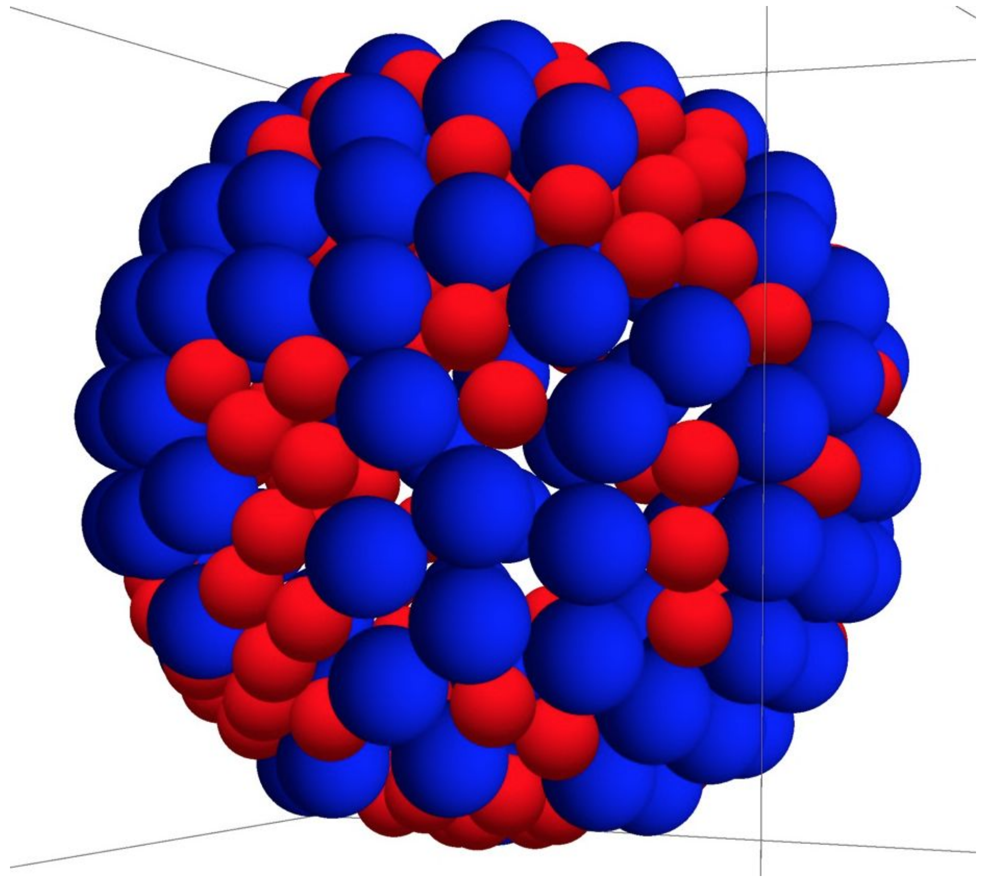
SIMONS FOUNDATION
Cracking the Glass Problem

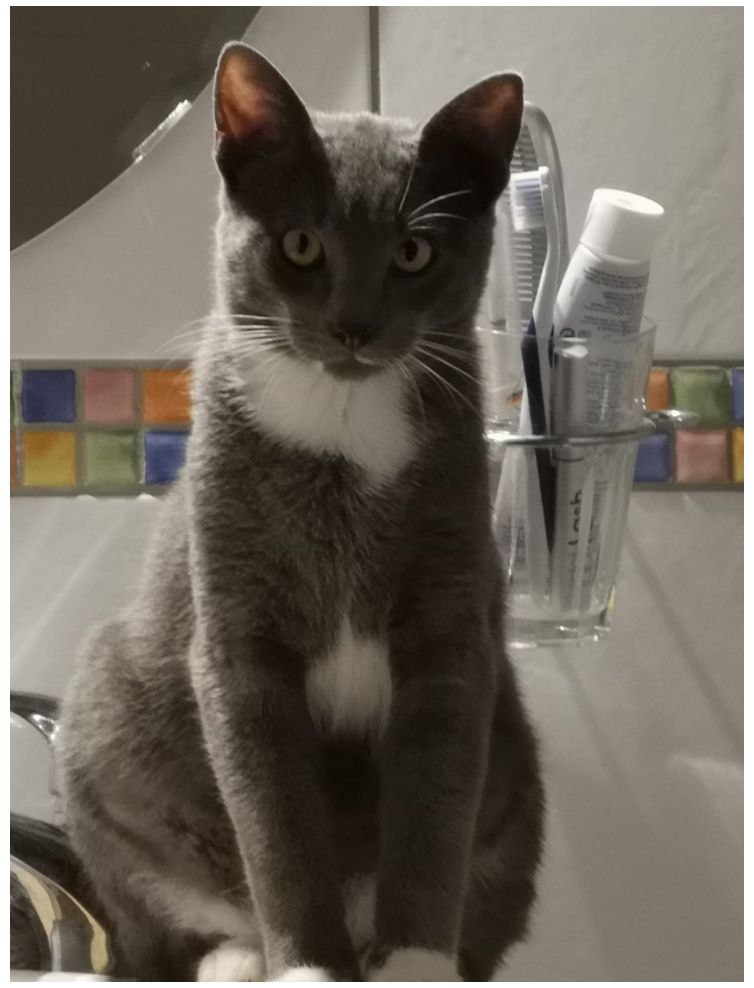
Revisiting the Capacity of Neural Networks

