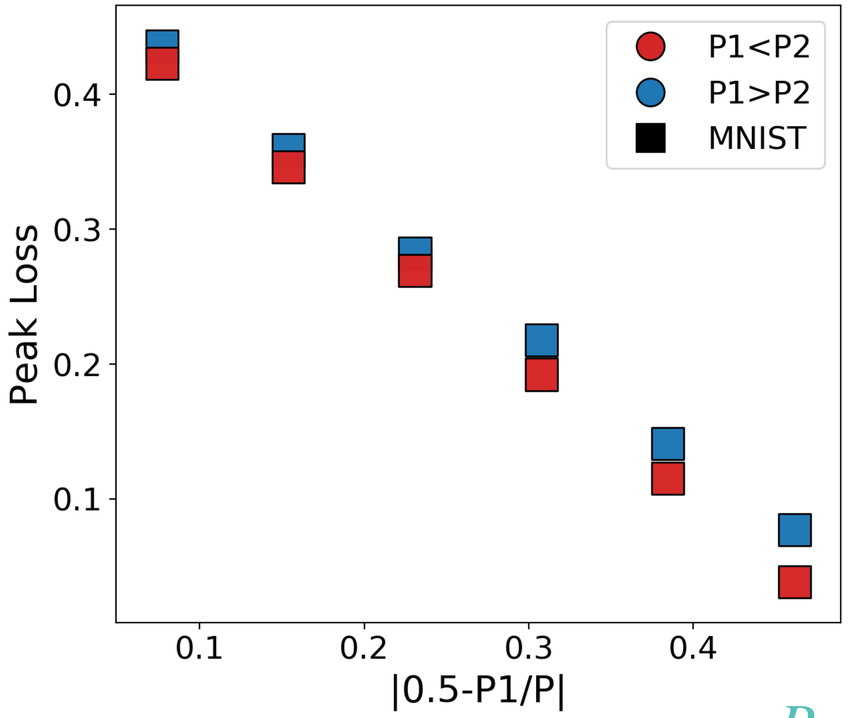
Mele, Ingrosso, Menichetti, RP, *in preparation*

Class unbalancing

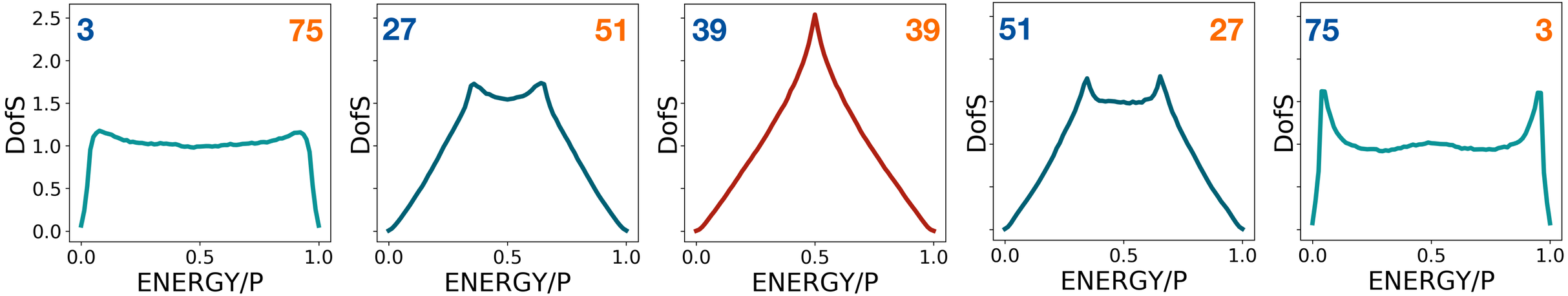


$$P_0 \neq P_1$$

$$\frac{P_1}{P} \neq 0.5$$



$P_1 \ll P_0$ $P_1 = P_0$ $P_1 \gg P_0$

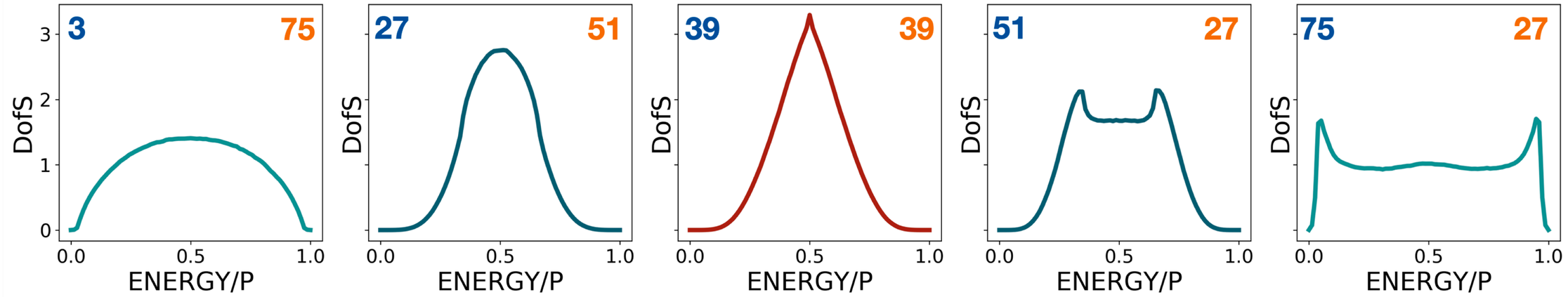


Mele, Ingrosso, Menichetti, RP, *in preparation*

Class unbalancing

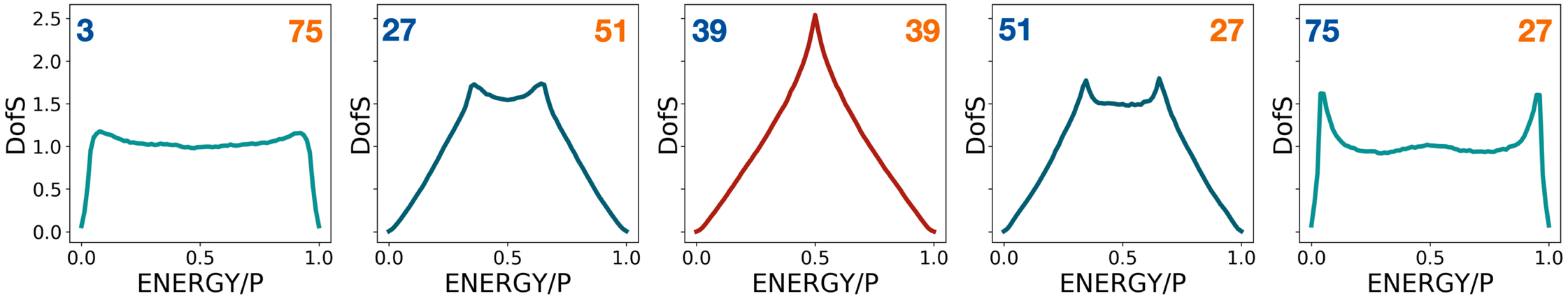
5 vs 1

$P_1 \ll P_5$ $P_1 \gg P_5$



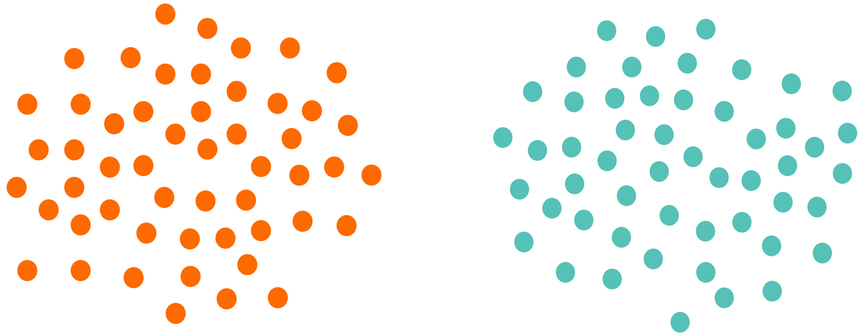
0 vs 1

$P_1 \ll P_0$ $P_1 \gg P_0$



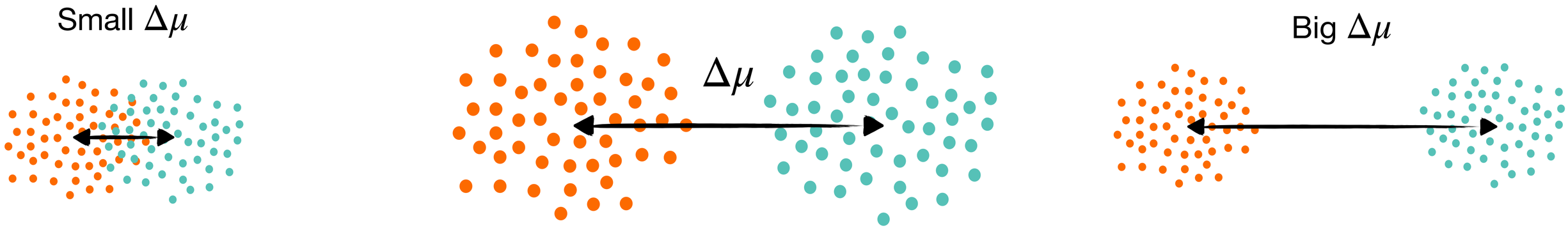
Mele, Ingrosso, Menichetti, RP, *in preparation*

Back to random data



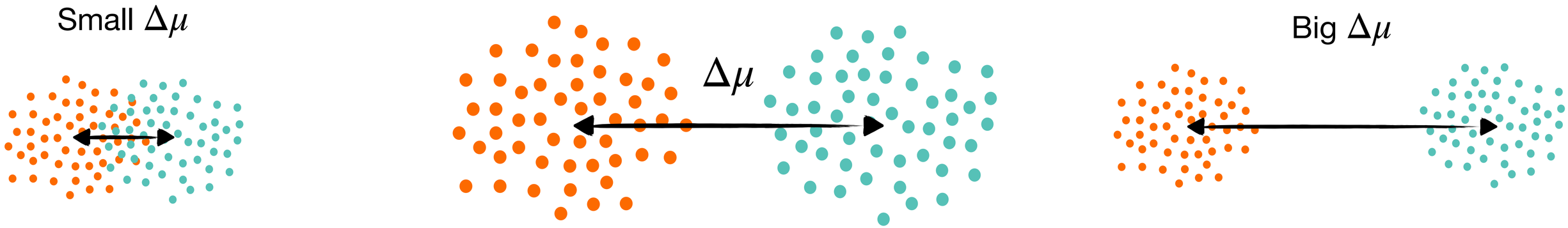
Mele, Ingrosso, Menichetti, RP, *in preparation*