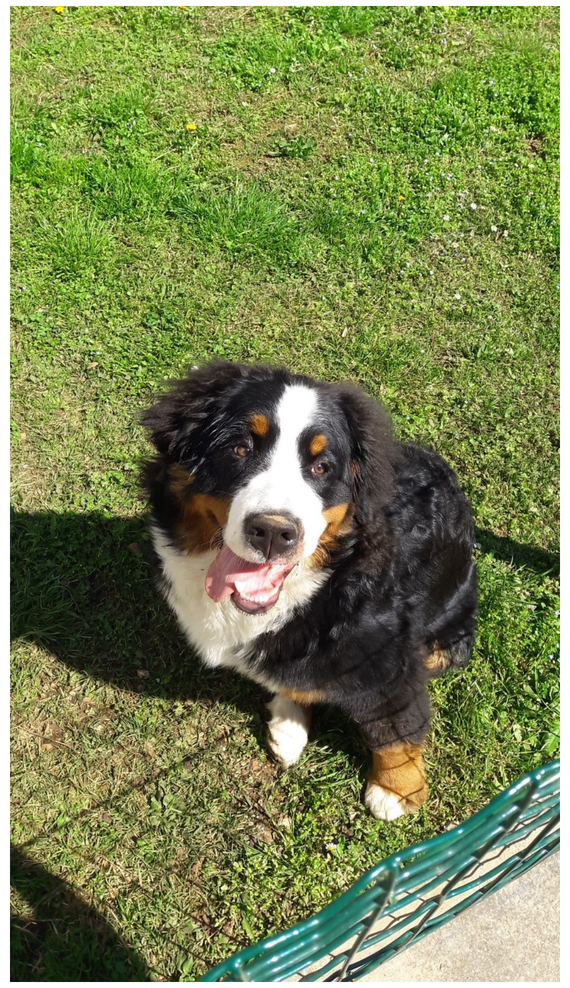
Silvio Franz
LPTMS ORSAY

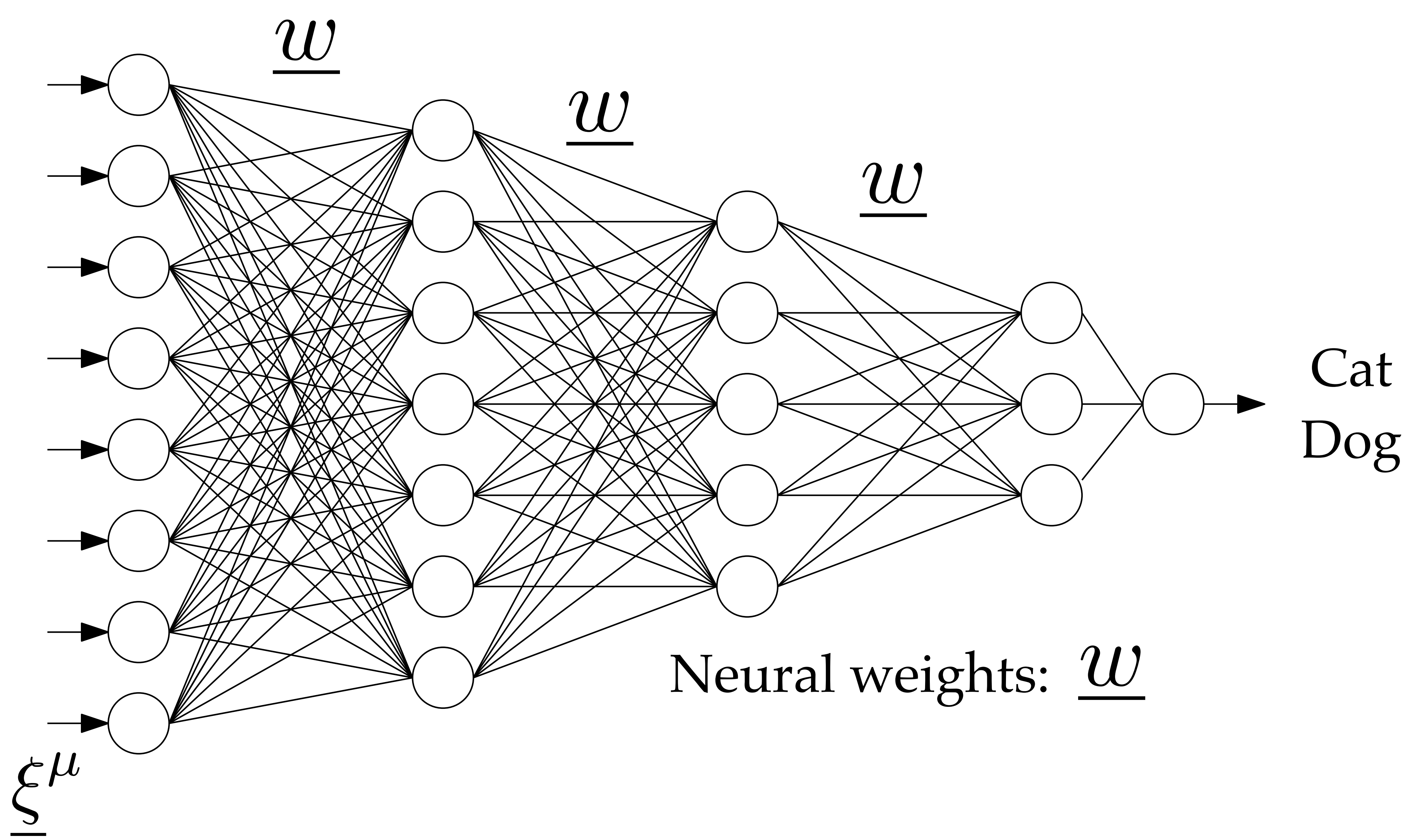


Common work with:
Sungmin Hwang
Giorgio Parisi
Maxim Sevelev
Antonio Sclocchi
Pierfrancesco Urbani
Francesco Zamponi

Supervised learning in Machine Learning







For each picture in the training set we want

$$\theta\left[f(\underline{w}, \underline{\xi}^\mu)\right] = 1 \quad \text{Dog}$$

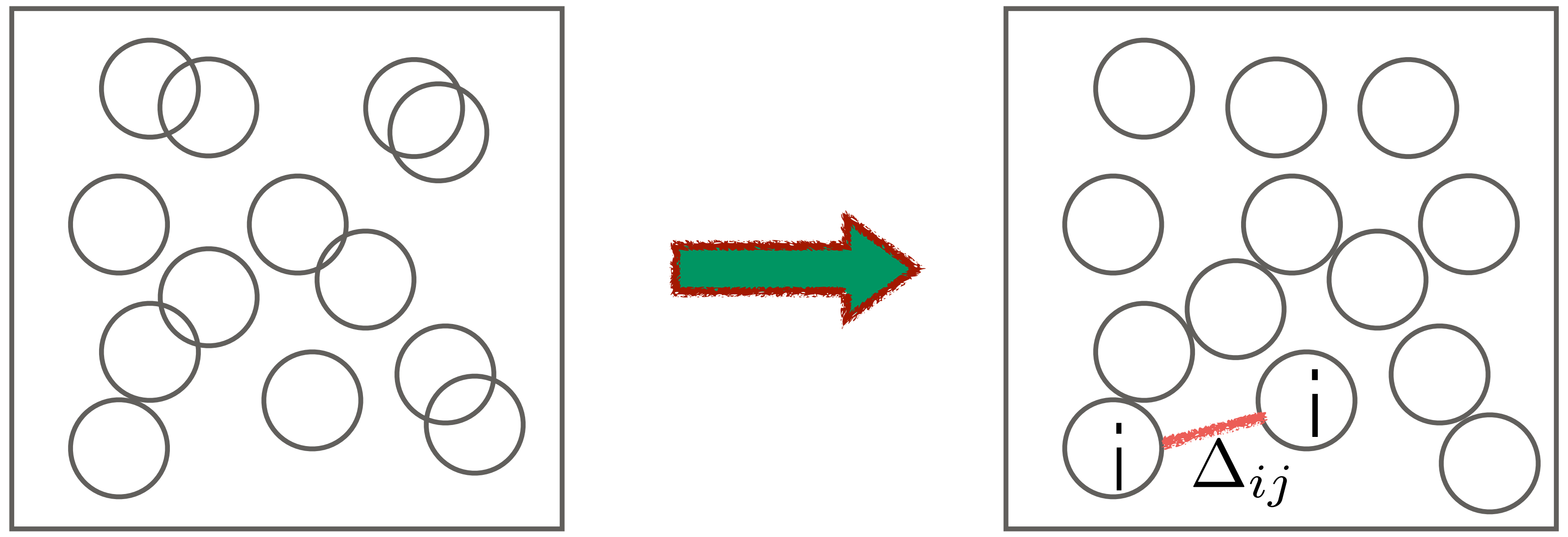
$$\theta\left[f(\underline{w}, \underline{\xi}^\mu)\right] = 0 \quad \text{Cat}$$

Find Weights such that each pattern is correctly classified : Constraint Satisfaction Pb.
Similar to Packing problems

See e.g. Engel Van den Broeck, Statistical Mechanics of Learning

Random Packings of Spheres

- ‘Agnostic’ algorithms
  no order



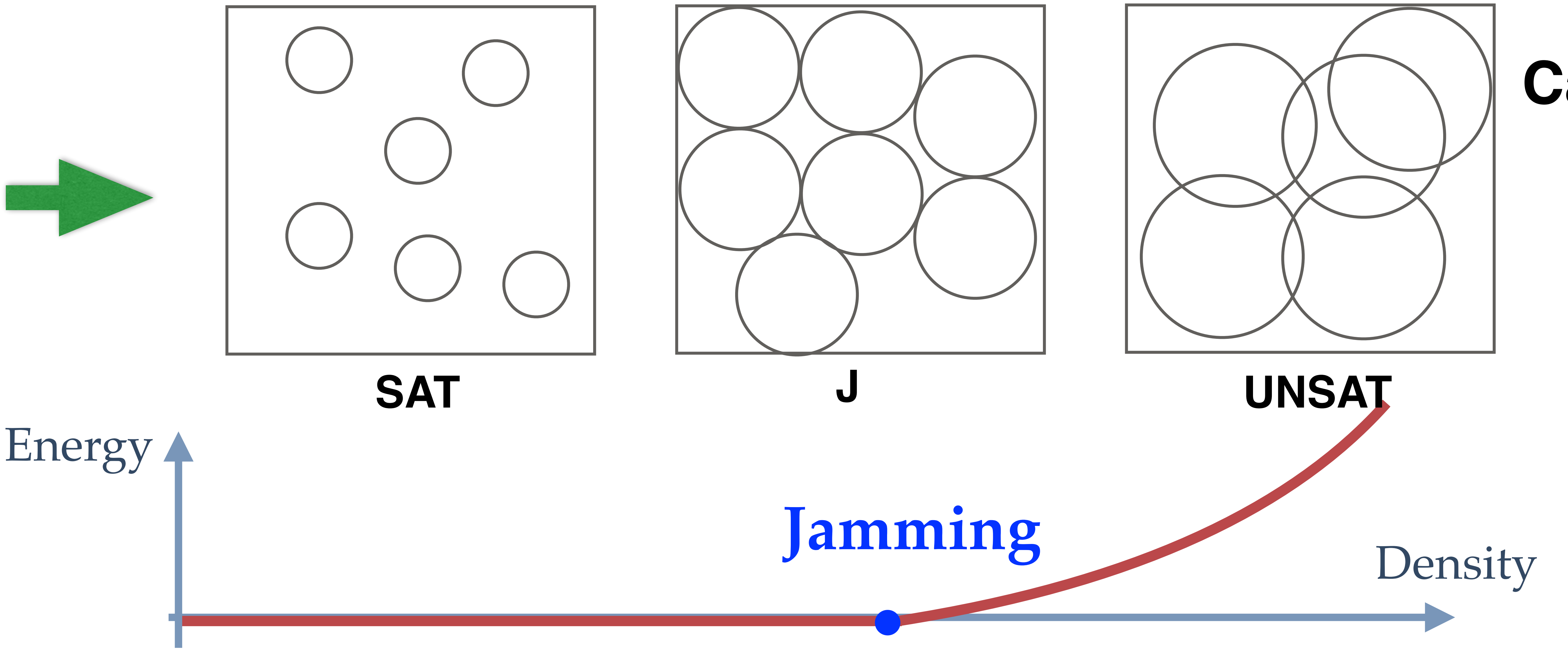
$$\Delta_{ij} = |\mathbf{r}_{ij}| - R_i - R_j$$

‘gap’ variable

Short range repulsion
Minimize H by gradient descent

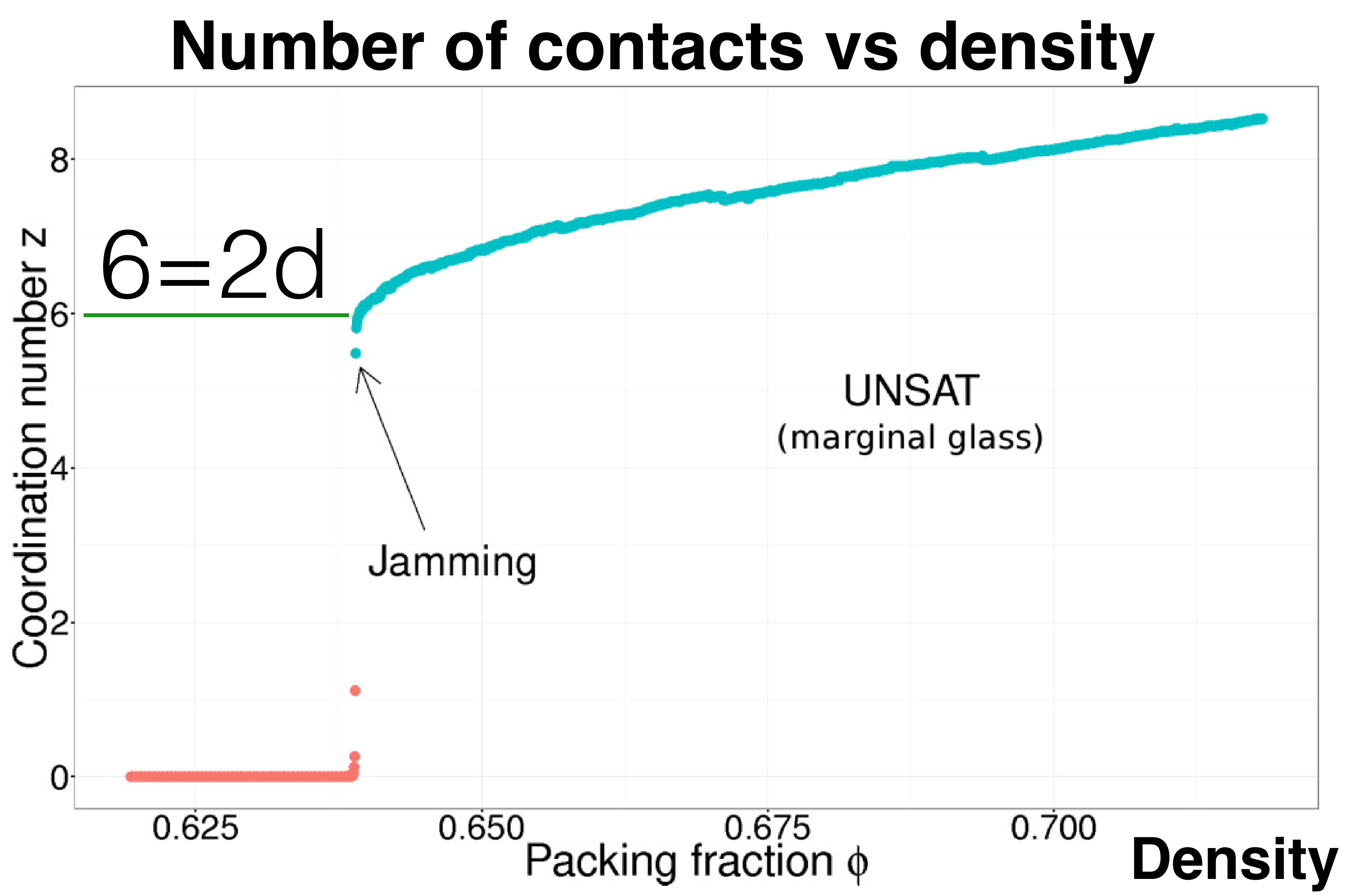
$$H = \frac{1}{2} \sum_{i,j} \theta(-\Delta_{ij}) \Delta_{ij}^2$$