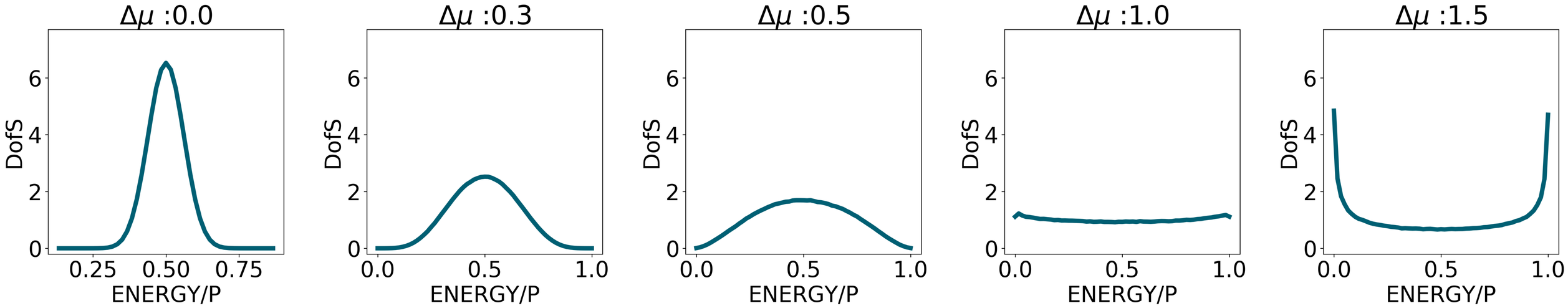
Back to random data



Mele, Ingrosso, Menichetti, RP, *in preparation*

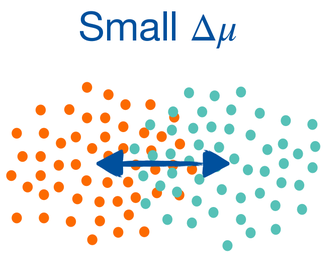
Back to random data



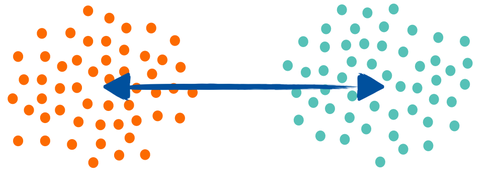


Mele, Ingrosso, Menichetti, RP, *in preparation*

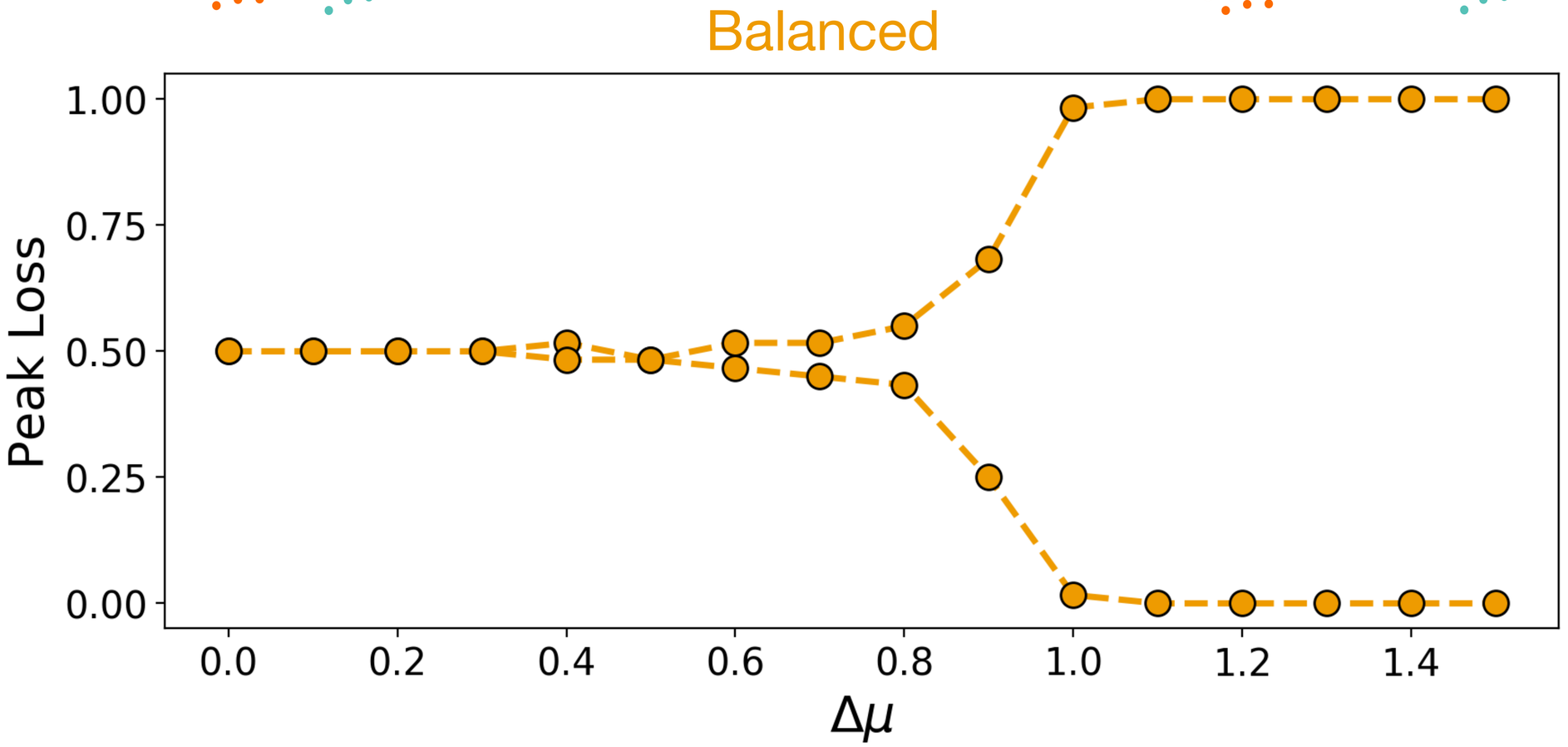
Back to random data



Small $\Delta\mu$

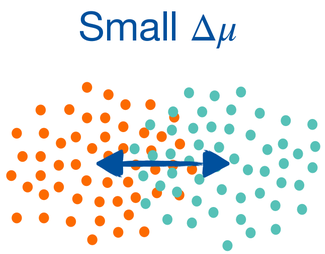


Big $\Delta\mu$

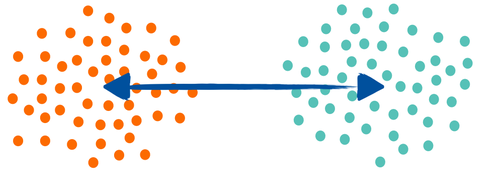


Mele, Ingrosso, Menichetti, RP, *in preparation*

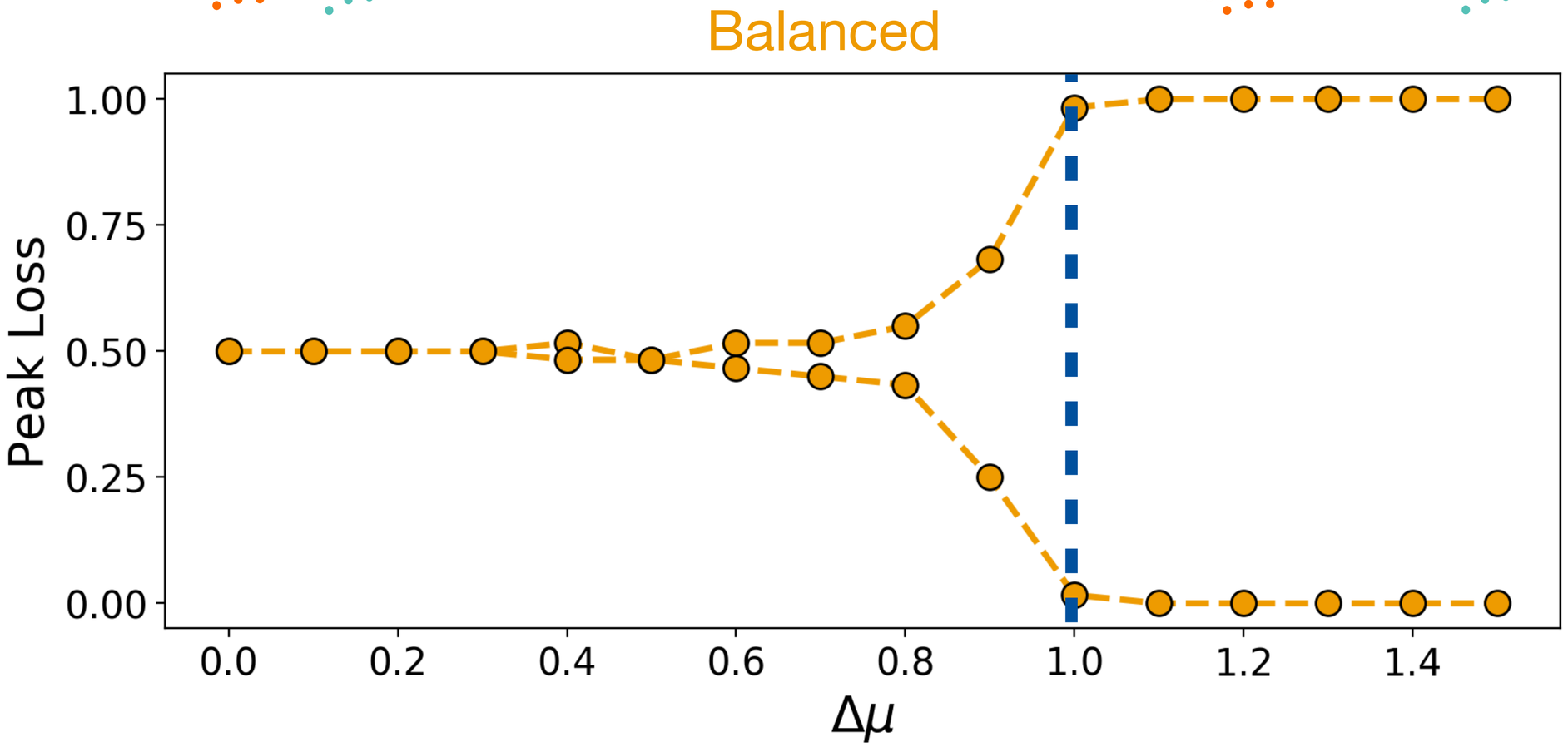
Back to random data



Small $\Delta\mu$

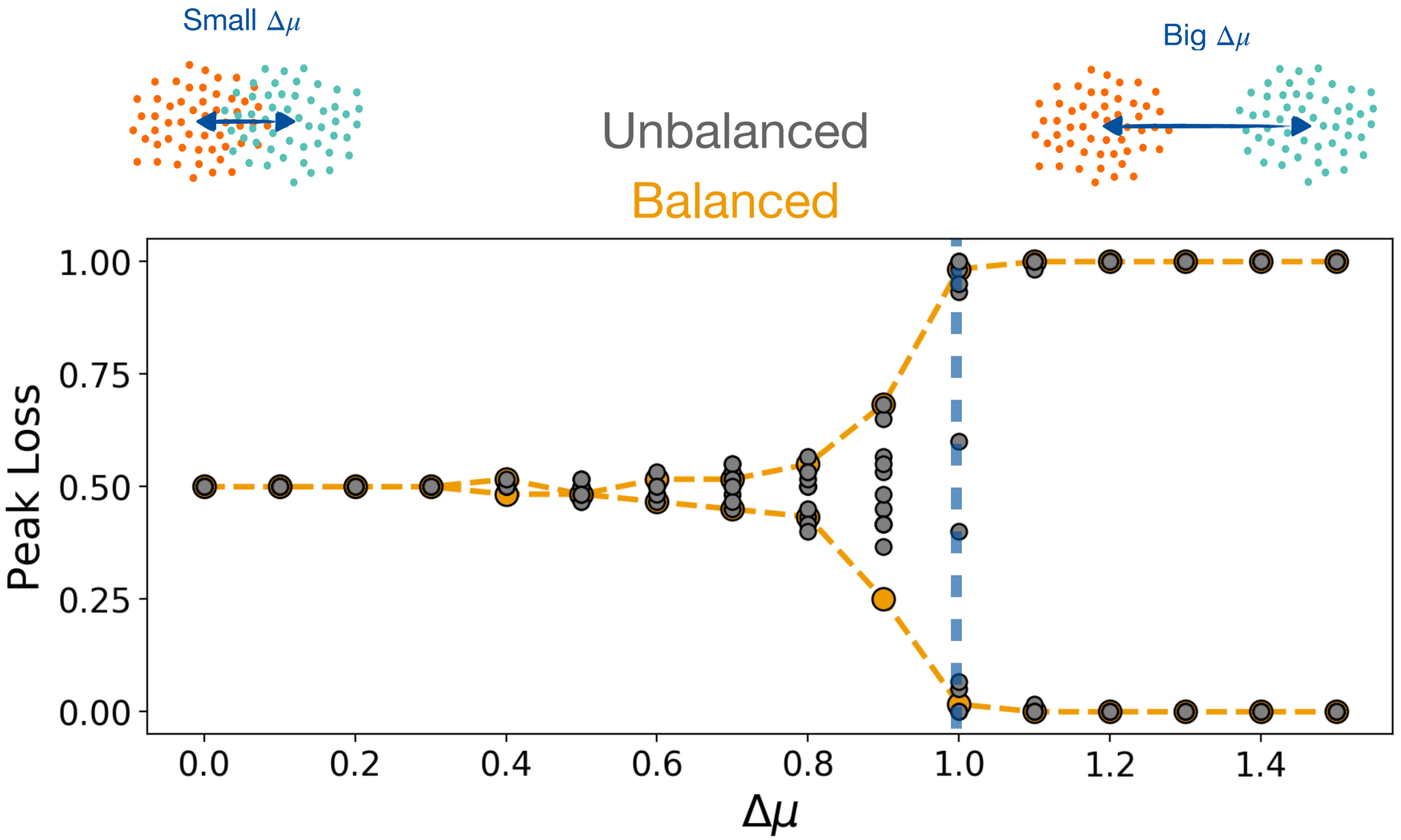


Big $\Delta\mu$



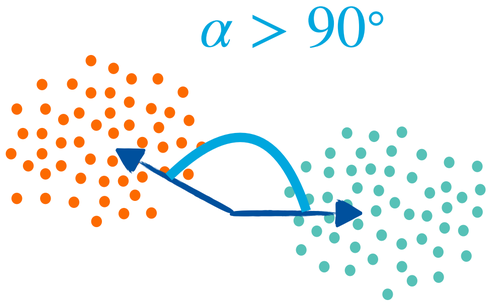
Mele, Ingrosso, Menichetti, RP, *in preparation*

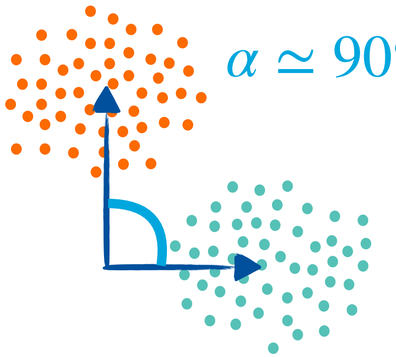
Back to random data

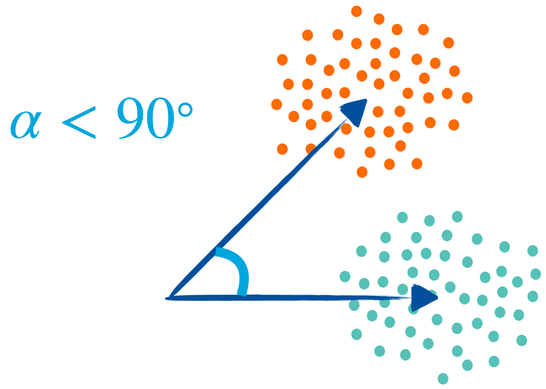


Mele, Ingrosso, Menichetti, RP, *in preparation*

Back to random data

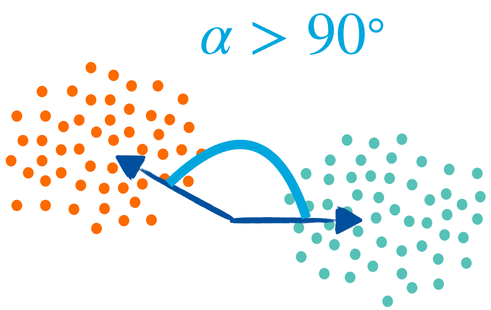


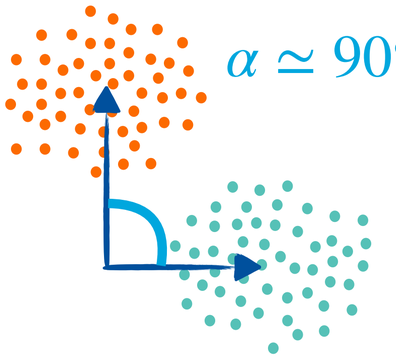


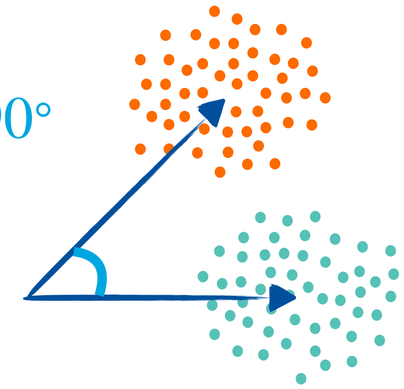


Mele, Ingrosso, Menichetti, RP, *in preparation*

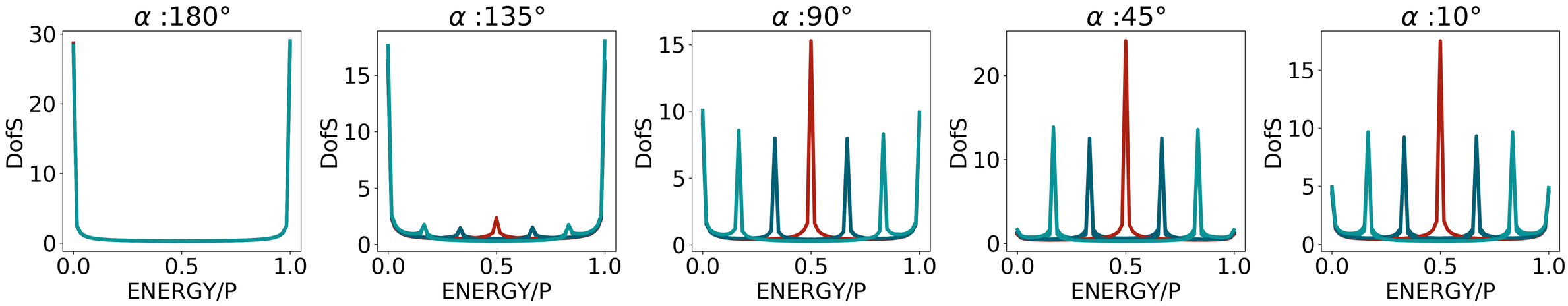
Back to random data





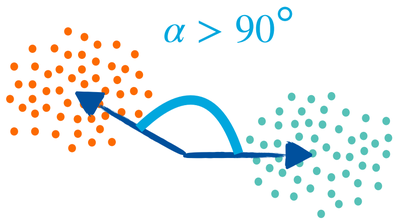


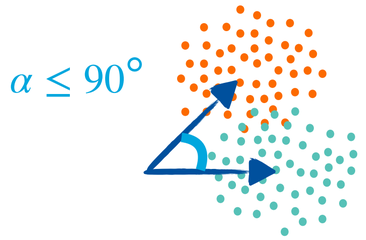
$\Delta\mu \gg 1$

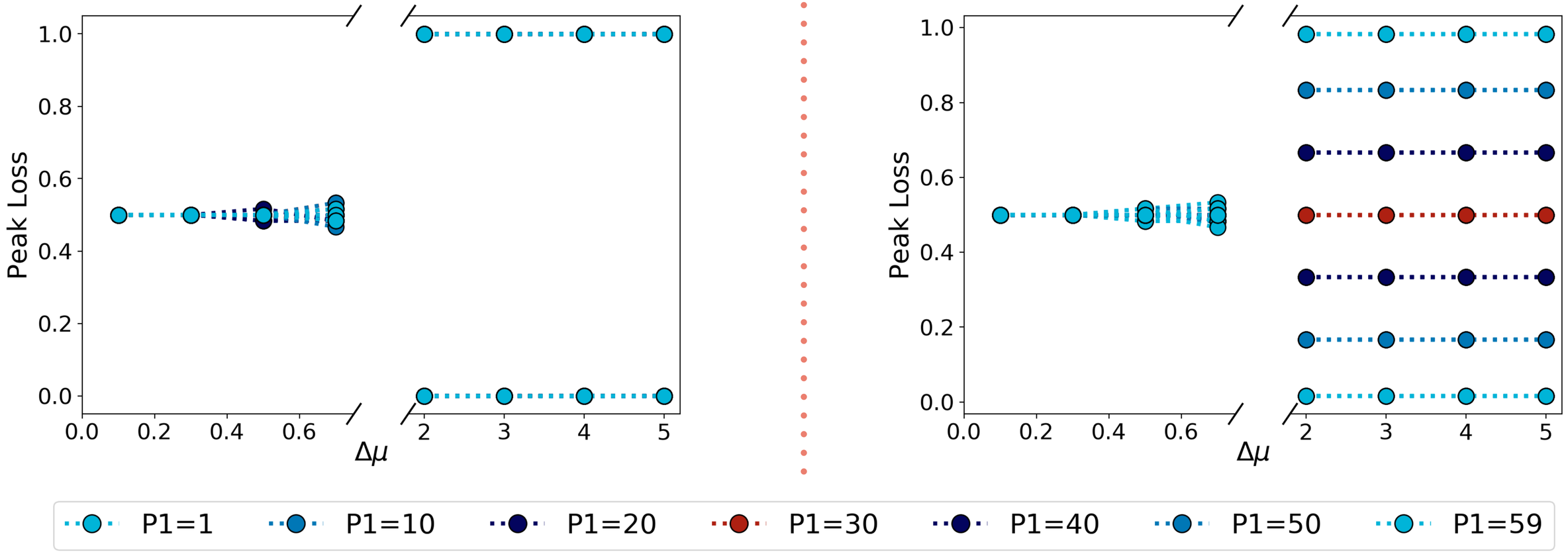


Mele, Ingrosso, Menichetti, RP, *in preparation*

Back to random data







Mele, Ingrosso, Menichetti, RP, *in preparation*

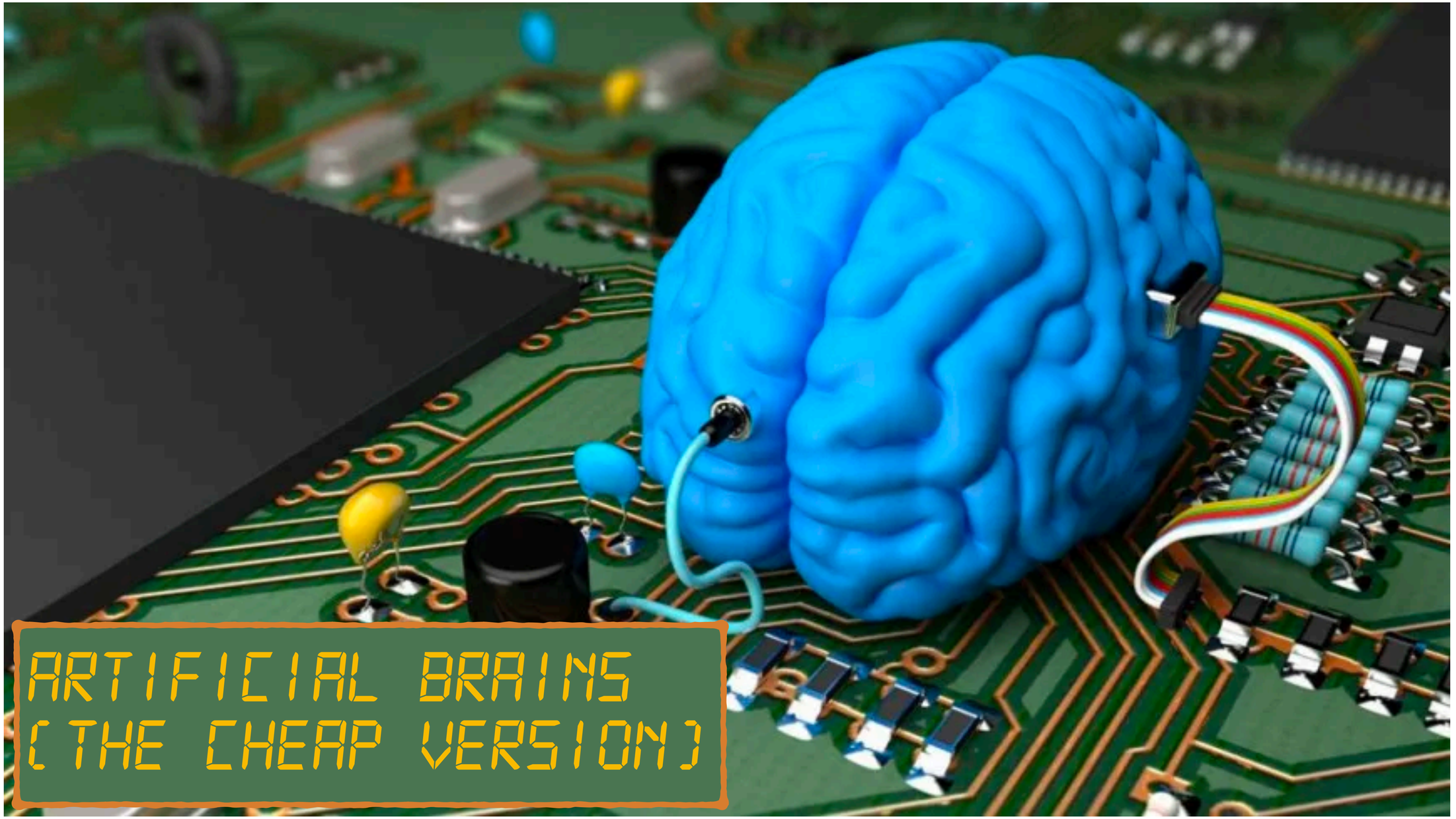
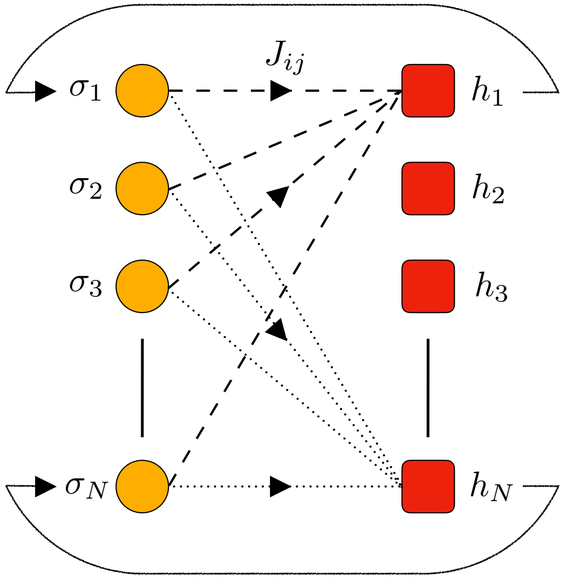


Image from https://khaleejmag.com/technology/artificial-brain-is-it-taking-up-human-roles/

Mapping optimisation of a Hopfield NN



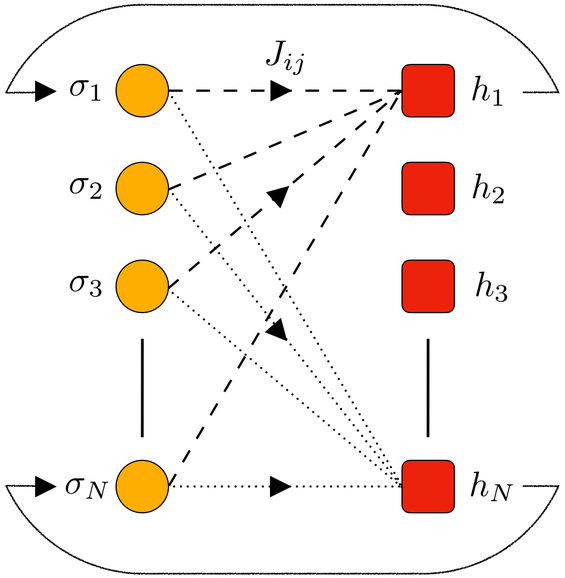
$$H(\sigma_1, ..., \sigma_N) = -\frac{1}{2} \sum_{i \neq j} J_{ij} \sigma_i \sigma_j$$

Couplings $J_{ij} = \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$

Memory patterns $\xi_i^\mu, \quad \mu = 1, ..., p$

Hopfield, PNAS (1982)
Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (1989)
Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN



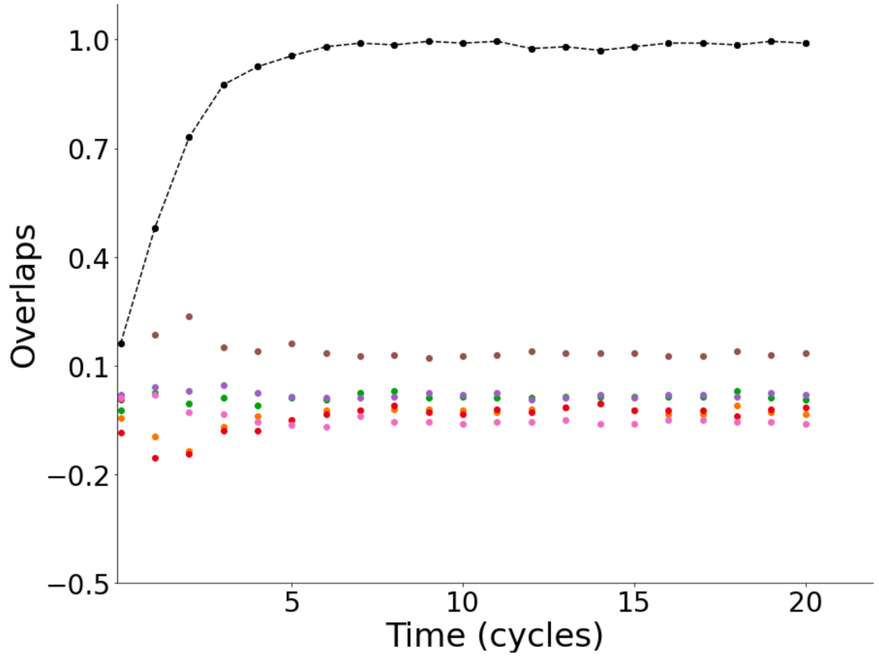
$$H(\sigma_1, ..., \sigma_N) = -\frac{1}{2}\sum_{i \neq j} J_{ij}\sigma_i\sigma_j$$

Couplings $J_{ij} = \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$

Memory patterns $\xi_i^\mu, \quad \mu = 1, ..., p$

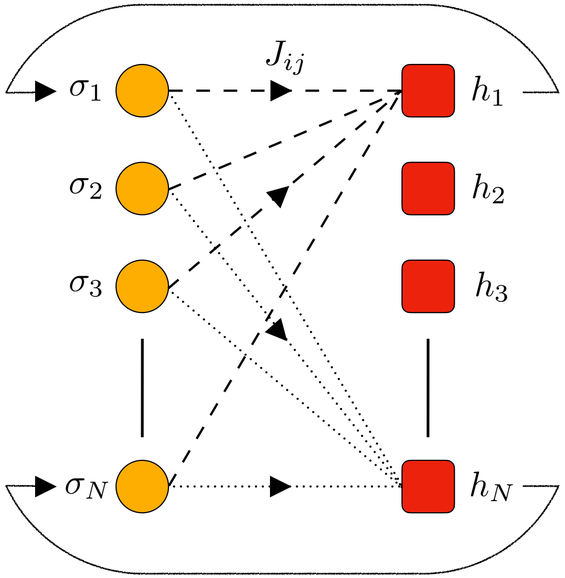
Model's dynamics retrieves memory patterns

Overlaps $m^\mu = \frac{1}{N}\sum_i \xi_i^\mu \sigma_i$



Hopfield, PNAS (1982)
Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (1989)
Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN



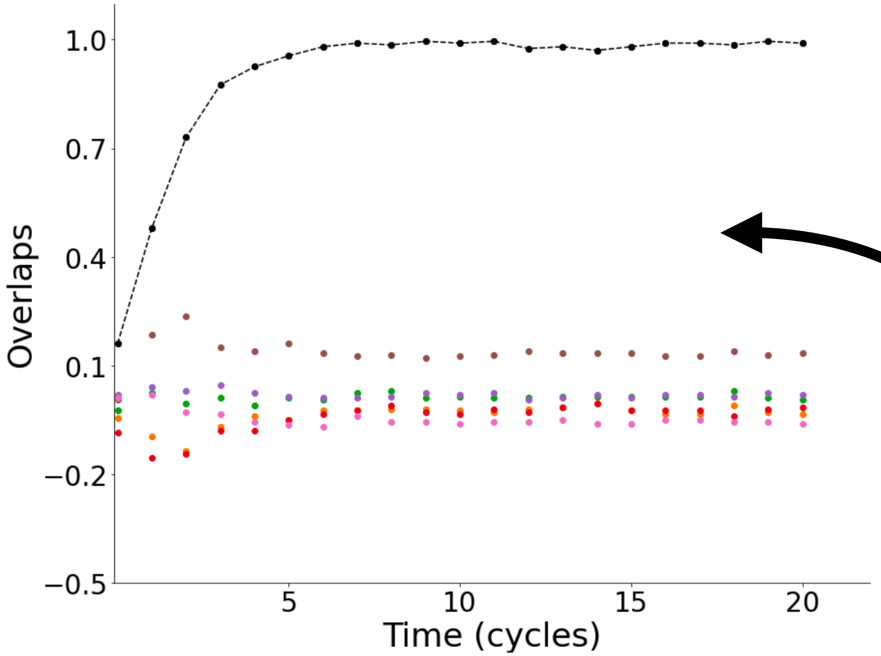
$$H(\sigma_1, ..., \sigma_N) = -\frac{1}{2}\sum_{i \neq j} J_{ij}\sigma_i\sigma_j$$

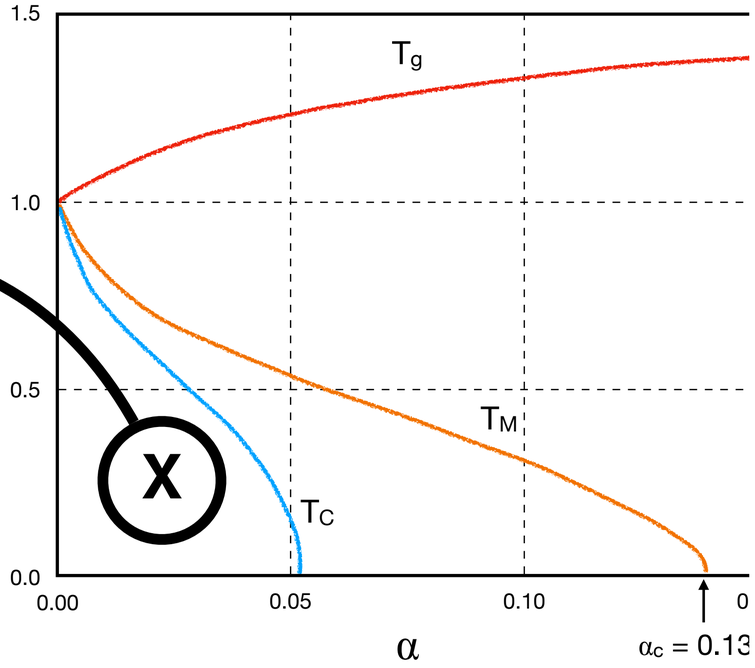
Couplings $J_{ij} = \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$

Memory patterns $\xi_i^\mu, \quad \mu = 1, ..., p$

Model's dynamics retrieves memory patterns

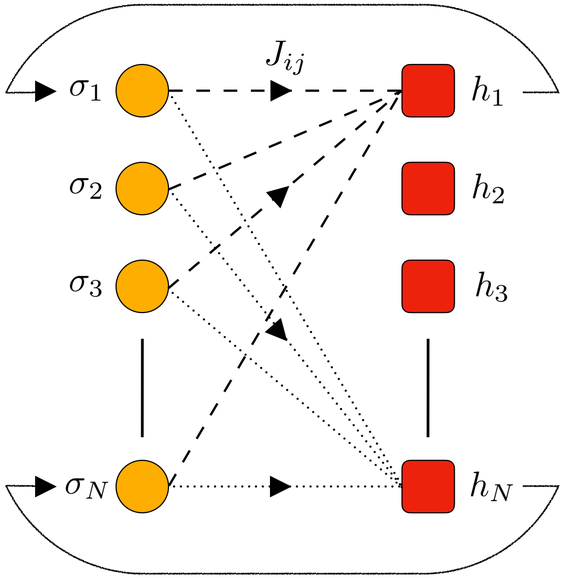
Overlaps $m^\mu = \frac{1}{N}\sum_i \xi_i^\mu \sigma_i$





Hopfield, PNAS (1982)
Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (1989)
Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN



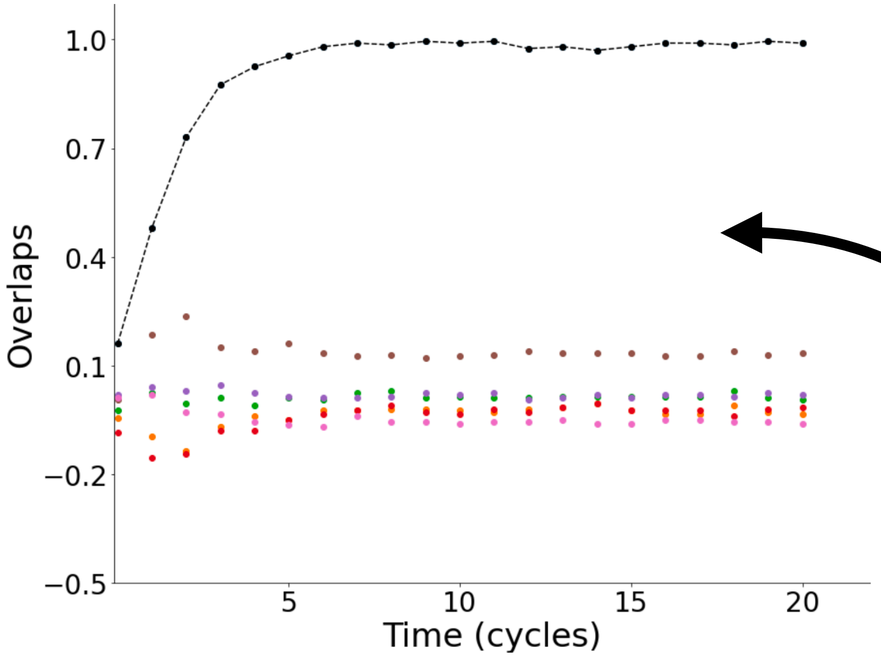
$$H(\sigma_1, ..., \sigma_N) = -\frac{1}{2}\sum_{i \neq j} J_{ij}\sigma_i\sigma_j$$

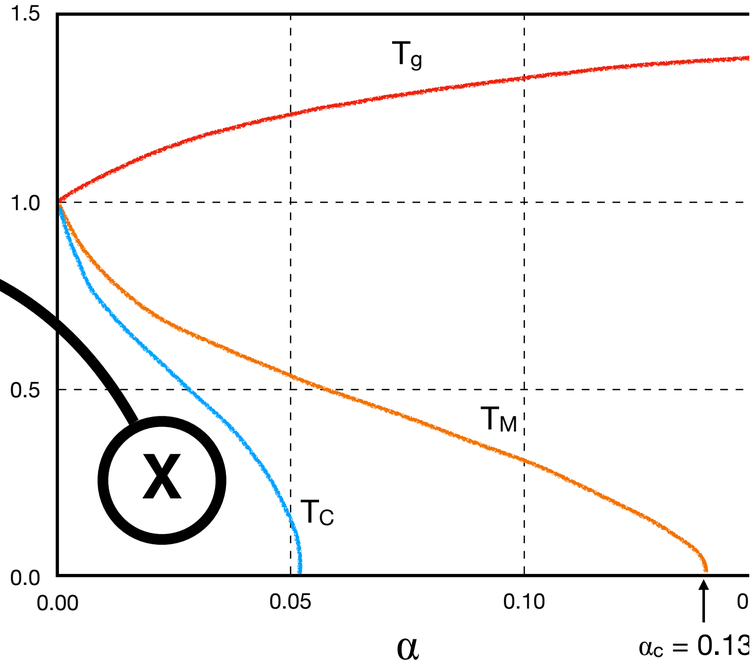
Couplings $J_{ij} = \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$

Memory patterns $\xi_i^\mu, \quad \mu = 1, ..., p$

Model's dynamics retrieves memory patterns

Overlaps $m^\mu = \frac{1}{N}\sum_i \xi_i^\mu \sigma_i$





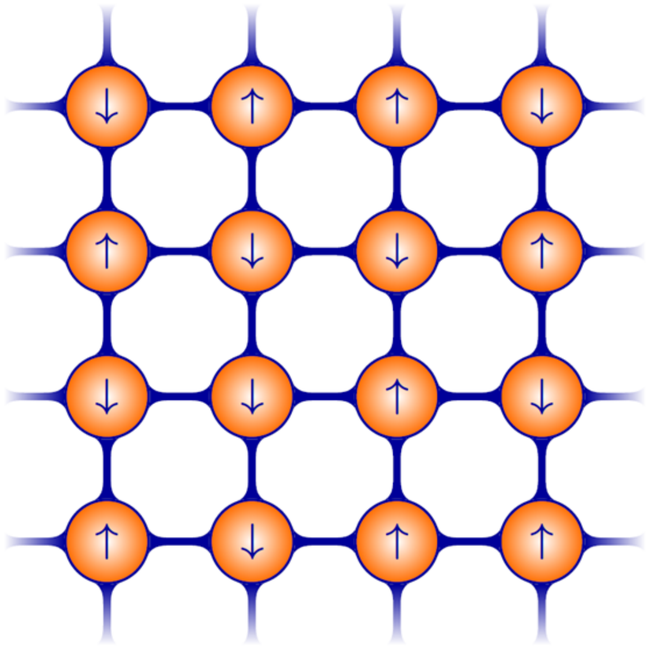
Look for the **most informative** neurons!

Hopfield, PNAS (1982)
Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (1989)
Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

CG'ing the Hopfield model

1) Simulate the Hopfield model



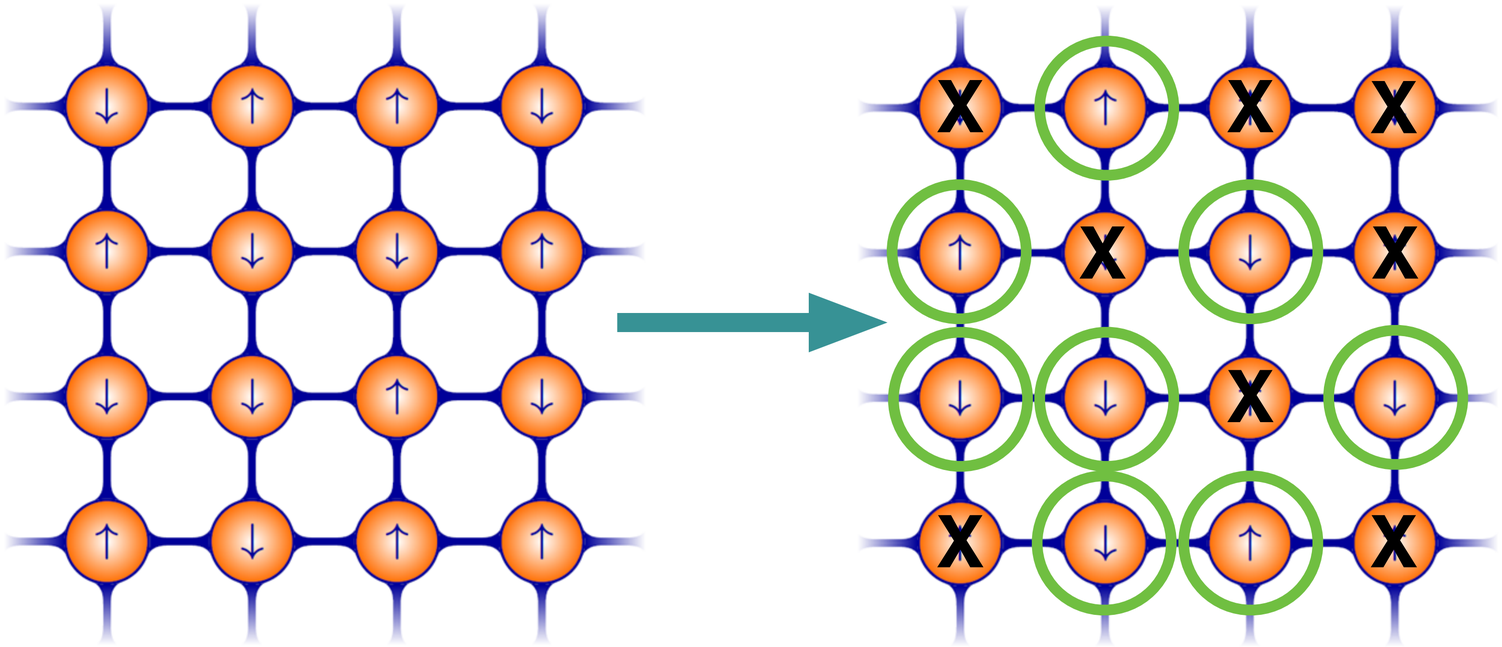
Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

CG'ing the Hopfield model

1) Simulate the Hopfield model

2) Select a subset of retained neurons



Aldrigo, Menichetti, RP, *arXiv* 2024

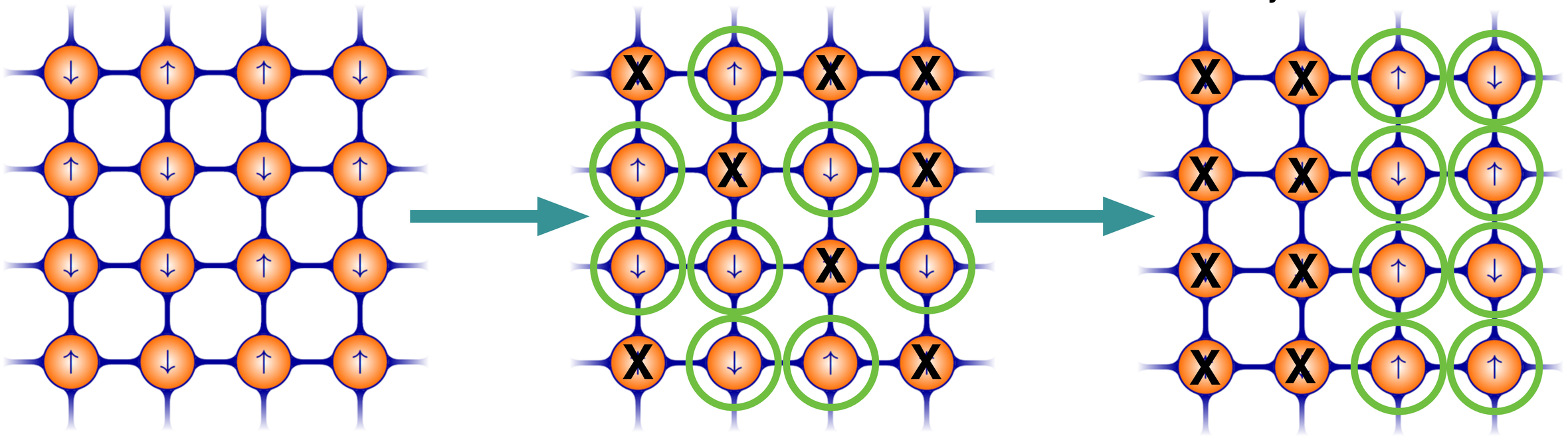
Mapping optimisation of a Hopfield NN

CG'ing the Hopfield model

1) Simulate the Hopfield model

2) Select a subset of retained neurons

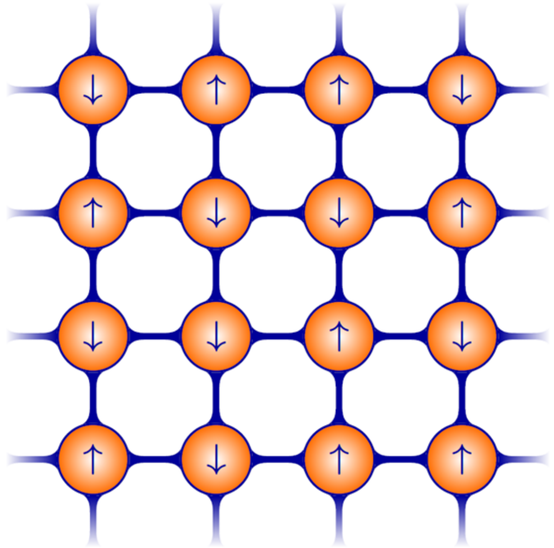
3) Optimise the selection to detect **maximally-informative** neurons



Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

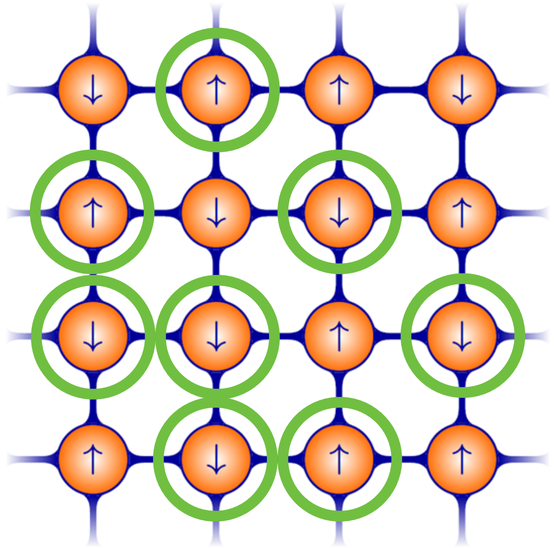
- Simulate the high-resolution Hopfield model
- **Empirical** “atomistic” probability $p(\sigma_1, ..., \sigma_N)$



Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

- Simulate the high-resolution Hopfield model
- **Empirical** “atomistic” probability $p(\sigma_1, ..., \sigma_N)$

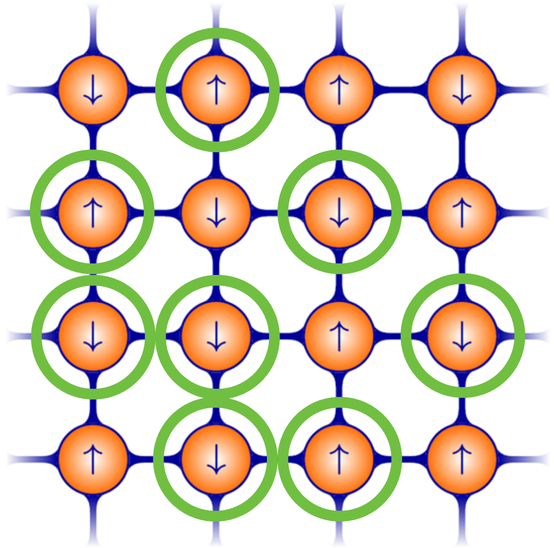


- Select n<N retained neurons S_i
- **Empirical** CG probability $P_M(S_1, ..., S_n)$
- **Empirical** backmapped probability $\bar{p}_M(\sigma_1, ..., \sigma_N)$

Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

- Simulate the high-resolution Hopfield model
- **Empirical** "atomistic" probability $p(\sigma_1, ..., \sigma_N)$



- Select n<N retained neurons S_i
- **Empirical** CG probability $P_M(S_1, ..., S_n)$
- **Empirical** backmapped probability $\bar{p}_M(\sigma_1, ..., \sigma_N)$

Resolution of the neuron selection

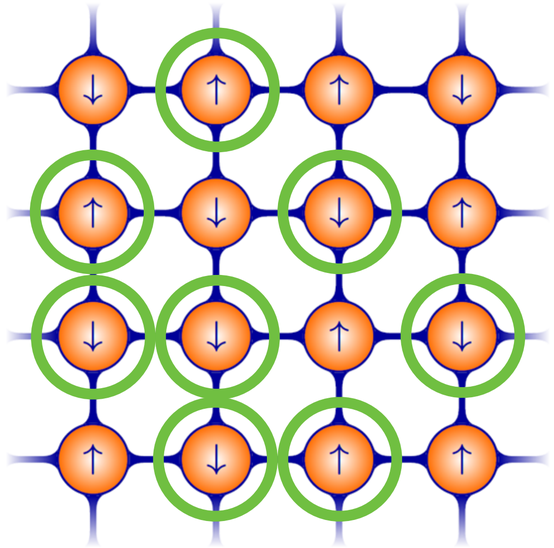
$$\mathcal{H}_M = -\sum_{\{S_i\}} P(S_1, ..., S_n) \ln P(S_1, ..., S_n)$$

- **Depends** on the specific selection
- **Decreases** by decreasing the number of retained neurons

Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

- Simulate the high-resolution Hopfield model
- **Empirical** “atomistic” probability $p(\sigma_1, ..., \sigma_N)$



- Select n<N retained neurons S_i
- **Empirical** CG probability $P_M(S_1, ..., S_n)$
- **Empirical** backmapped probability $\bar{p}_M(\sigma_1, ..., \sigma_N)$

Resolution of the neuron selection

$$\mathcal{H}_M = -\sum_{\{S_i\}} P(S_1, ..., S_n) \ln P(S_1, ..., S_n)$$

- **Depends** on the specific selection
- **Decreases** by decreasing the number of retained neurons

Information loss generated by the selection: mapping entropy

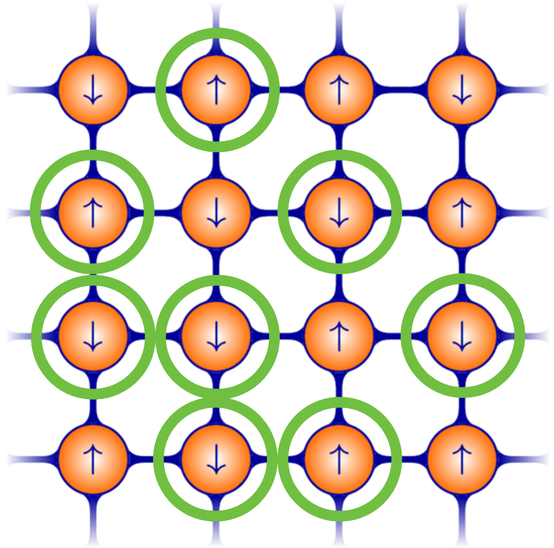
$$S_M^{map} = \sum_{\{\sigma_i\}} p(\{\sigma_i\}) \ln \left(\frac{p(\{\sigma_i\})}{\bar{p}_M(\{\sigma_i\})} \right)$$

- **Depends** on the specific selection
- **Increases** by decreasing the number of retained neurons

Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

- Simulate the high-resolution Hopfield model
- **Empirical** “atomistic” probability $p(\sigma_1, ..., \sigma_N)$



- Select n<N retained neurons S_i
- **Empirical** CG probability $P_M(S_1, ..., S_n)$
- **Empirical** backmapped probability $\bar{p}_M(\sigma_1, ..., \sigma_N)$

Resolution of the neuron selection

$$\mathcal{H}_M = -\sum_{\{S_i\}} P(S_1, ..., S_n) \ln P(S_1, ..., S_n)$$

- **Depends** on the specific selection
- **Decreases** by decreasing the number of retained neurons

Maximally informative neurons
Minimise the mapping entropy
in the space of possible selections!

Information loss generated by the selection: mapping entropy

$$S_M^{map} = \sum_{\{\sigma_i\}} p(\{\sigma_i\}) \ln \left(\frac{p(\{\sigma_i\})}{\bar{p}_M(\{\sigma_i\})} \right)$$

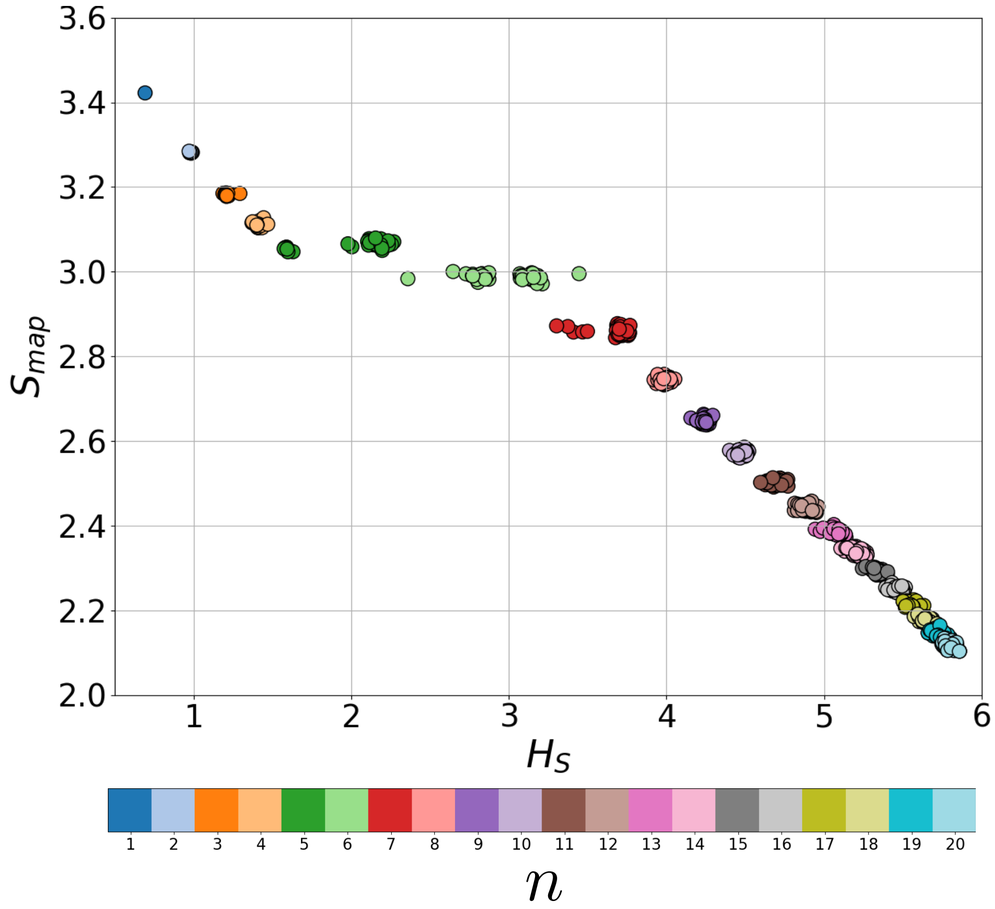
- **Depends** on the specific selection
- **Increases** by decreasing the number of retained neurons

Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

Maximally informative selection of neurons that **minimise the mapping entropy**

Hopfield model with N=100 neurons and 5 memory patterns

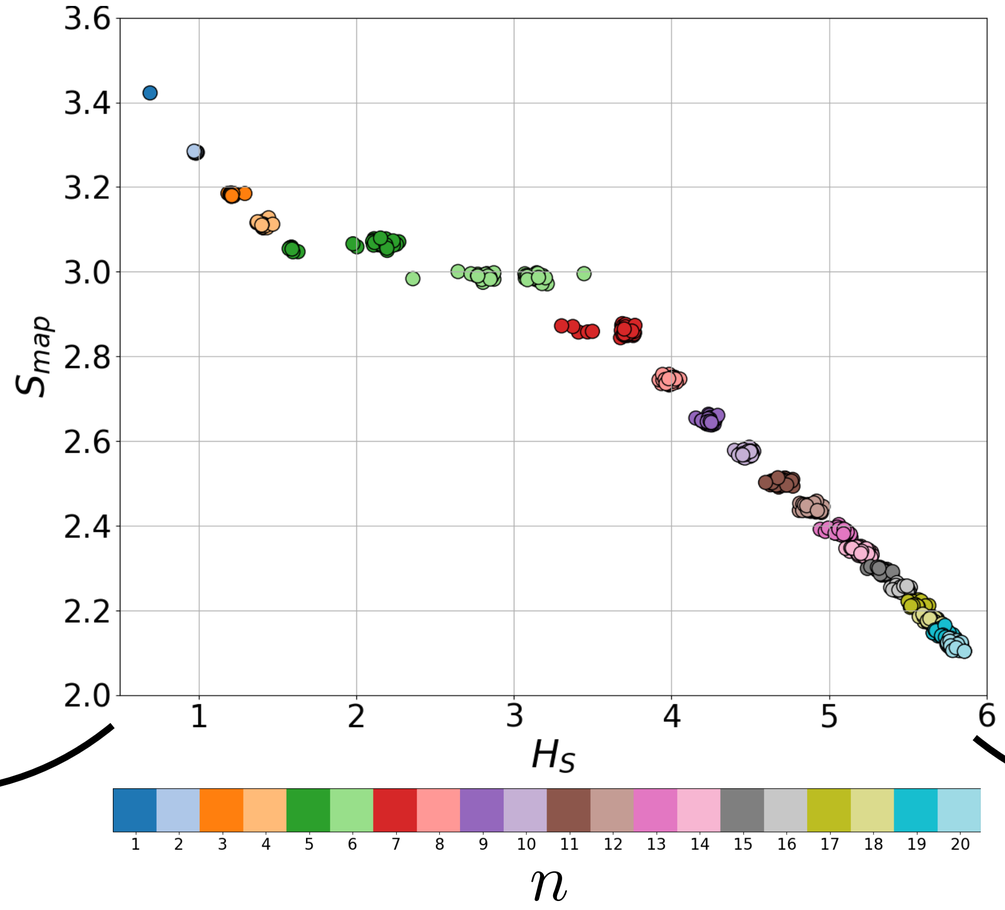


Aldrigo, Menichetti, RP, *arXiv* 2024

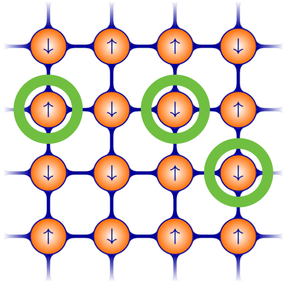
Mapping optimisation of a Hopfield NN

Maximally informative selection of neurons that **minimise the mapping entropy**

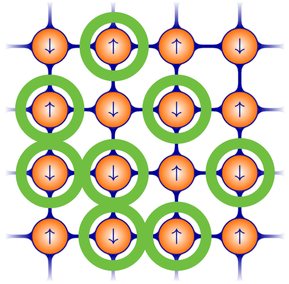
Hopfield model with N=100 neurons and 5 memory patterns