Physical space



Capacity limit in NN ~ Jamming

Nature of the jamming point



Maxwell counting argument

dN_c # of degrees of freedom

$zN_c/2$ # of contact forces

mechanical equilibrium

$$z \geq 2d$$

Jamming is an 'isostatic' marginal point

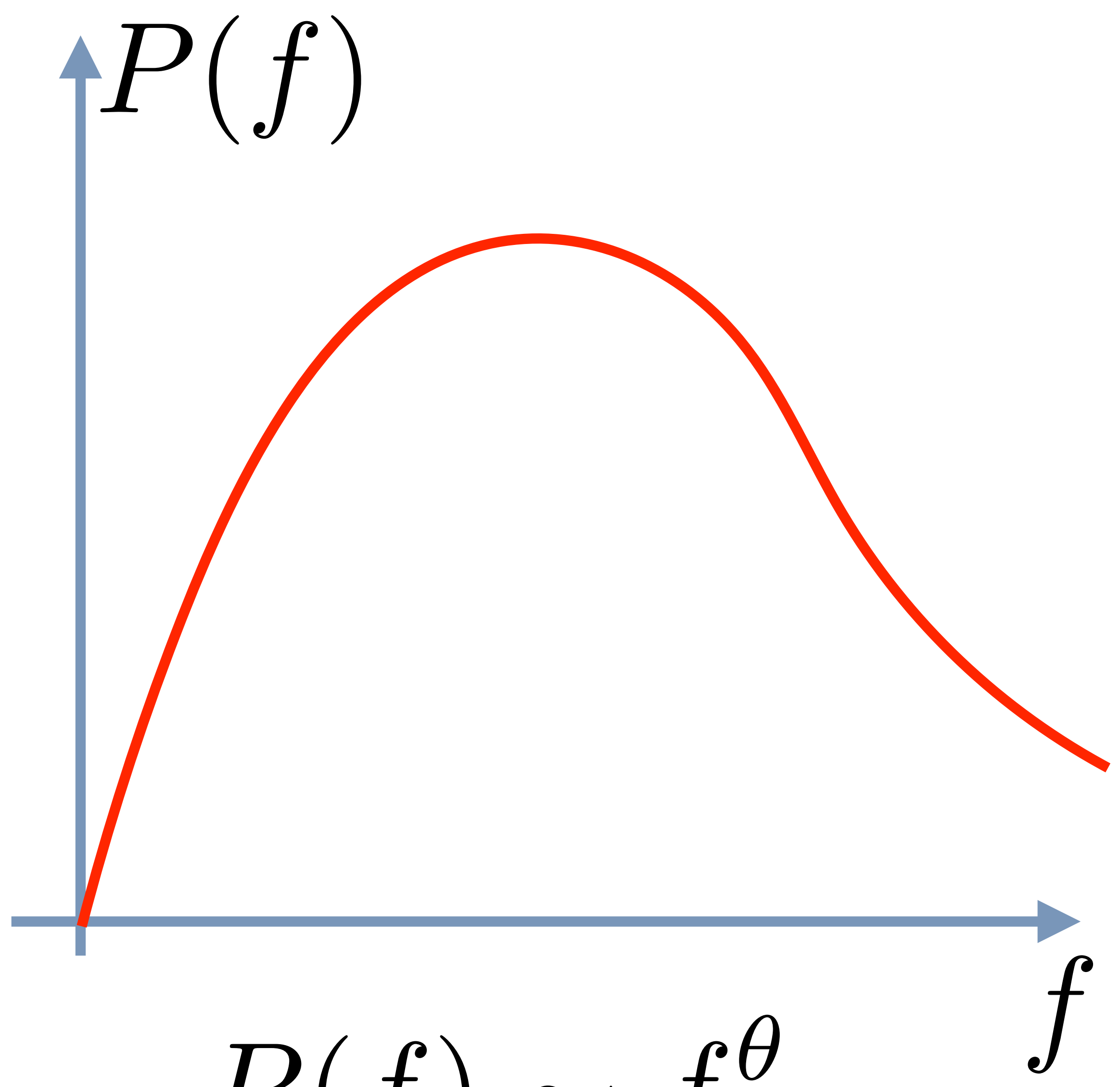
$$z_c = 2d$$

Moukarzel PRL 1988
Tkachenko, Witten PRE 1999
Lubensky et al 2015

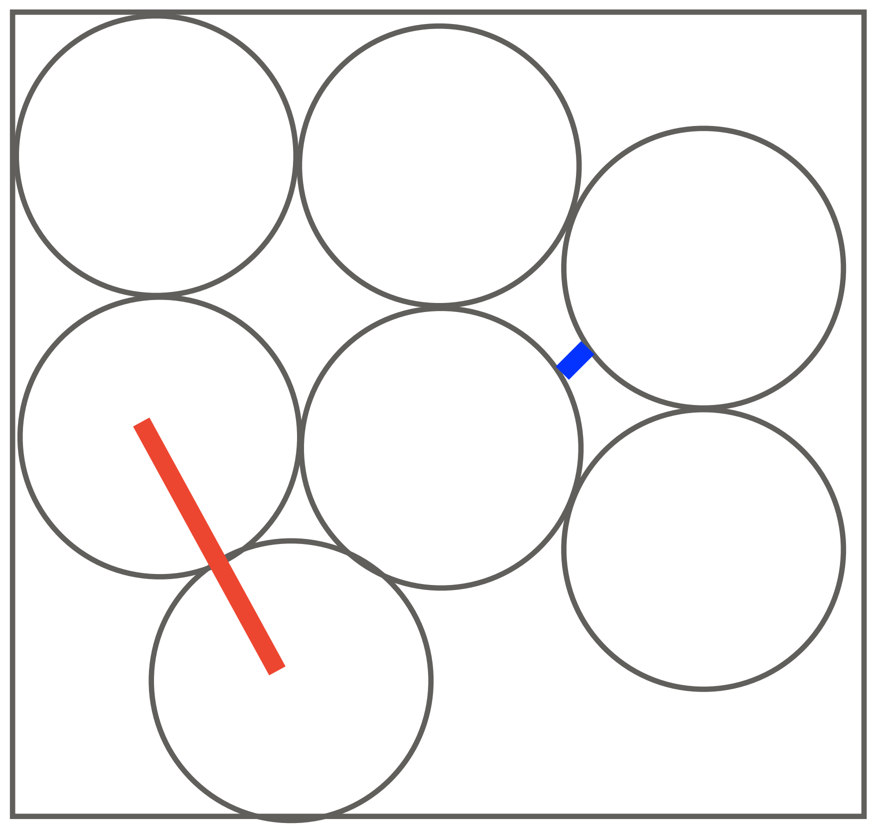
Critical microstructure of packings

Smoking gun of criticality

Force distribution

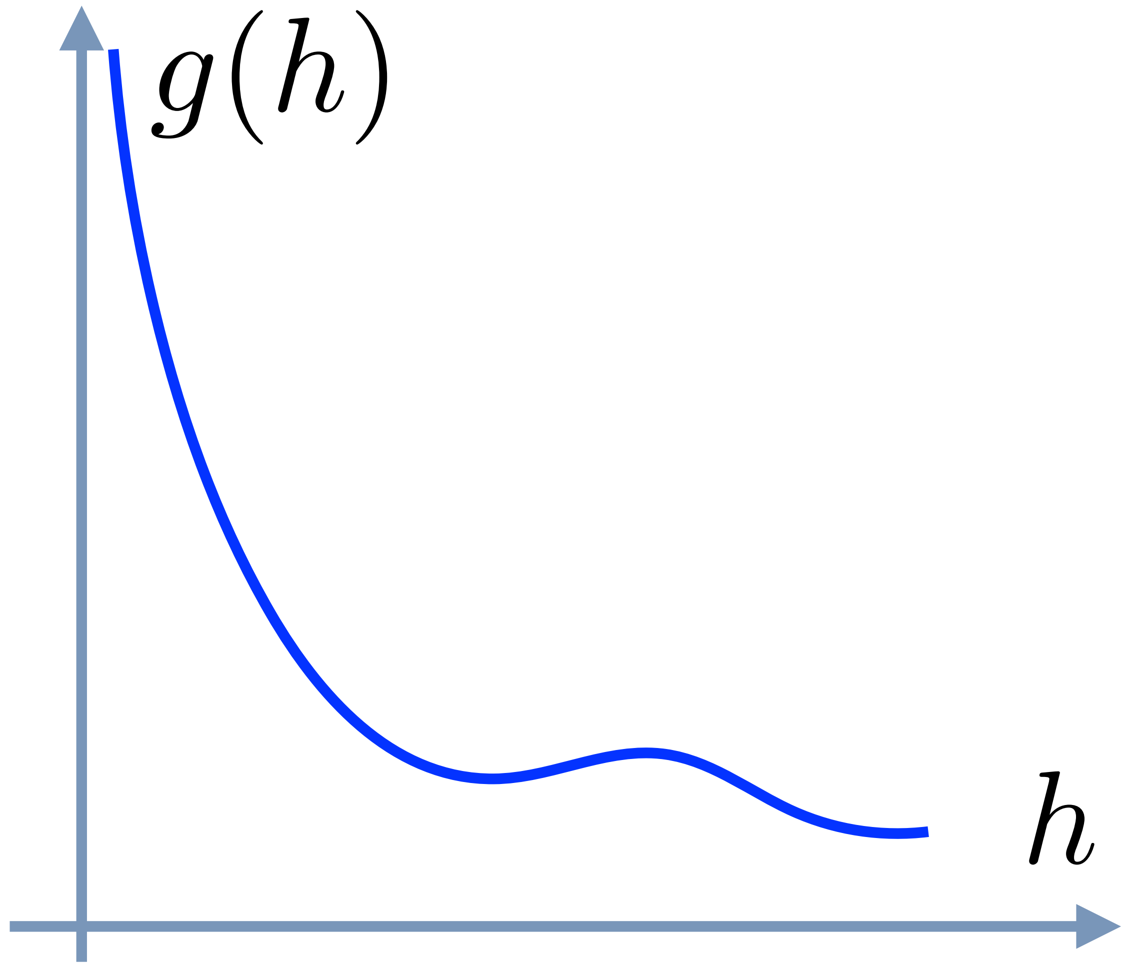


$$P(f) \sim f^{\theta}$$



$$\gamma = 1/(2+\theta)$$

Gap distribution



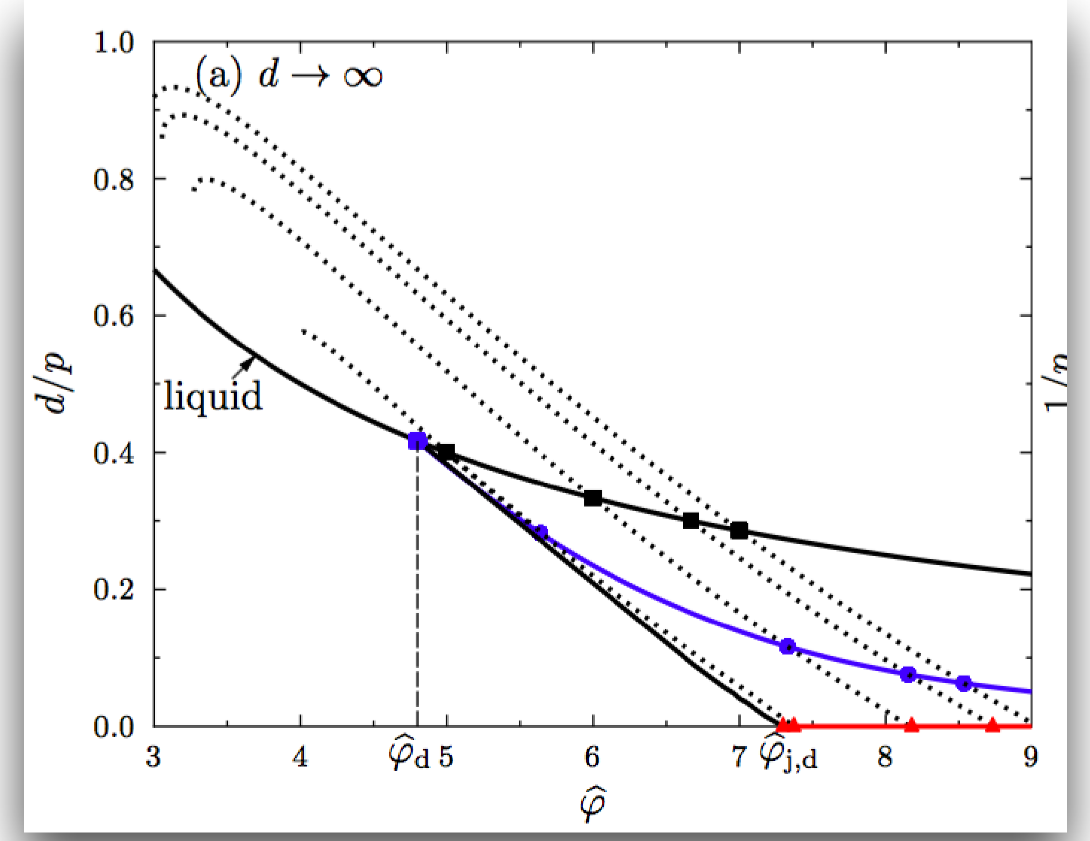
$$g(h) \sim h^{-\gamma}$$

θ & γ Non trivial exponents

Wyart, PRL 109 (12), 125502
Lerner, Düring, Wyart, Soft Matter 9 (34), 8252-8263
Charbonneau, Corwin, Parisi, Zamponi, PRL 114 (12), 125504

High Dimensional Soft Spheres

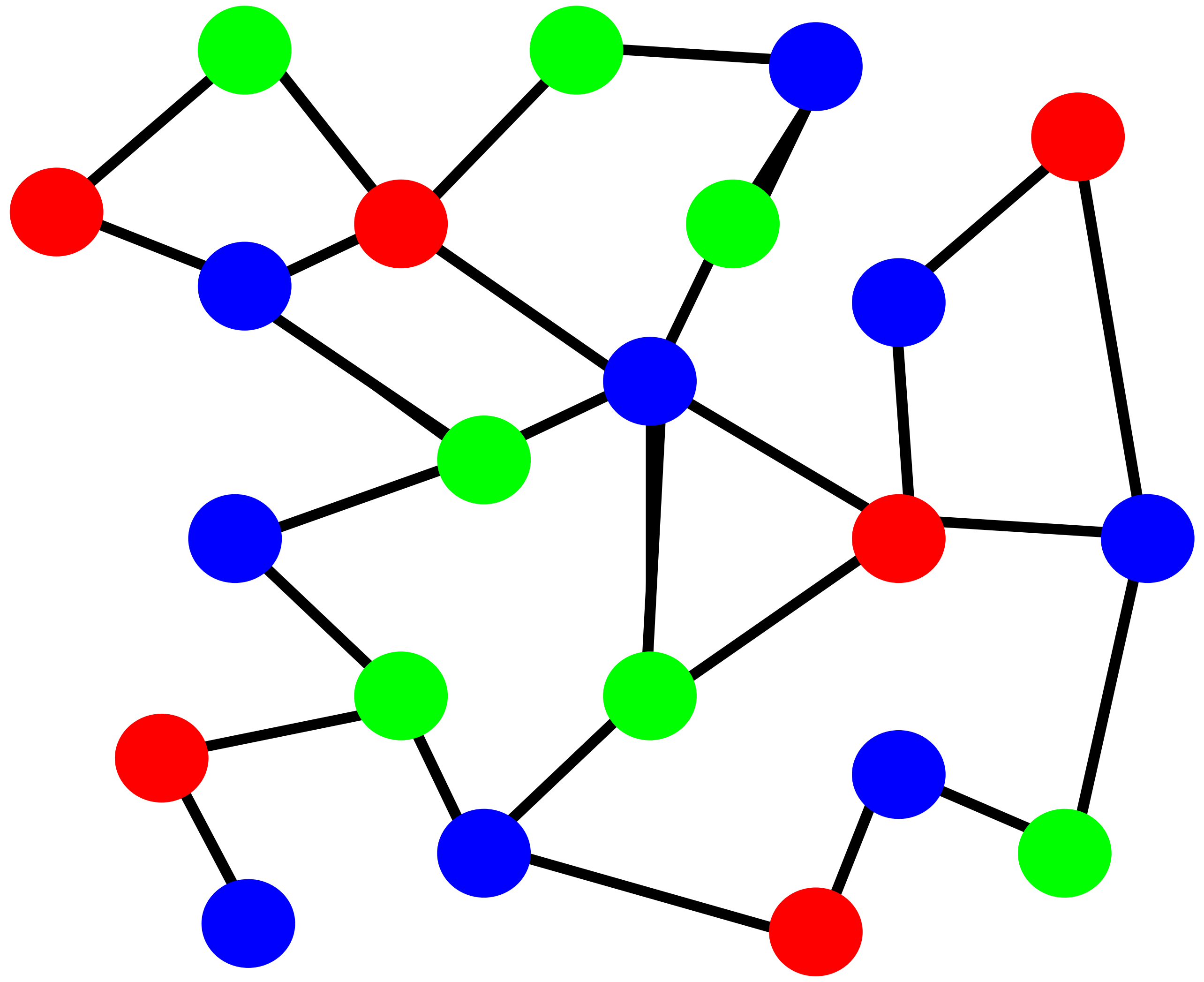
- Exact Solution for $D \to \infty$
- Ideal glass transition & Jamming
- Existence of a low temperature marginal glass
- Computation of Jamming Exponents $\theta = 0.42311$ $\gamma = 0.41269$
- Surprisingly exponents do not depend on D in simulations



Charbonneau Parisi Kurchan Urbani Zamponi 2014

Stat-Phys of Constraint Satisfaction

- **Graph Coloring (COL):** Given a graph and q colors, colour the vertexes such that no edges insists on vertexes of the same colour — Ground State of a Potts Antiferromagnet —



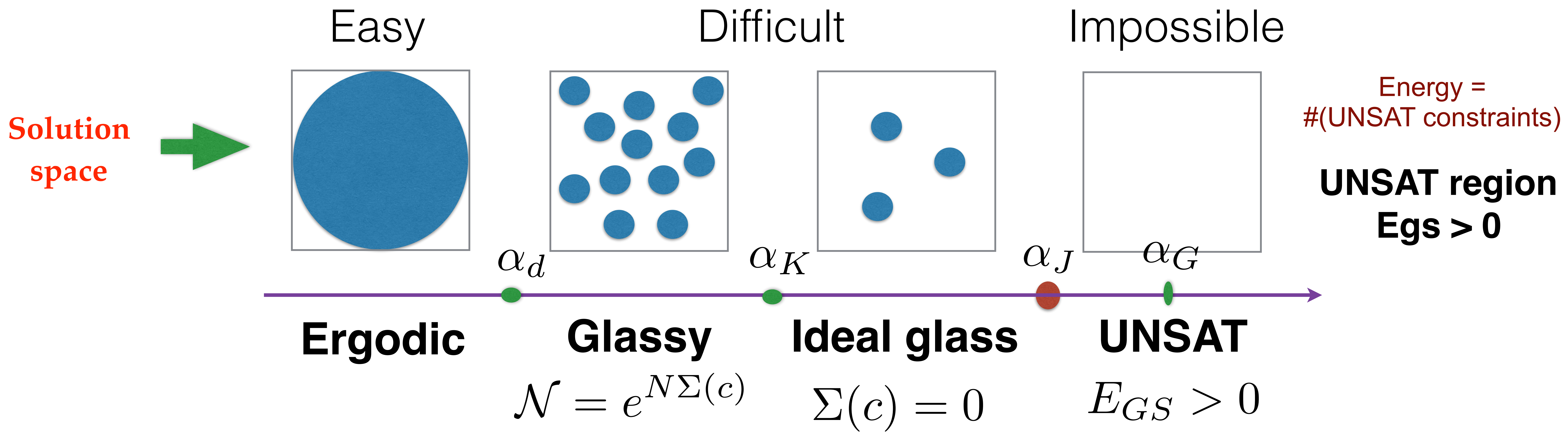
$$H[x] = \#(\text{violated constraints})$$

- More generally, **Constraint Satisfaction Problems:** given N variables and M constraints
  (1) Does it exist a SAT assignment
  (2) Find it.