Low resolution



High resolution

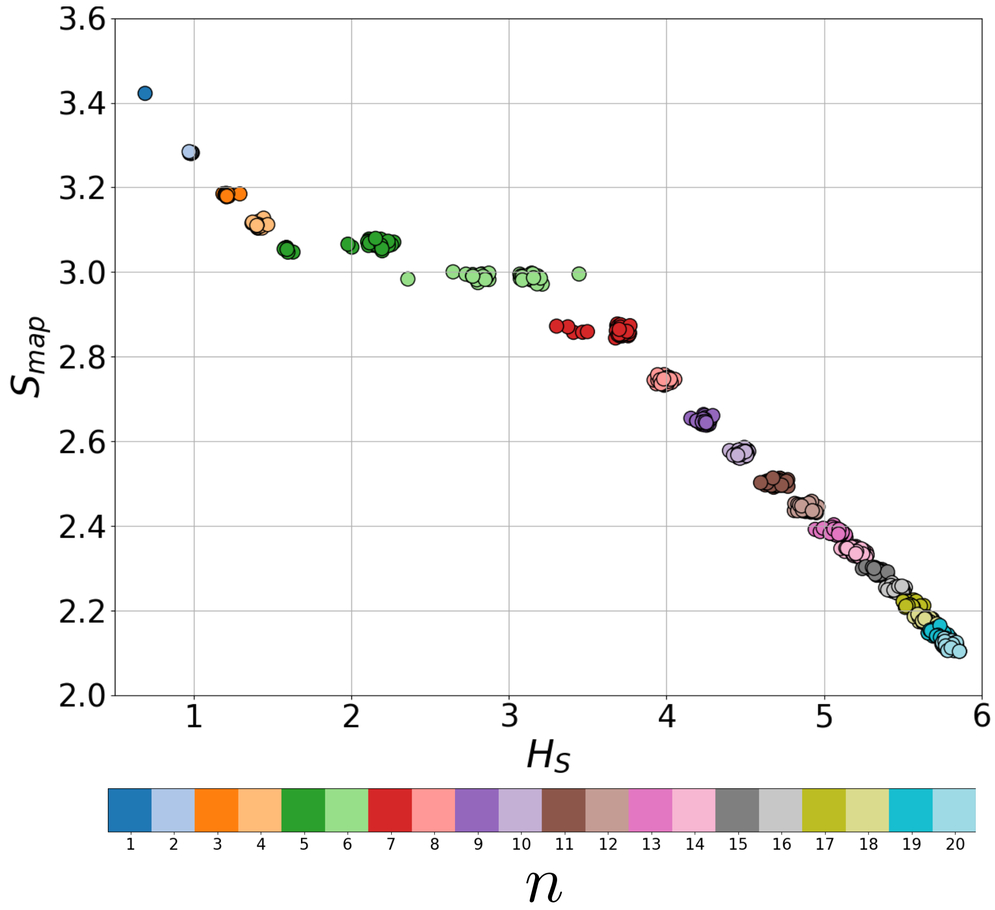


Aldrigo, Menichetti, RP, *arXiv* 2024

Mapping optimisation of a Hopfield NN

Maximally informative selection of neurons that **minimise the mapping entropy**

Hopfield model with N=100 neurons and 5 memory patterns



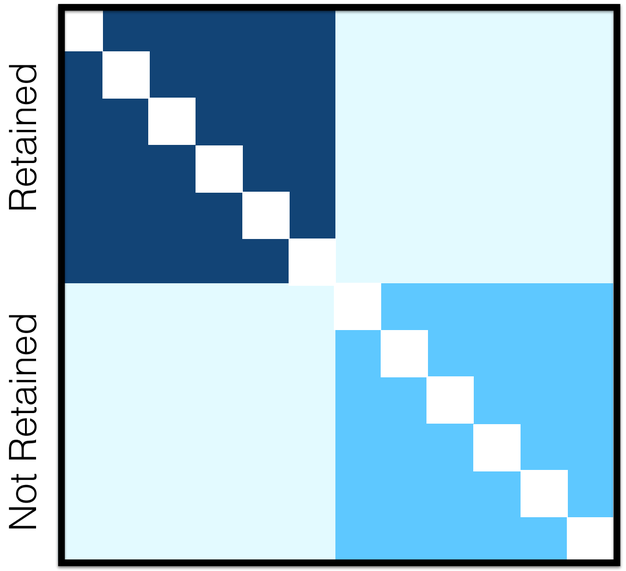
Aldrigo, Menichetti, RP, *arXiv* 2024

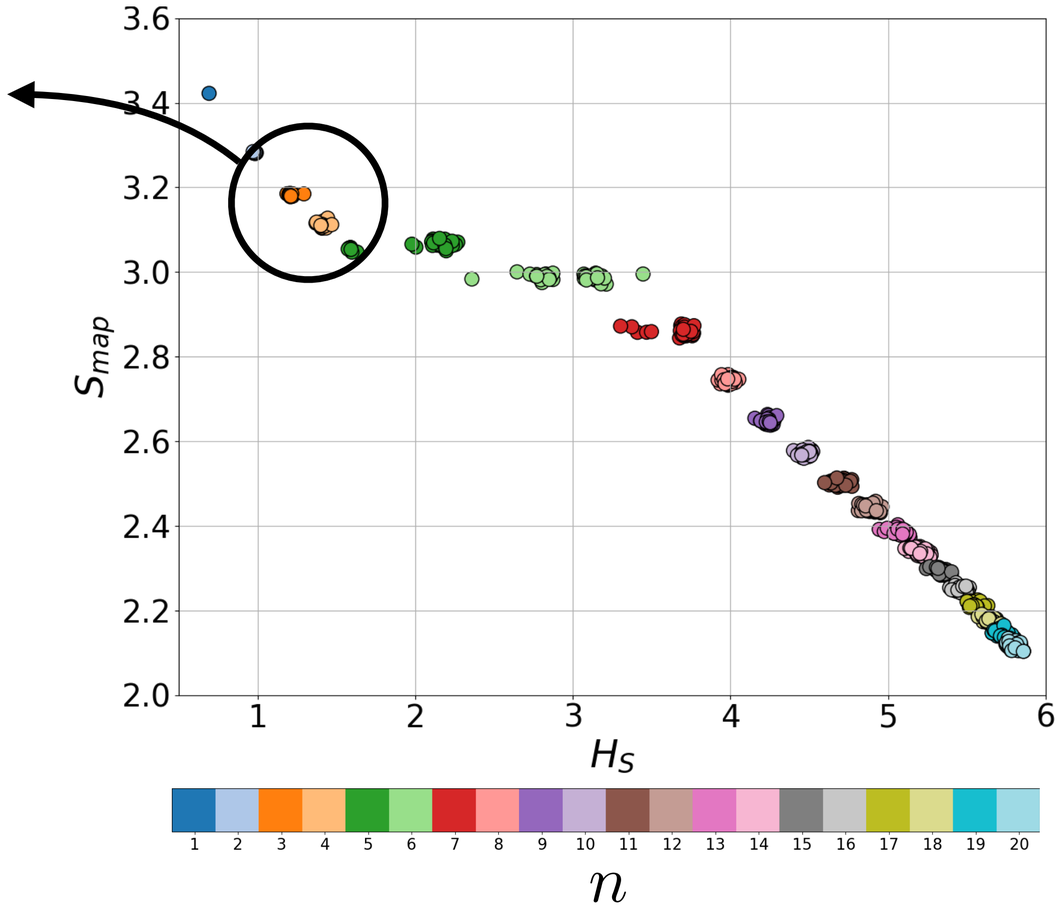
Mapping optimisation of a Hopfield NN

Maximally informative selection of neurons that **minimise the mapping entropy**

Hopfield model with N=100 neurons and 5 memory patterns

Ordered coupling matrix $|J_{ij}|$





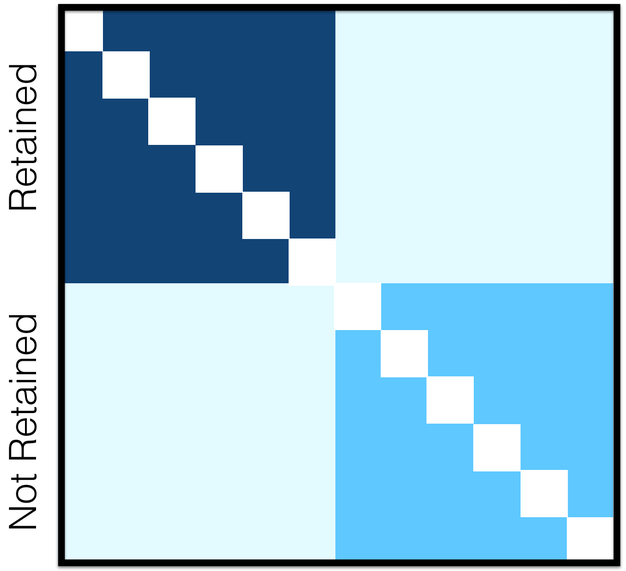
Aldrigo, Menichetti, RP, *arXiv* 2024

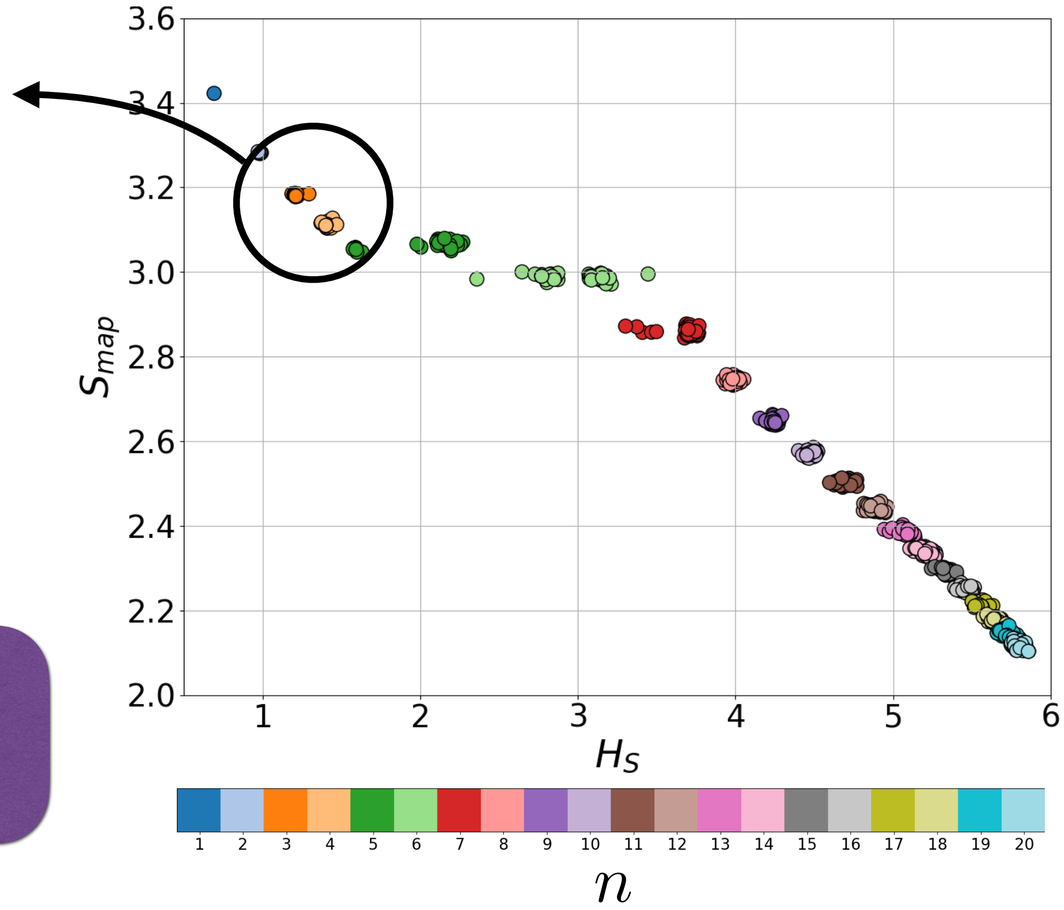
Mapping optimisation of a Hopfield NN

Maximally informative selection of neurons that **minimise the mapping entropy**

Hopfield model with N=100 neurons and 5 memory patterns

Ordered coupling matrix $|J_{ij}|$





Strongly interacting retained neurons, weakly coupled with the integrated ones

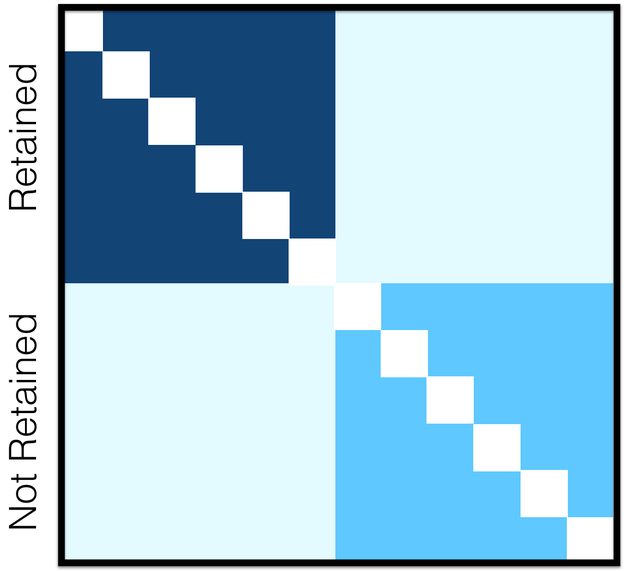
Aldrigo, Menichetti, RP, *arXiv* 2024

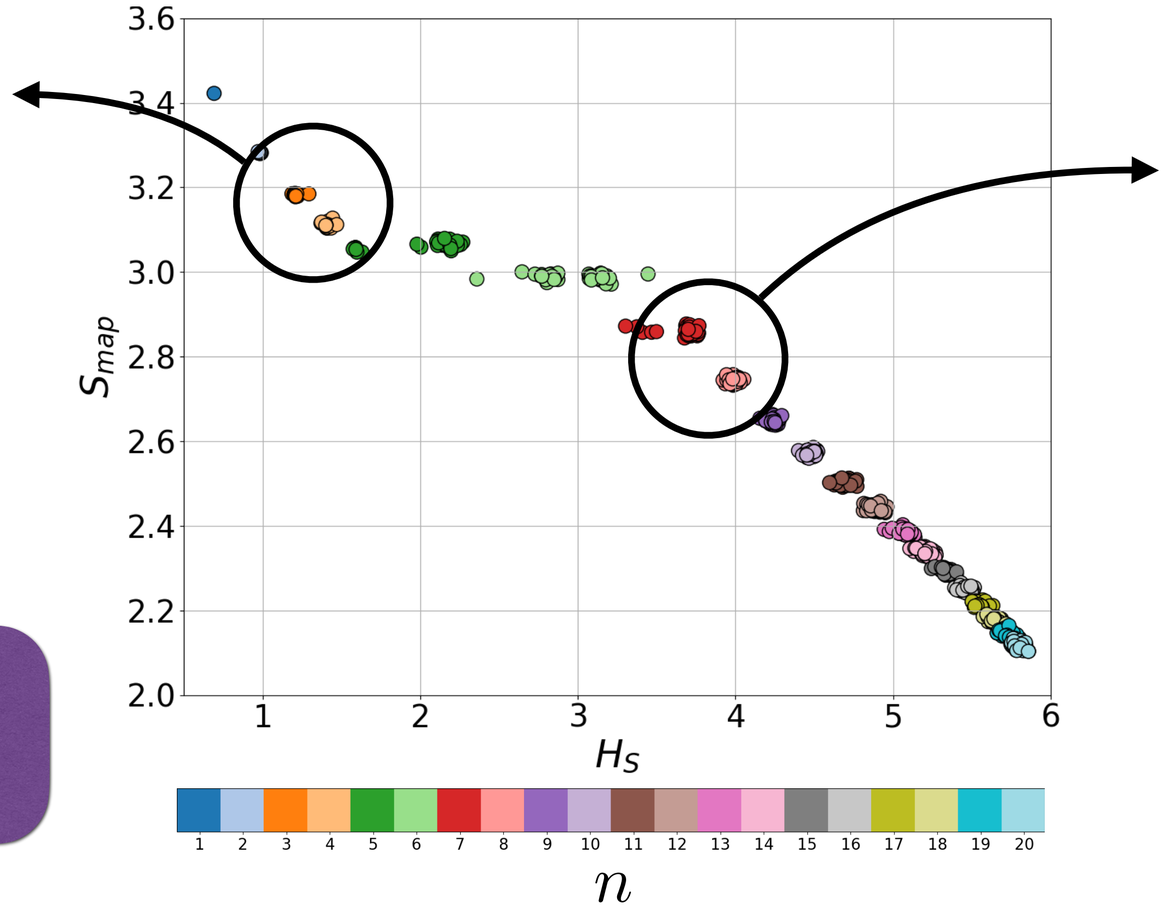
Mapping optimisation of a Hopfield NN

Maximally informative selection of neurons that **minimise the mapping entropy**

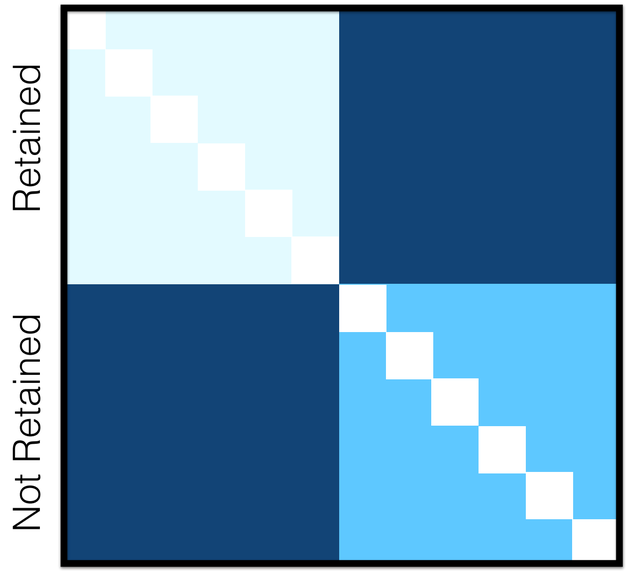
Hopfield model with N=100 neurons and 5 memory patterns

Ordered coupling matrix $|J_{ij}|$





Ordered coupling matrix $|J_{ij}|$



Strongly interacting retained neurons, weakly coupled with the integrated ones

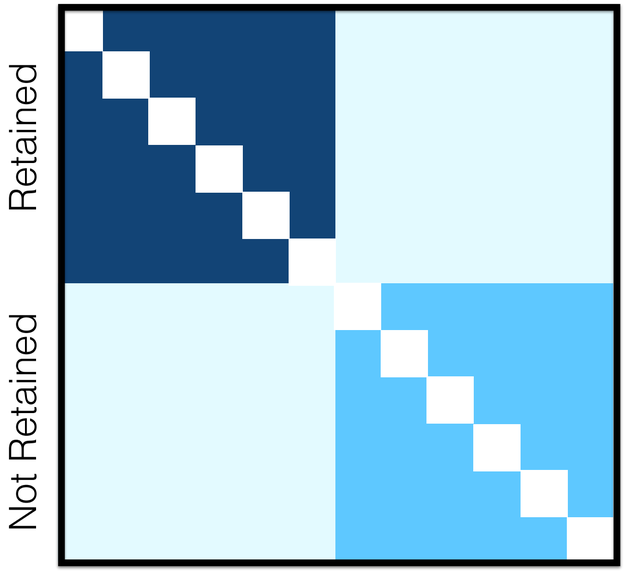
Aldrigo, Menichetti, RP, *arXiv* 2024

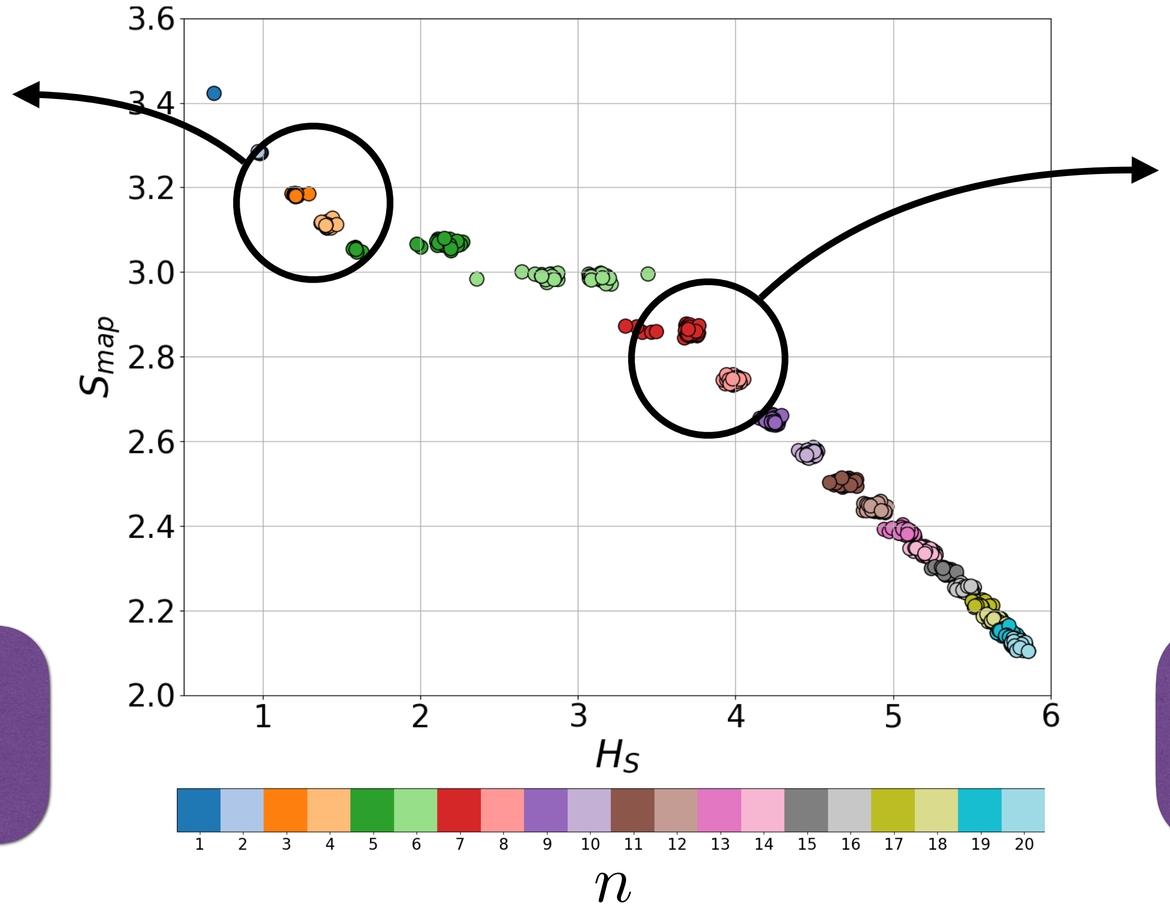
Mapping optimisation of a Hopfield NN

Maximally informative selection of neurons that **minimise the mapping entropy**

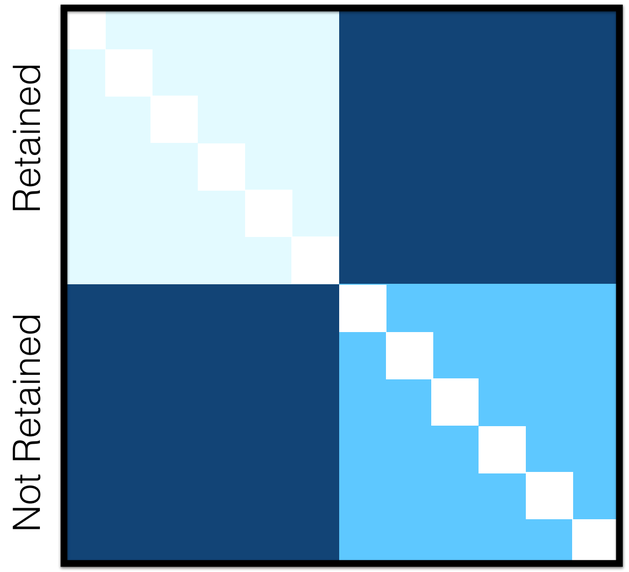
Hopfield model with N=100 neurons and 5 memory patterns