Discrete Problems : Important in Computer Science

Random CSP

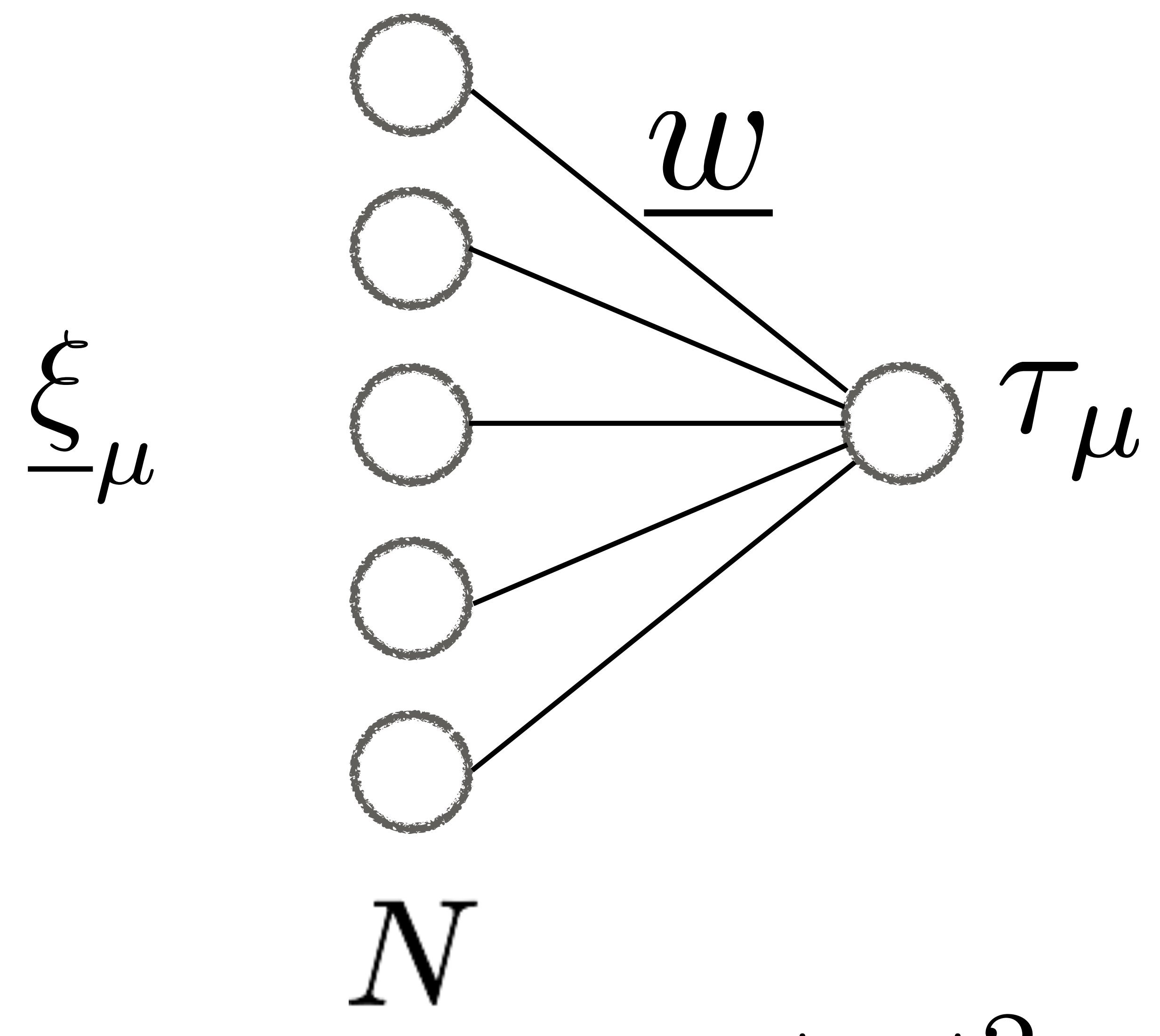
- N Variables, M Constraints. $N \to \infty$ relevant variable $\alpha = M/N$ (Density of constraints)



Exact Solution with Cavity-RSB techniques Mezard Parisi Zecchina Science 297, 812 (2002)
Krzakała, Montanari Ricci-Tersenghi Semerjian Zdeborova PNAS 104.25 (2007): 10318 + many others

The Perceptron

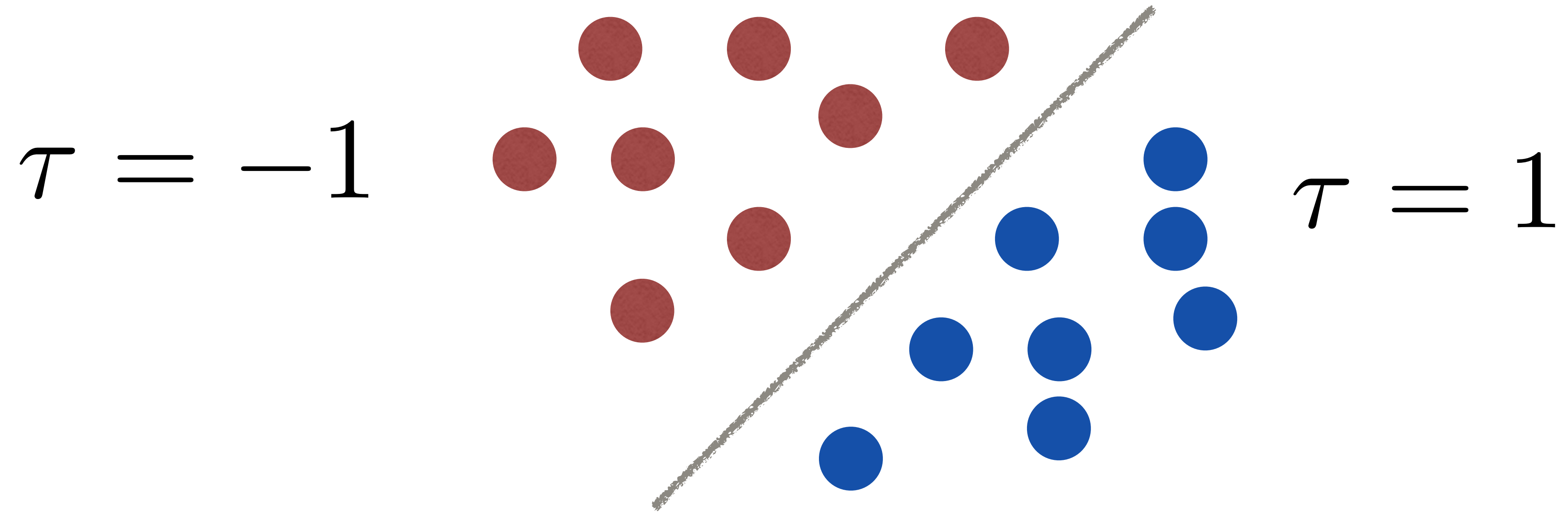
Model for a Single Neuron — Linear Classifier



$$\tau_\mu = \mathrm{sgn}\left(\frac{1}{\sqrt{N}}\underline{\xi}^\mu \cdot \underline{w}\right)$$

$$\tau_\mu \frac{1}{\sqrt{N}}\underline{\xi}^\mu \cdot \underline{w} > 0$$

$$\tau_\mu \frac{1}{\sqrt{N}}\underline{\xi}^\mu \cdot \underline{w} > \sigma$$ add stability



$$|\underline{w}|^2 = N$$

$$h_\mu = \frac{1}{\sqrt{N}}\tau_\mu \underline{\xi}_\mu \cdot \underline{w} - \sigma$$