Ordered coupling matrix $|J_{ij}|$





Ordered coupling matrix $|J_{ij}|$



Strongly interacting retained neurons, weakly coupled with the integrated ones

Weakly interacting retained neurons, strongly coupled with the integrated ones

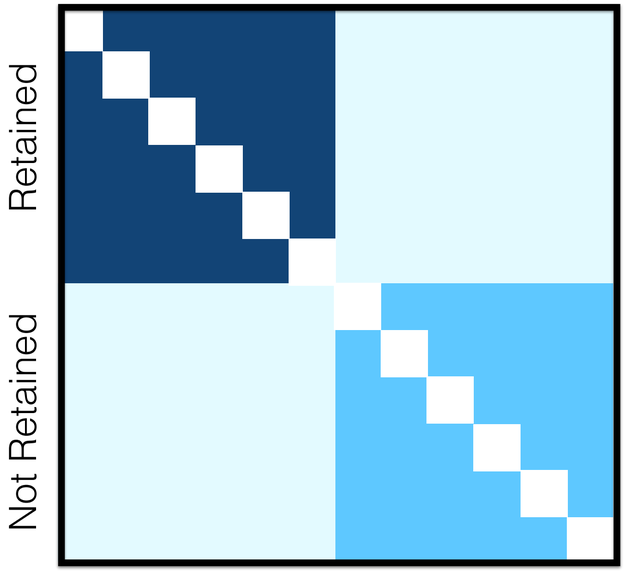
Aldrigo, Menichetti, RP, *arXiv* 2024

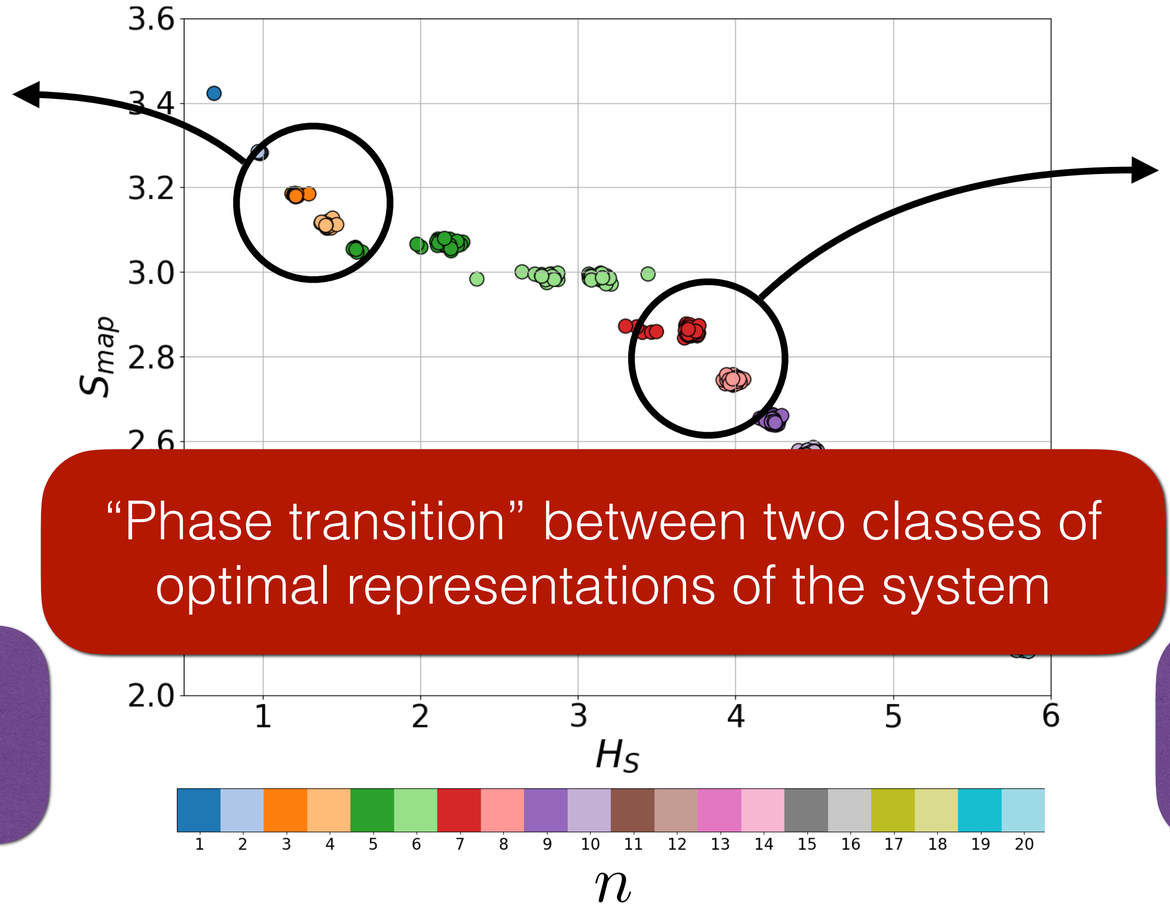
Mapping optimisation of a Hopfield NN

Maximally informative selection of neurons that **minimise the mapping entropy**

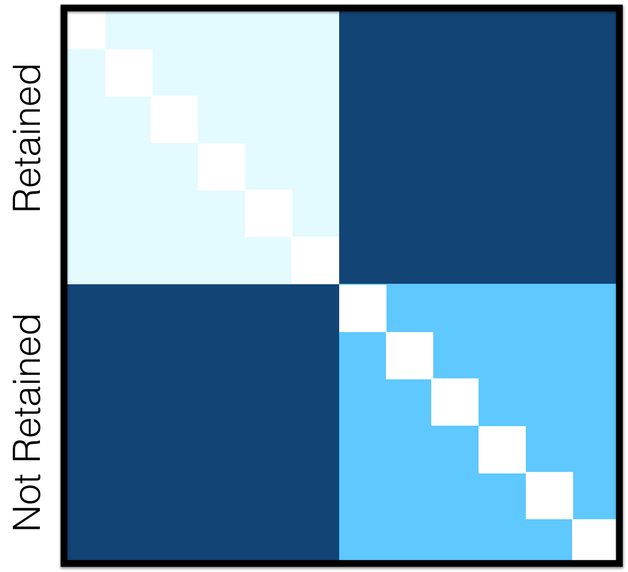
Hopfield model with N=100 neurons and 5 memory patterns

Ordered coupling matrix $|J_{ij}|$





Ordered coupling matrix $|J_{ij}|$



"Phase transition" between two classes of optimal representations of the system

Strongly interacting retained neurons, weakly coupled with the integrated ones

Weakly interacting retained neurons, strongly coupled with the integrated ones

Aldrigo, Menichetti, RP, *arXiv* 2024

Food for thought

Food for thought

Coarse-grained modelling in SM is essentially a one-step RG

Food for thought

Coarse-grained modelling in SM is essentially a one-step RG

Differences: importance of mapping in SM, discrete value of the RG "scale"

Food for thought

Coarse-grained modelling in SM is essentially a one-step RG

Differences: importance of mapping in SM, discrete value of the RG "scale"

ML offers new solutions to old problems in SM and RG,
SM and RG can provide novel tools to understand ML

Food for thought

Coarse-grained modelling in SM is essentially a one-step RG

Differences: importance of mapping in SM, discrete value of the RG "scale"

ML offers new solutions to old problems in SM and RG,
SM and RG can provide novel tools to understand ML

Open, common challenges
- RG / CGing phenomena out of equilibrium
- absence of spectral gaps ("too multi-scale" systems)
- sufficient sampling (SM), sufficient training sets (ML)
- theoretical understanding of NN's far from the Th. limit

More on the interactions in CG models

Get the interaction I: Relative entropy

Goal: CG and AA distribution minimise KL divergence

$$S_{rel} \equiv D_{KL}(p_r(\mathbf{r})||P_r(\mathbf{r}|U)) = \int d\mathbf{r}\ p_r(\mathbf{r}) \ln\left(\frac{p_r(\mathbf{r})}{P_r(\mathbf{r}|U)}\right)$$

$$P_r(\mathbf{r}|U) = P_R(\mathbf{M}(\mathbf{r})|U)\frac{w(\mathbf{r})}{\Omega(\mathbf{M}(\mathbf{r}))}$$

Prob. that the CG model samples an AA configuration

$$\Omega(\mathbf{R}) \equiv \int d\mathbf{r}\ w(\mathbf{r})\delta(\mathbf{M}(\mathbf{r}) - \mathbf{R})$$

Weighted # of AA states that map on CG state R

$$w(\mathbf{r}) = V^{-n}\exp(-\beta u(\mathbf{r}))$$

Chosen weight is Boltzmann

$$P_r(\mathbf{r}|U) = P_R(\mathbf{M}(\mathbf{r})|U)\frac{p_r(\mathbf{r})}{p_R(\mathbf{M}(\mathbf{r}))}$$

Normalised prob. as in the first line

Get the interaction I: Relative entropy

Relative entropy (general expression)

$$
\begin{aligned}
S_{rel} &= \int d\mathbf{r}\; p_r(\mathbf{r}) \ln\left(V^{n-N} \frac{p_r(\mathbf{r})}{P_R(\mathbf{M}(\mathbf{r})|U)} \right) - \int d\mathbf{r}\; p_r(\mathbf{r}) \ln\left(V^{n-N} \frac{p_r(\mathbf{r})}{p_R(\mathbf{M}(\mathbf{r}))} \right) \\
&= \int d\mathbf{r}\; p_r(\mathbf{r}) \ln\left(V^{n-N} \frac{p_r(\mathbf{r})}{P_R(\mathbf{M}(\mathbf{r})|U)} \right) - S_{map}
\end{aligned}
$$

Relative entropy with Boltzmann weights

$$
\begin{aligned}
S_{rel} &= \int d\mathbf{R}\; p_R(\mathbf{R}) \ln\left(\frac{p_R(\mathbf{R})}{P_R(\mathbf{R}|U)} \right) \\
&= -\beta \int d\mathbf{R}\; p_R(\mathbf{R}) \left((U^0(\mathbf{R}) - U(\mathbf{R})) - (F[U^0] - F[U]) \right) \\
&= -\beta \langle U^0(\mathbf{R}) - U(\mathbf{R}) \rangle_{AA} + \beta (F[U^0] - F[U]) \geq 0
\end{aligned}
$$

Get the interaction II: Force matching

Formal definition of the problem

Goal: CG forces equal AA forces on average

$$\chi^2[\mathbf{F}] = \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle$$

Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

Formal definition of the problem

Goal: CG forces equal AA forces on average

$$\chi^2[\mathbf{F}] = \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle$$

$$= \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle + \frac{1}{3N}\left\langle \sum_I |\mathbf{F}_I(\mathbf{M}(\mathbf{r})) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle$$

Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

Formal definition of the problem

Goal: CG forces equal AA forces on average

$$
\begin{aligned}
\chi^2[\mathbf{F}] &= \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle \\
&= \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle + \frac{1}{3N}\left\langle \sum_I |\mathbf{F}_I(\mathbf{M}(\mathbf{r})) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle \\
&= \chi^2[\mathbf{F}^0] + \frac{1}{3N}\left\langle \sum_I |\mathbf{F}_I(\mathbf{M}(\mathbf{r})) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle
\end{aligned}
$$

Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

Formal definition of the problem

Goal: CG forces equal AA forces on average

$$
\begin{aligned}
\chi^2[\mathbf{F}] &= \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle \\
&= \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle + \frac{1}{3N}\left\langle \sum_I |\mathbf{F}_I(\mathbf{M}(\mathbf{r})) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle \\
&= \chi^2[\mathbf{F}^0] + \frac{1}{3N}\left\langle \sum_I |\mathbf{F}_I(\mathbf{M}(\mathbf{r})) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle
\end{aligned}
$$

Distance from AA ff to MB ff

Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

Formal definition of the problem

Goal: CG forces equal AA forces on average

$$\chi^2[\mathbf{F}] = \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle$$

$$= \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle + \frac{1}{3N}\left\langle \sum_I |\mathbf{F}_I(\mathbf{M}(\mathbf{r})) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle$$

$$= \chi^2[\mathbf{F}^0] + \frac{1}{3N}\left\langle \sum_I |\mathbf{F}_I(\mathbf{M}(\mathbf{r})) - \mathbf{F}^0_I(\mathbf{M}(\mathbf{r}))|^2 \right\rangle$$