Gaps

$$H[\underline{w}] = \frac{1}{2}\sum_{\mu=1}^{\alpha N} h_\mu^2 \theta(-h_\mu)$$

Cost function:
harmonic perceptron

McCulloch & Pitts 1936, Rosenblatt 1958

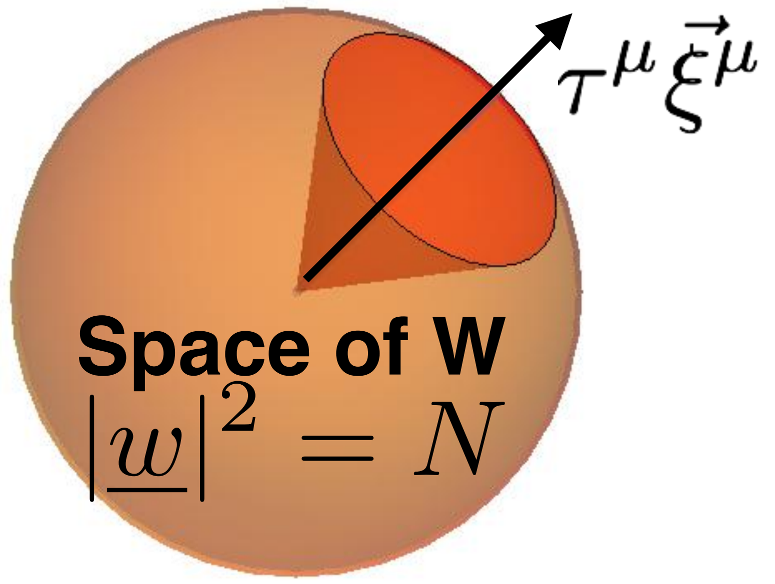
Forces= ‘slack variables’ = Lagrange multipliers