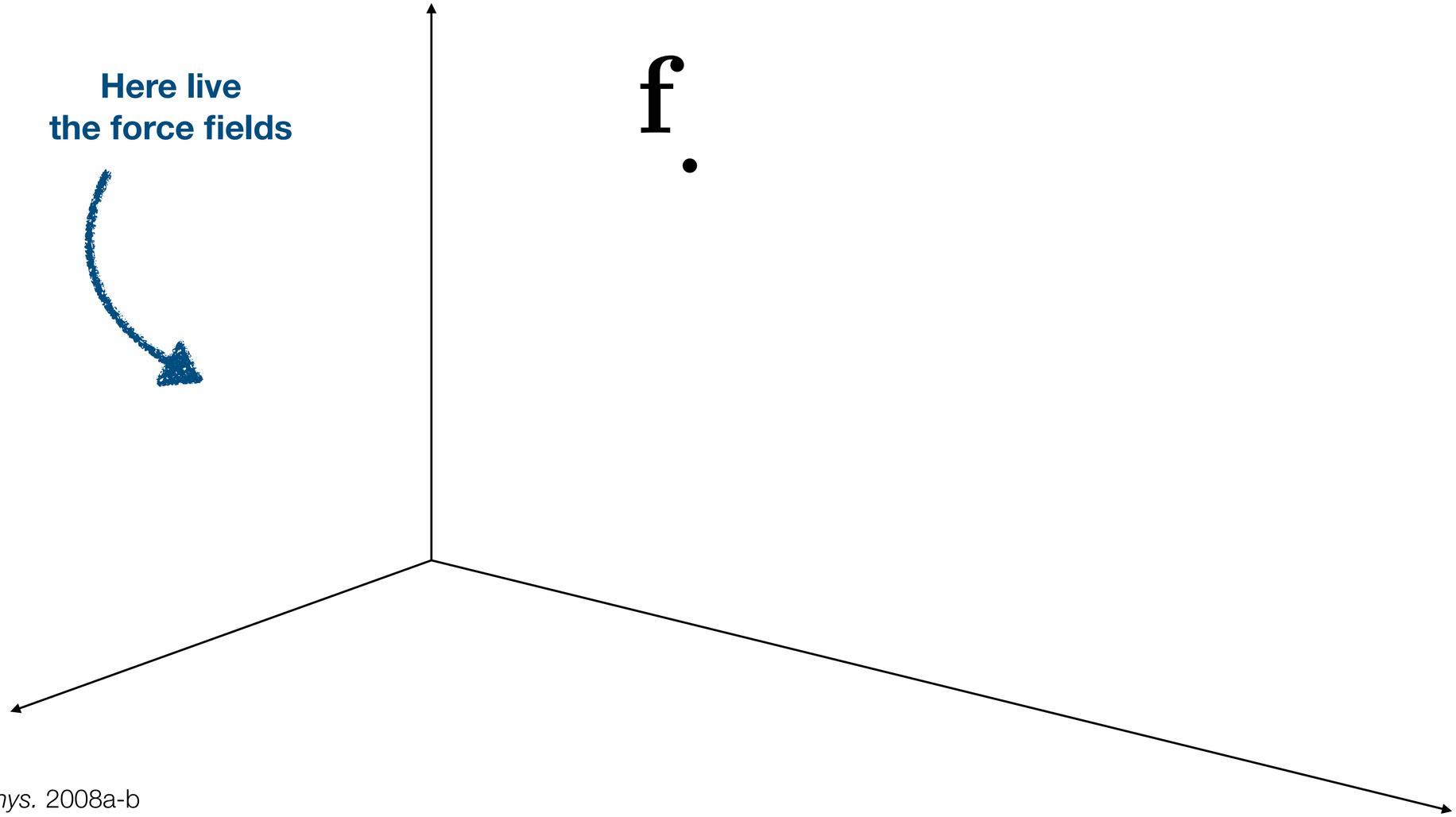
Distance from AA ff to MB ff

Distance from MB ff to CG ff

Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

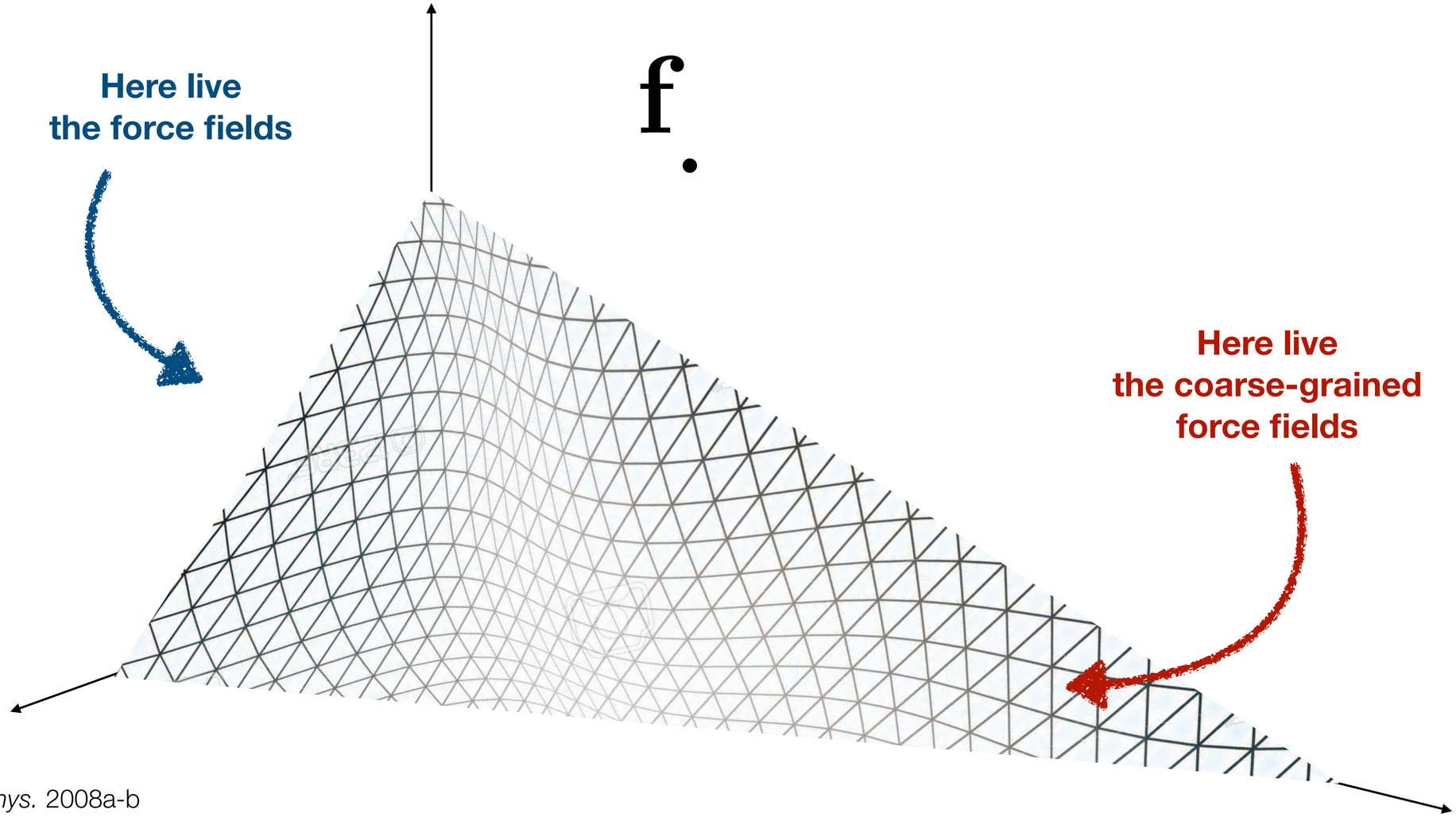
Geometric interpretation



Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

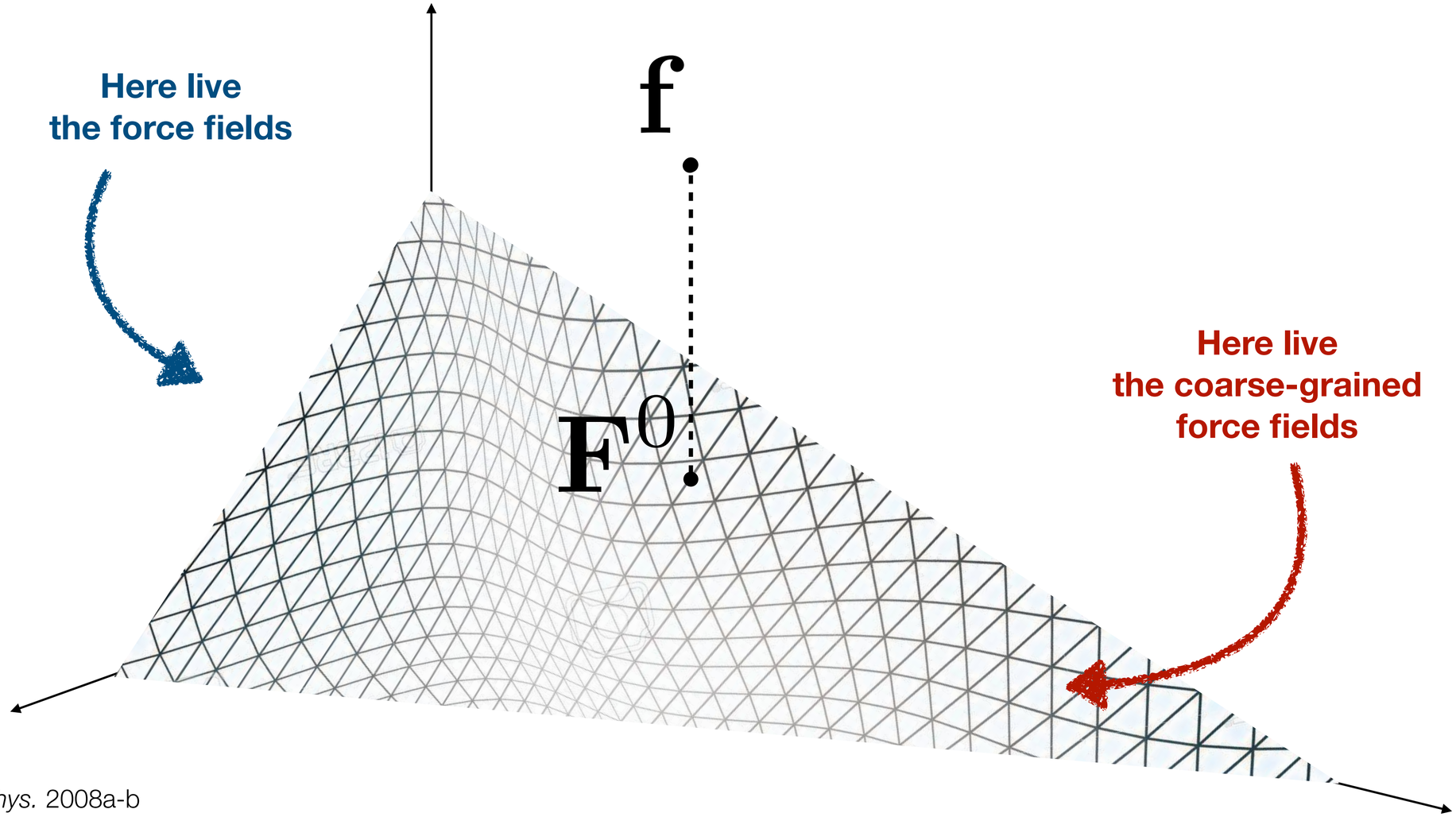
Geometric interpretation



Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

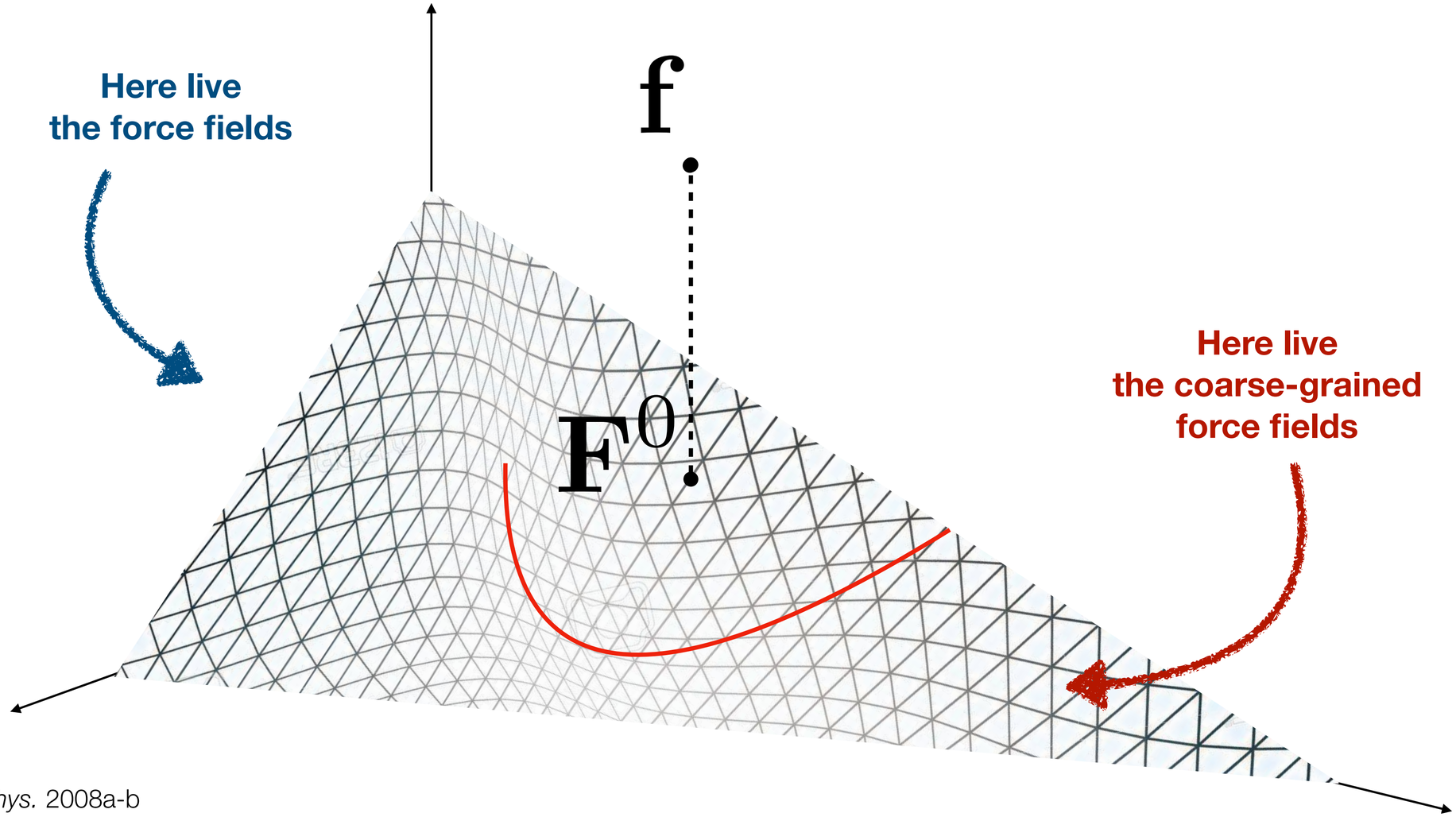
Geometric interpretation



Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

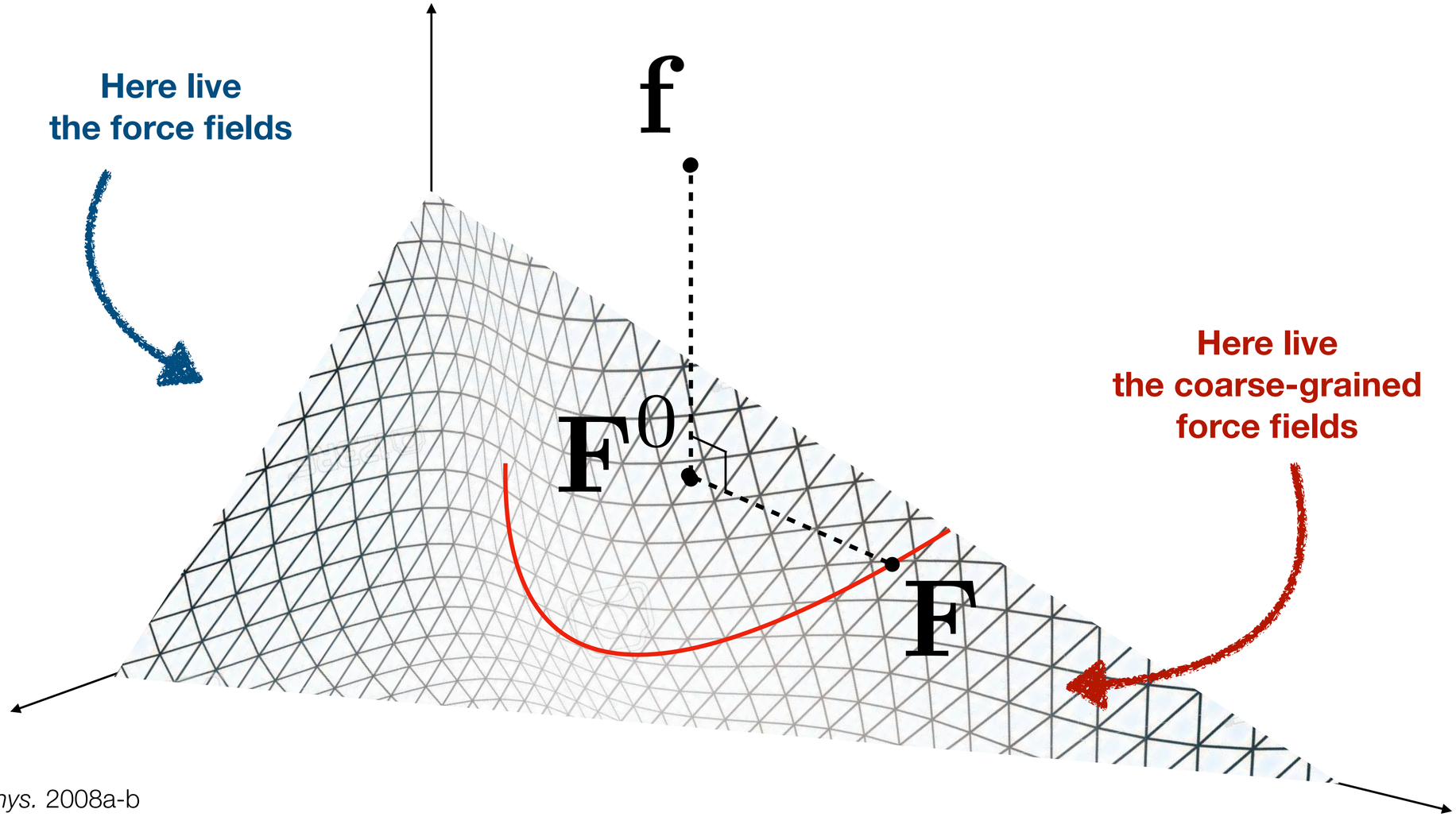
Geometric interpretation



Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

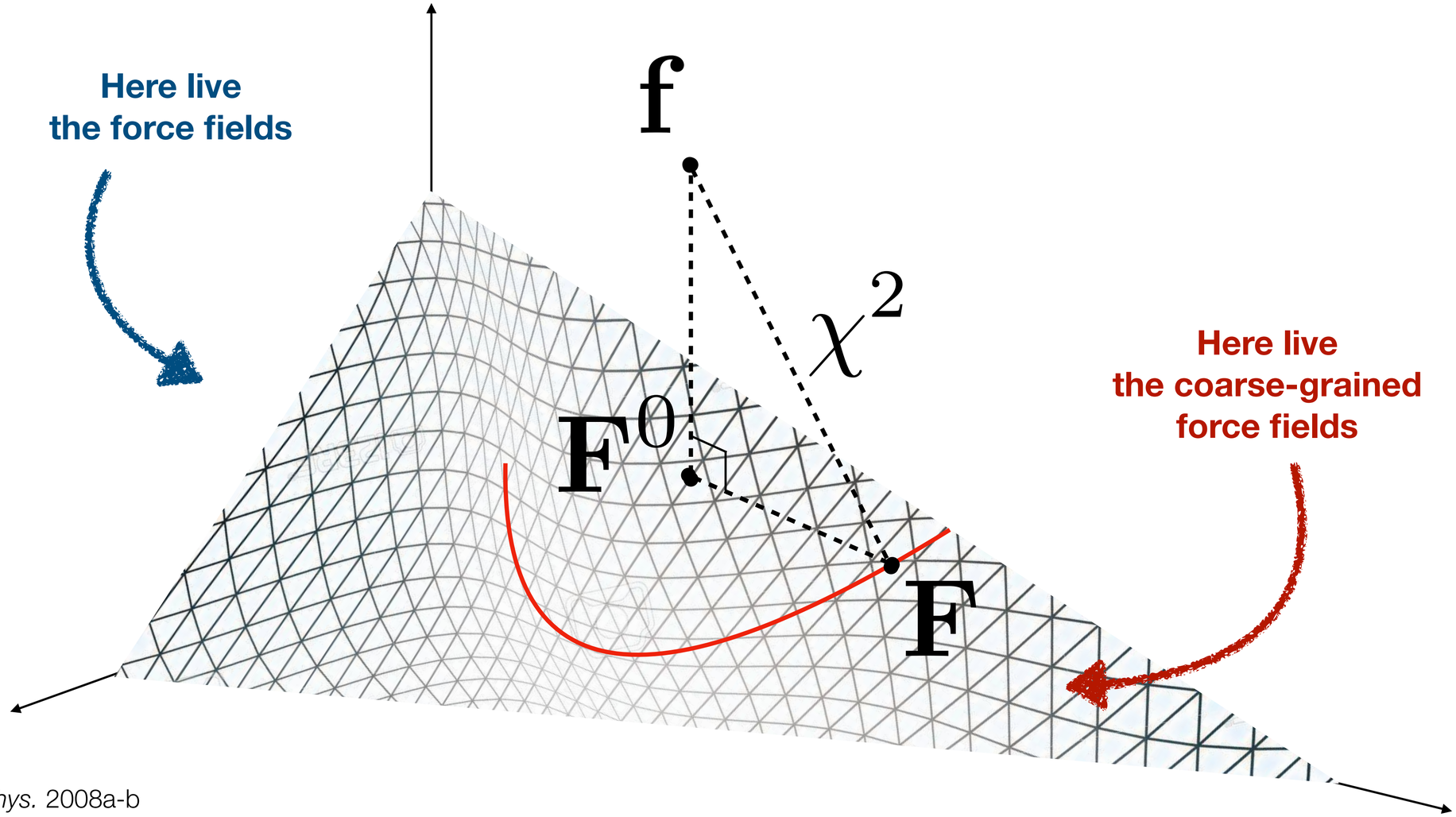
Geometric interpretation



Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

Get the interaction II: Force matching

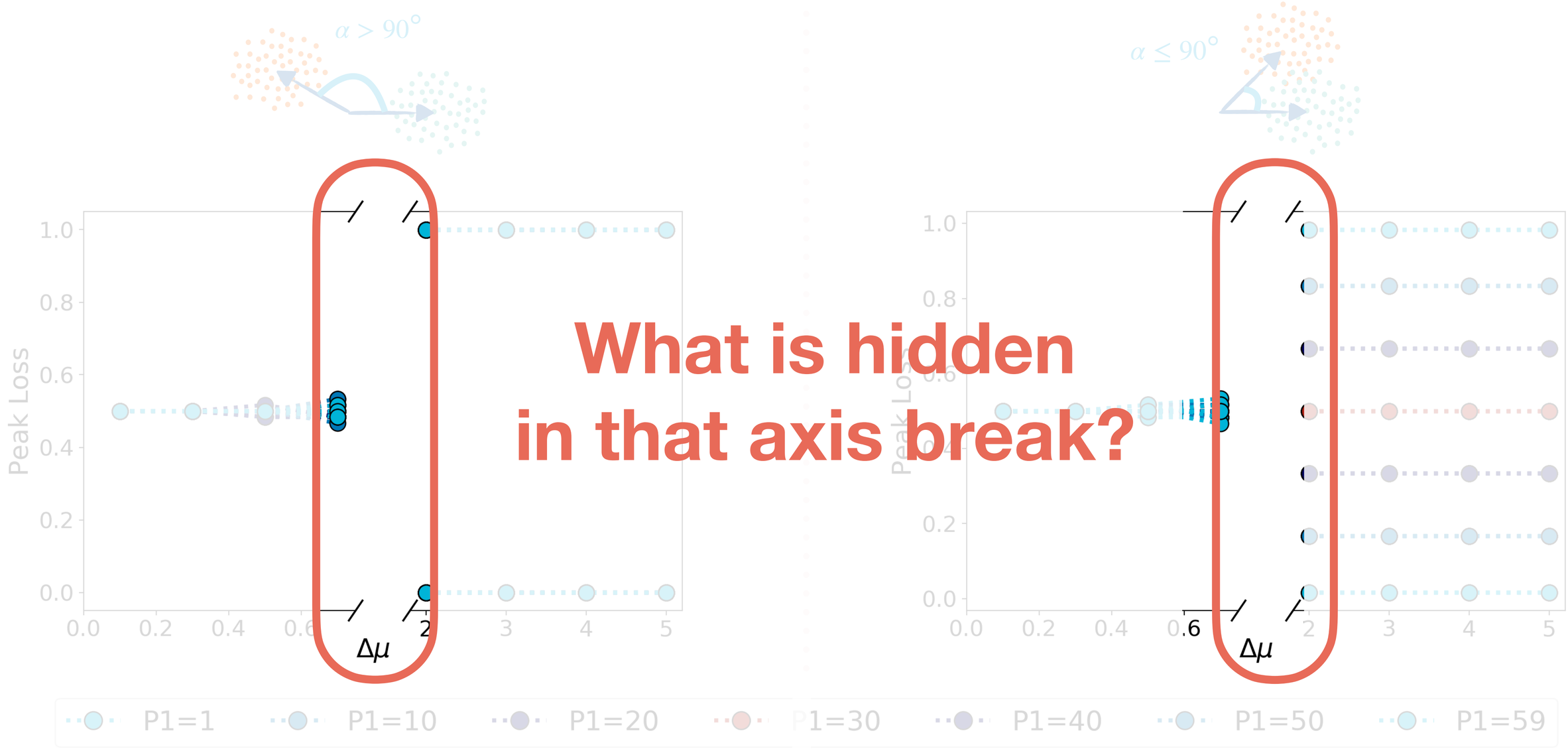
Geometric interpretation



Noid *et al.*, *J. Chem. Phys.* 2008a-b
Noid, *J. Chem. Phys.* 2013
RP, Peter, Kremer, *Entropy* 2014

More on the "transition" in the perceptron

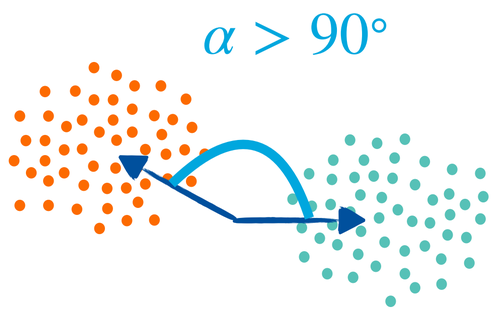
Back to random data

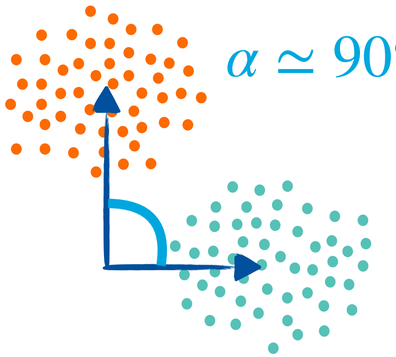


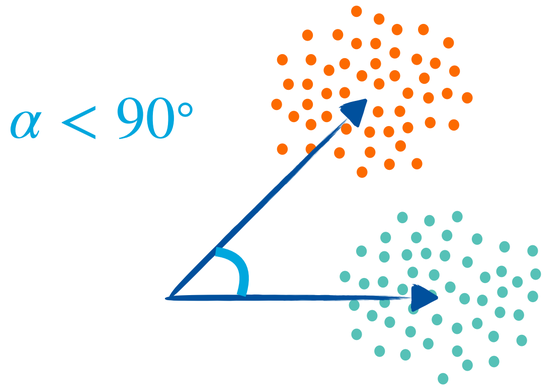
What is hidden in that axis break?

Mele, Ingrosso, Menichetti, RP, *in preparation*

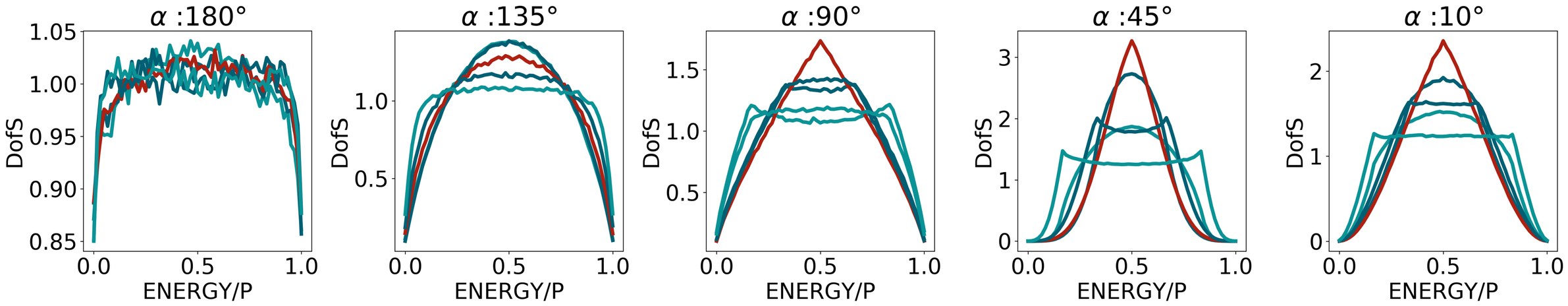
Back to random data





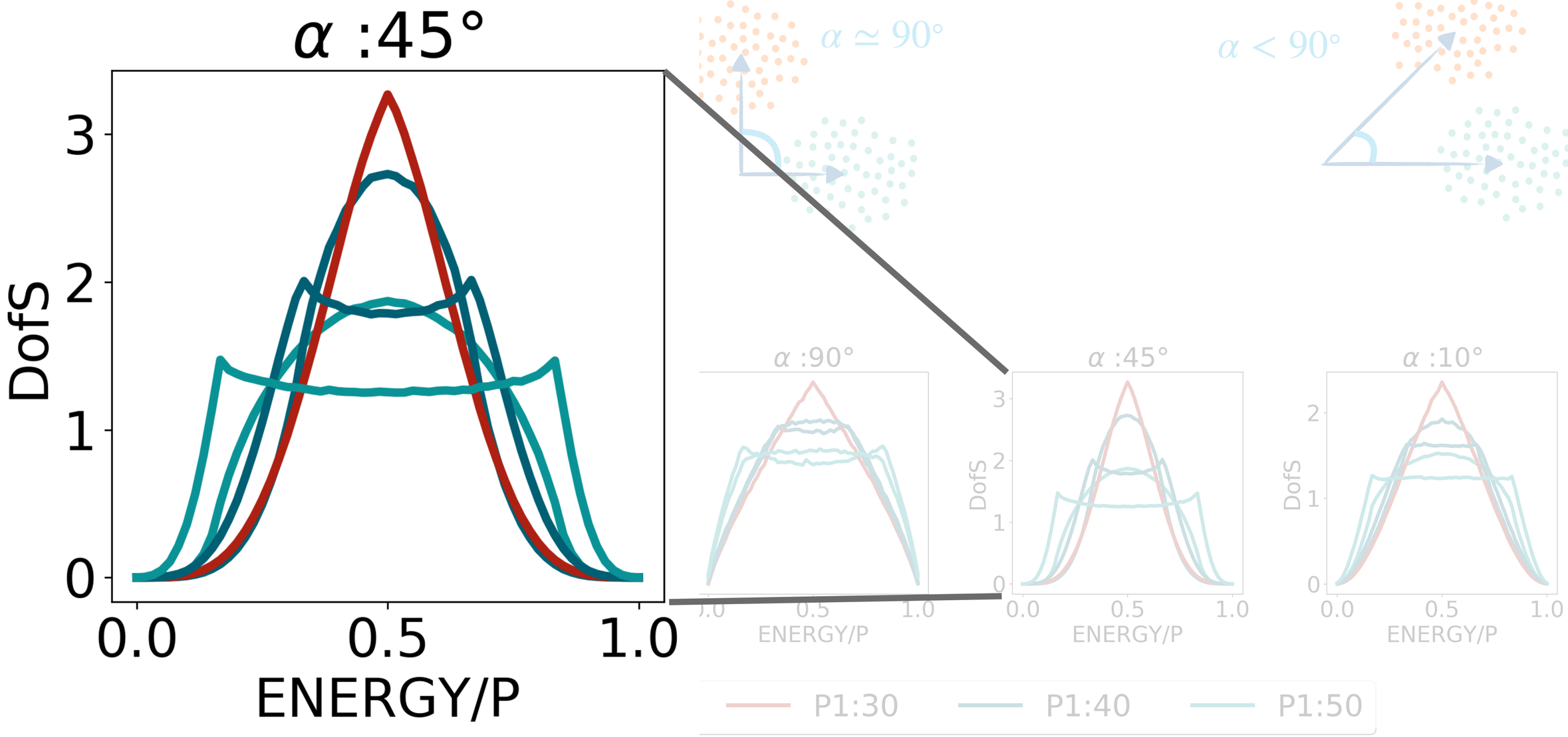


$\Delta\mu \simeq 1$



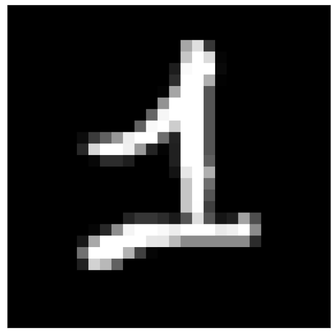
Mele, Ingrosso, Menichetti, RP, *in preparation*

Back to random data



Mele, Ingrosso, Menichetti, RP, *in preparation*

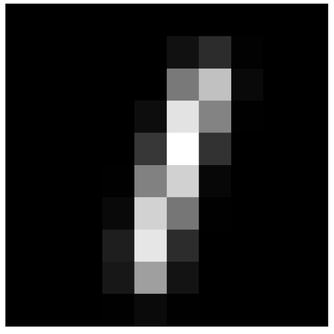
Back again to realistic data



N=784

Reduce Dimensions

N=100

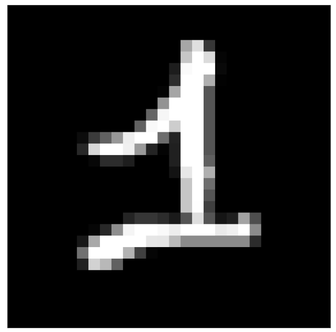


Gaussian Clones

$$N(x,\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

Mele, Ingrosso, Menichetti, RP, *in preparation*

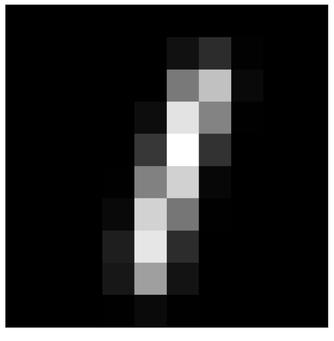
Back again to realistic data



N=784

Reduce Dimensions

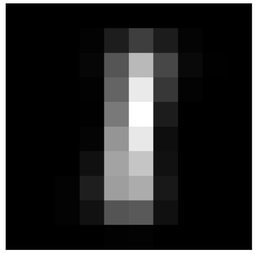
N=100

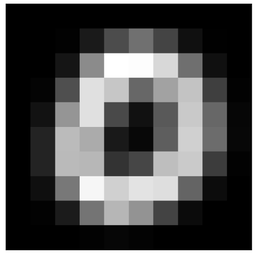


Gaussian Clones

$$N(x,\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

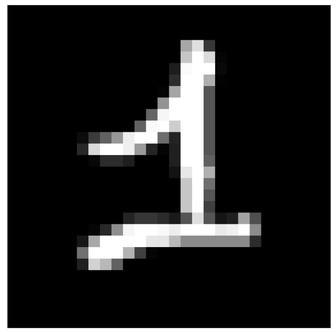
Mean - μ





Mele, Ingrosso, Menichetti, RP, *in preparation*

Back again to realistic data

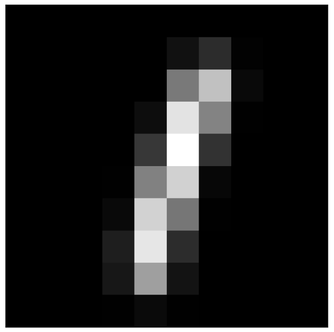


28

N=784

Reduce Dimensions

N=100

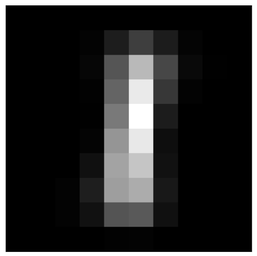


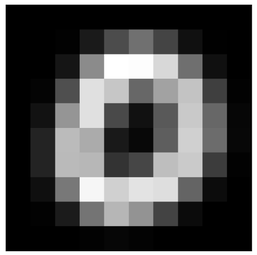
10

Gaussian Clones

$$N(x,\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

Mean - μ



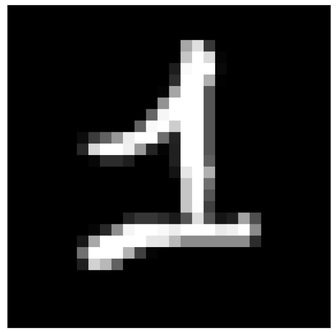


+

Covariance - Σ

Mele, Ingrosso, Menichetti, RP, *in preparation*

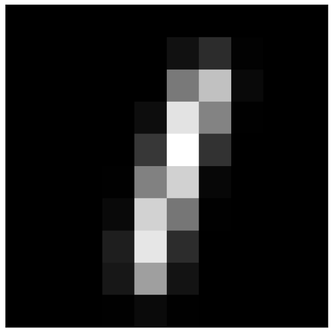
Back again to realistic data



N=784

Reduce Dimensions

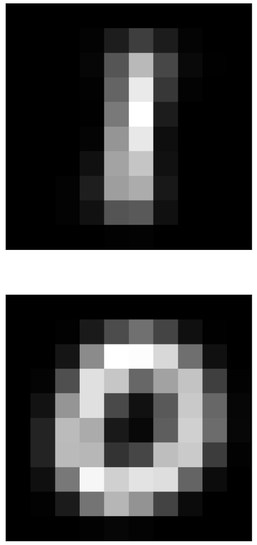
N=100



Gaussian Clones

$$N(x,\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

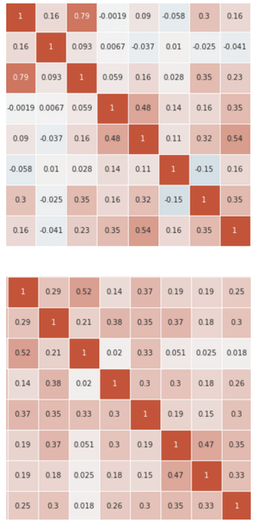
Mean - μ



+

Covariance - Σ

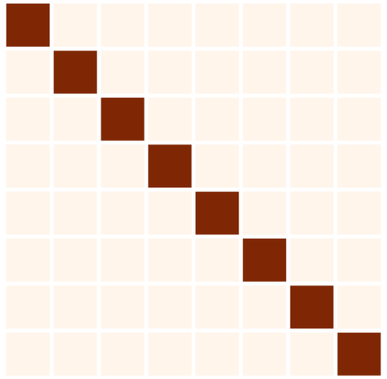
GM



2ISO

diagonal matrix

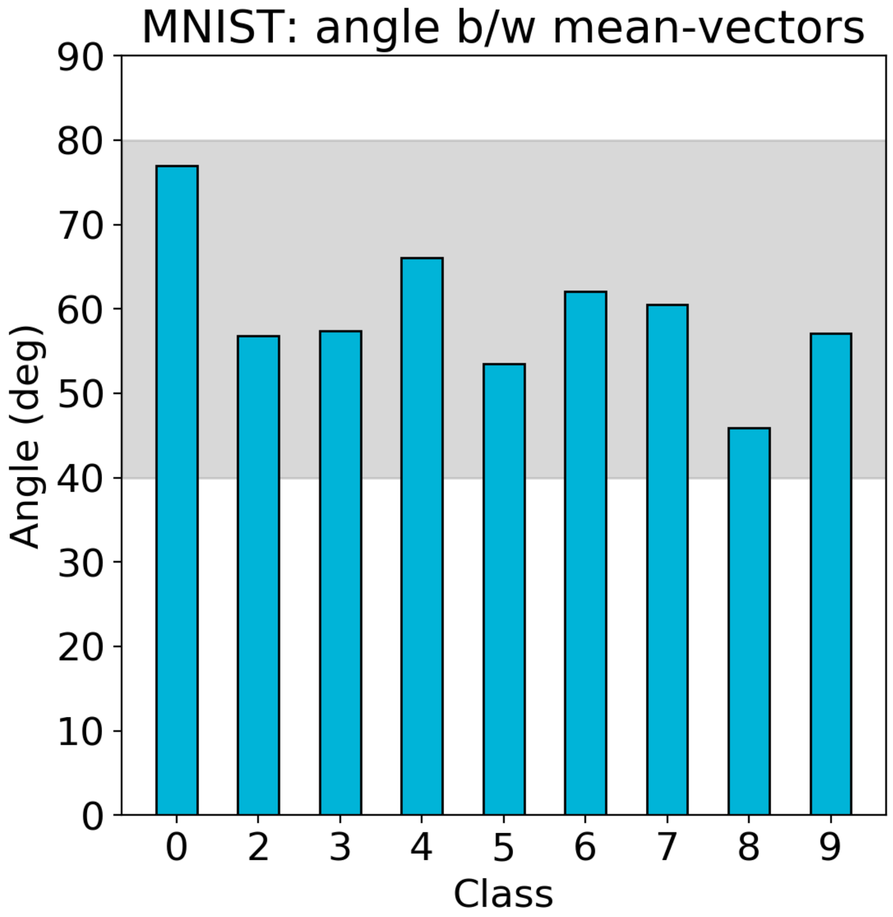
$$v = \sqrt{v_1 \cdot v_2}$$



Mele, Ingrosso, Menichetti, RP, *in preparation*

Back again to realistic data

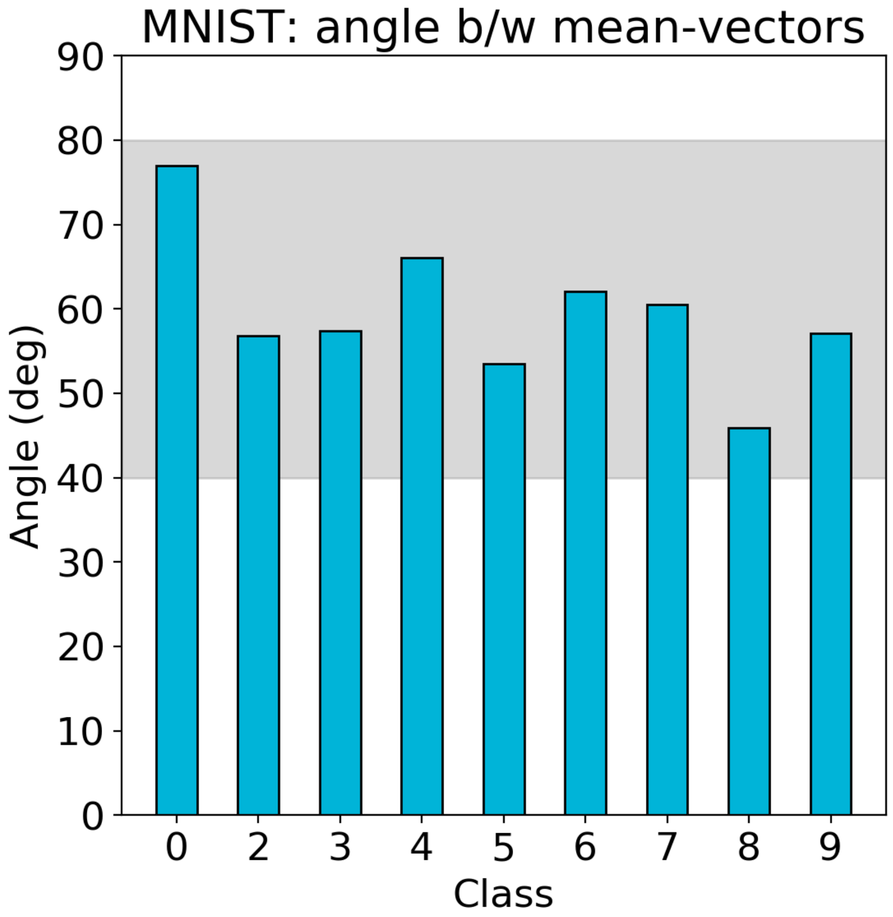
1 vs ✱



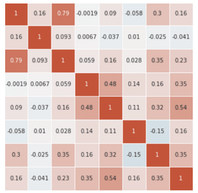
Mele, Ingrosso, Menichetti, RP, *in preparation*

Back again to realistic data

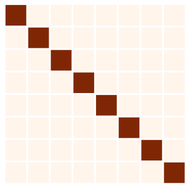
1 vs ∗

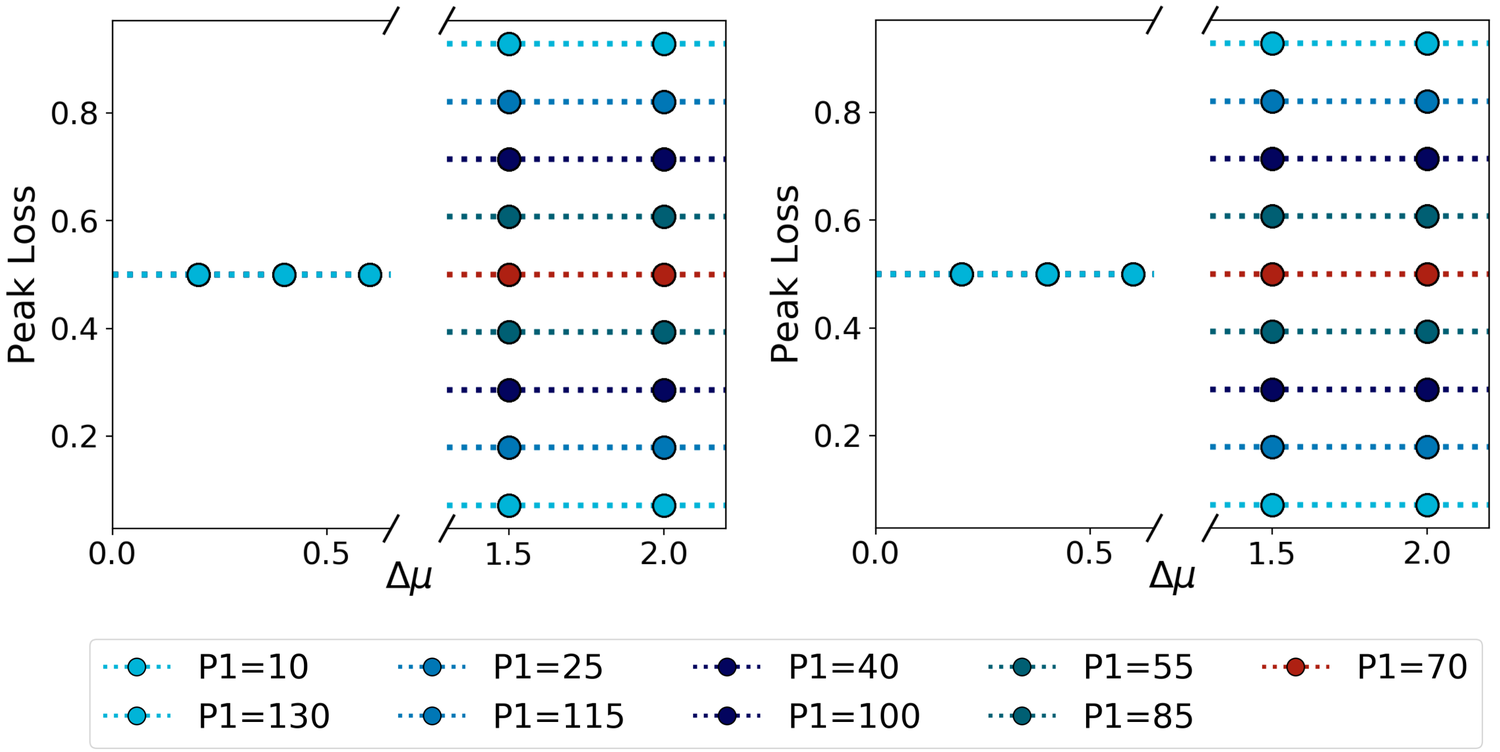


GM



2ISO





Mele, Ingrosso, Menichetti, RP, *in preparation*