Information Storage Problem

- Data without structure ξ_i^μ τ^μ Random associations

$$\tau_\mu \frac{1}{\sqrt{N}} \underline{\xi}^\mu \cdot \underline{w} > \sigma$$

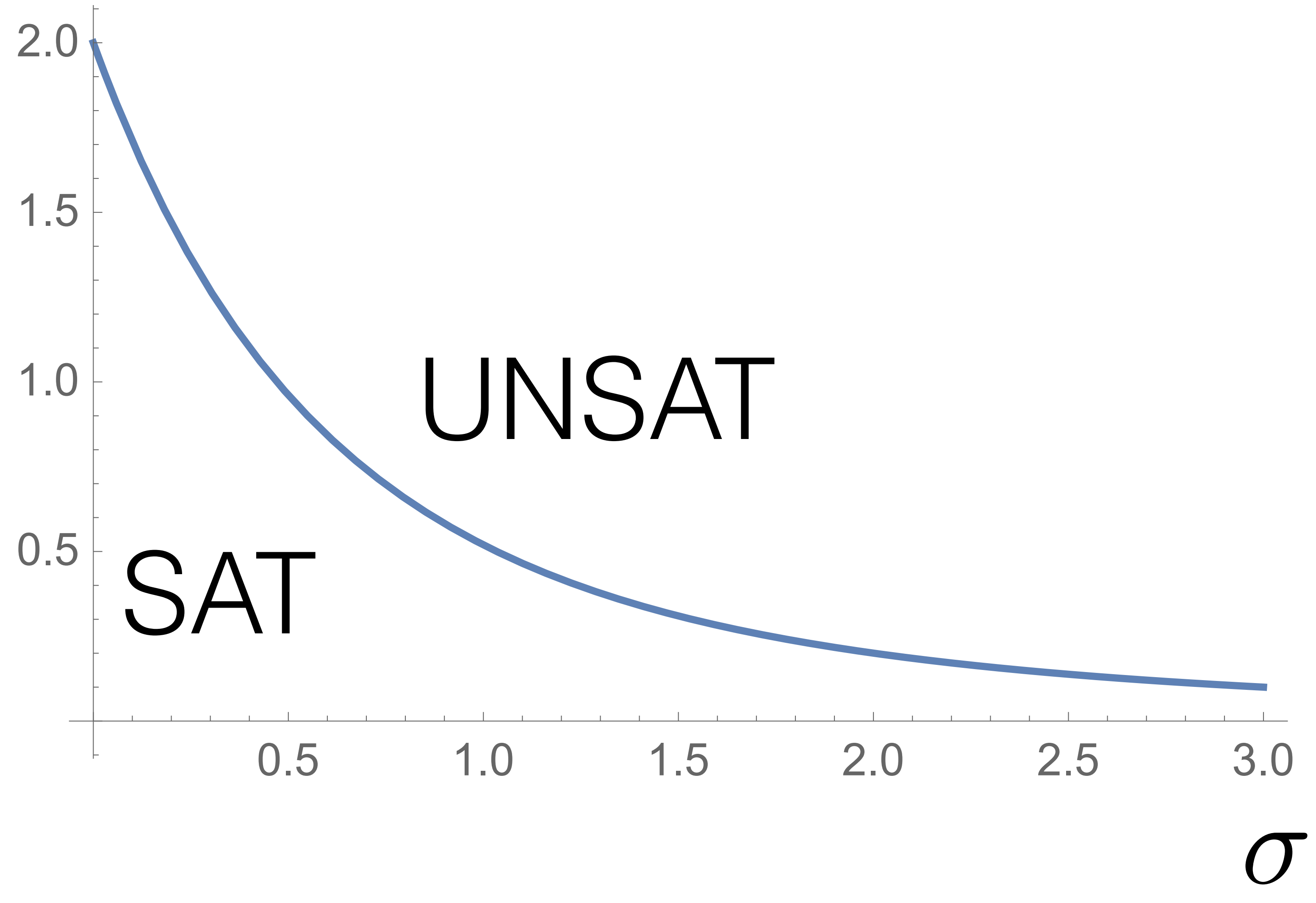
$\sigma > 0$



Exactly Solvable

Easy Optimisation
Convex

$\alpha = M/N$

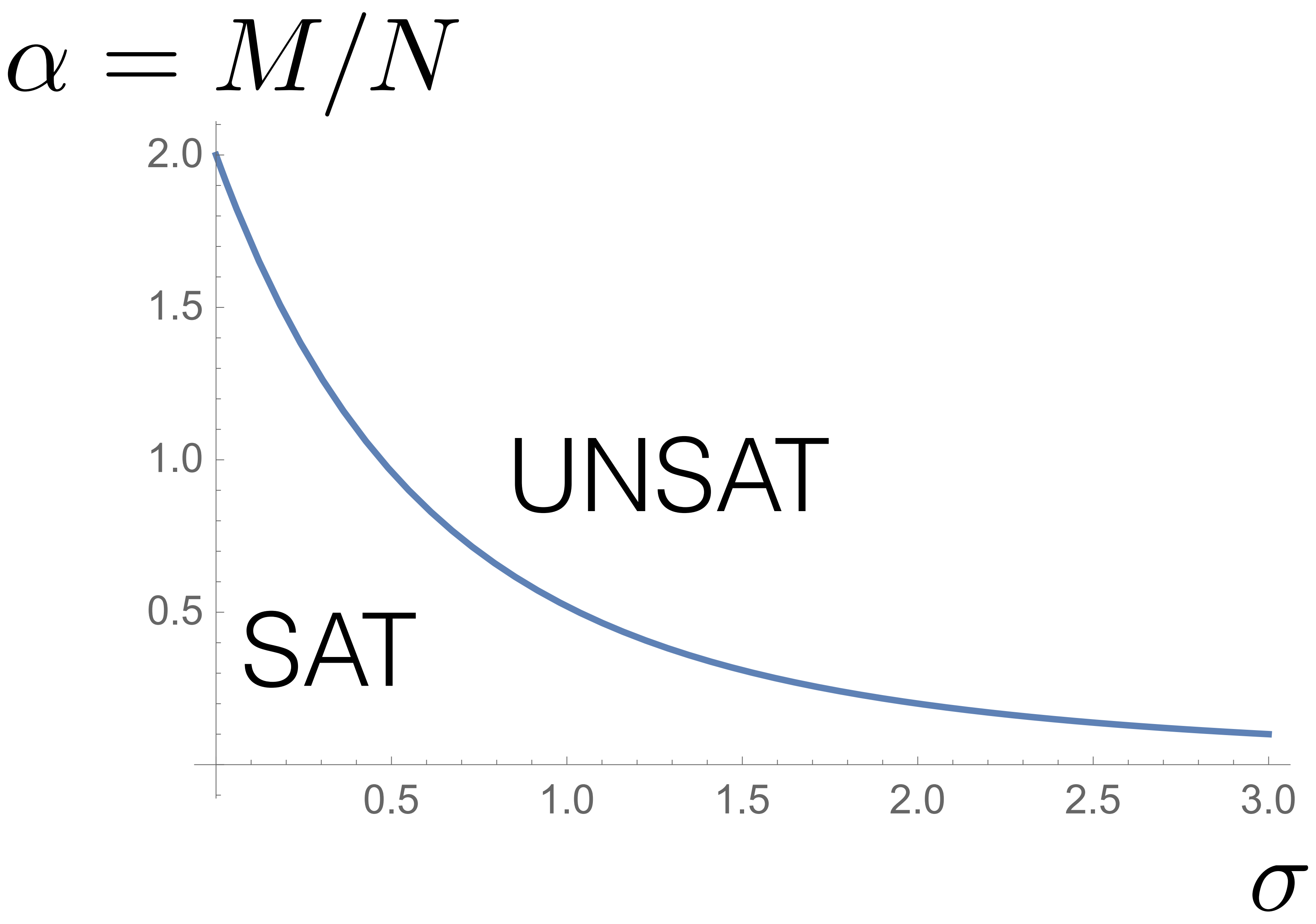


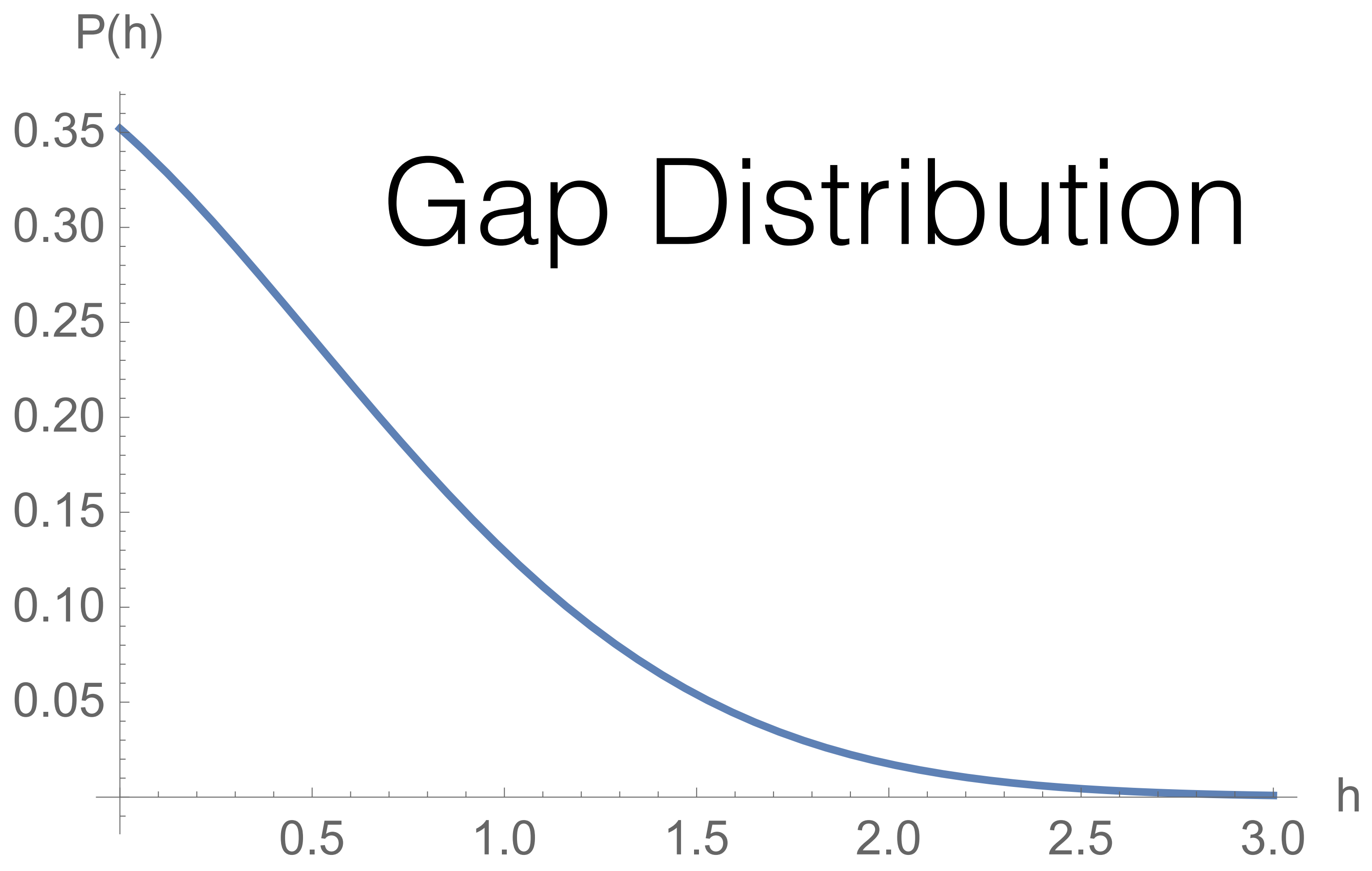
- Limit of capacity α_c Jamming SAT-UNSAT

SAT phase: Unique cluster of solutions
UNSAT phase: Unique minimum

Gardner + Gardner Derrida 1988

The Capacity limit of the Perceptron

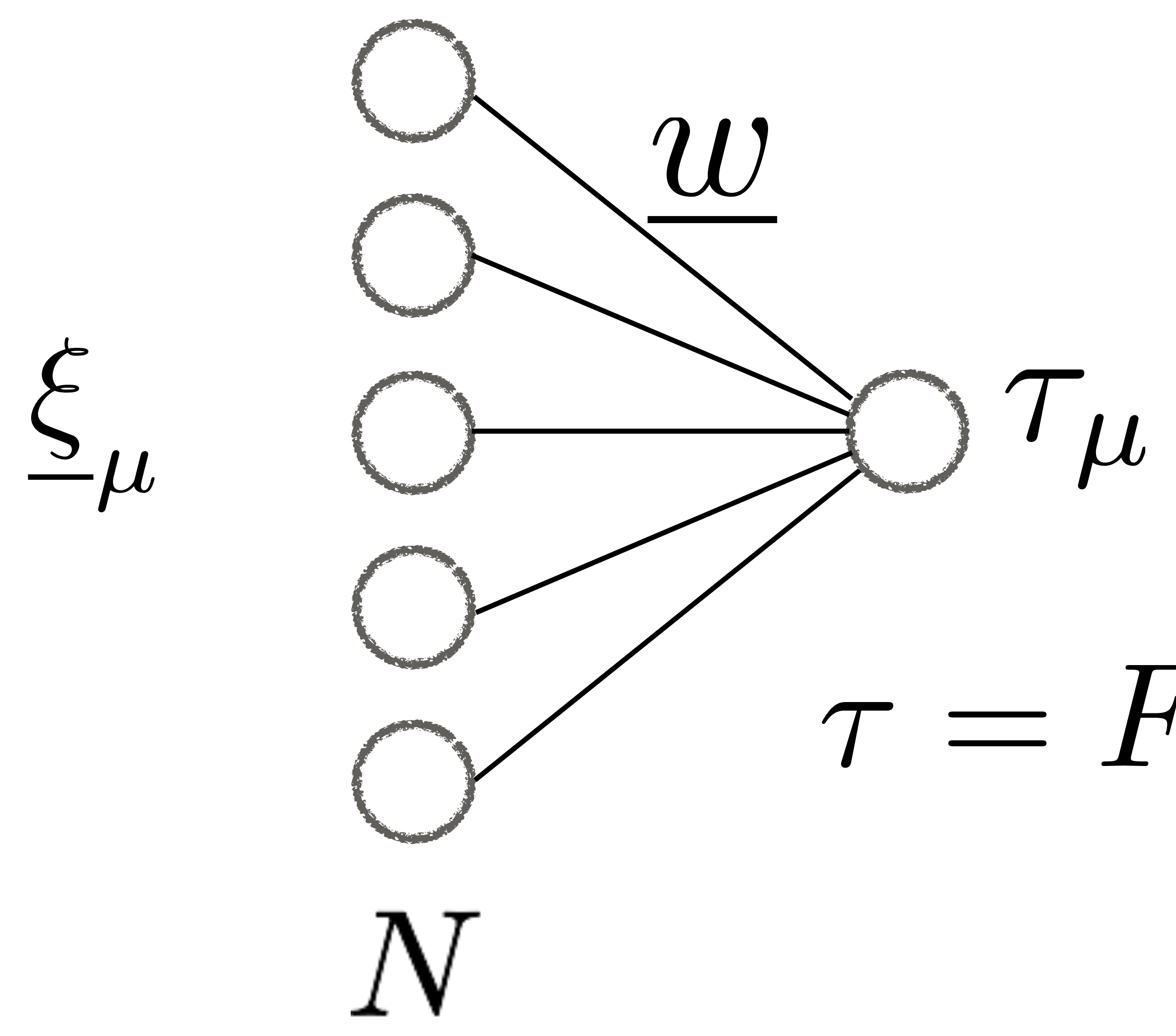




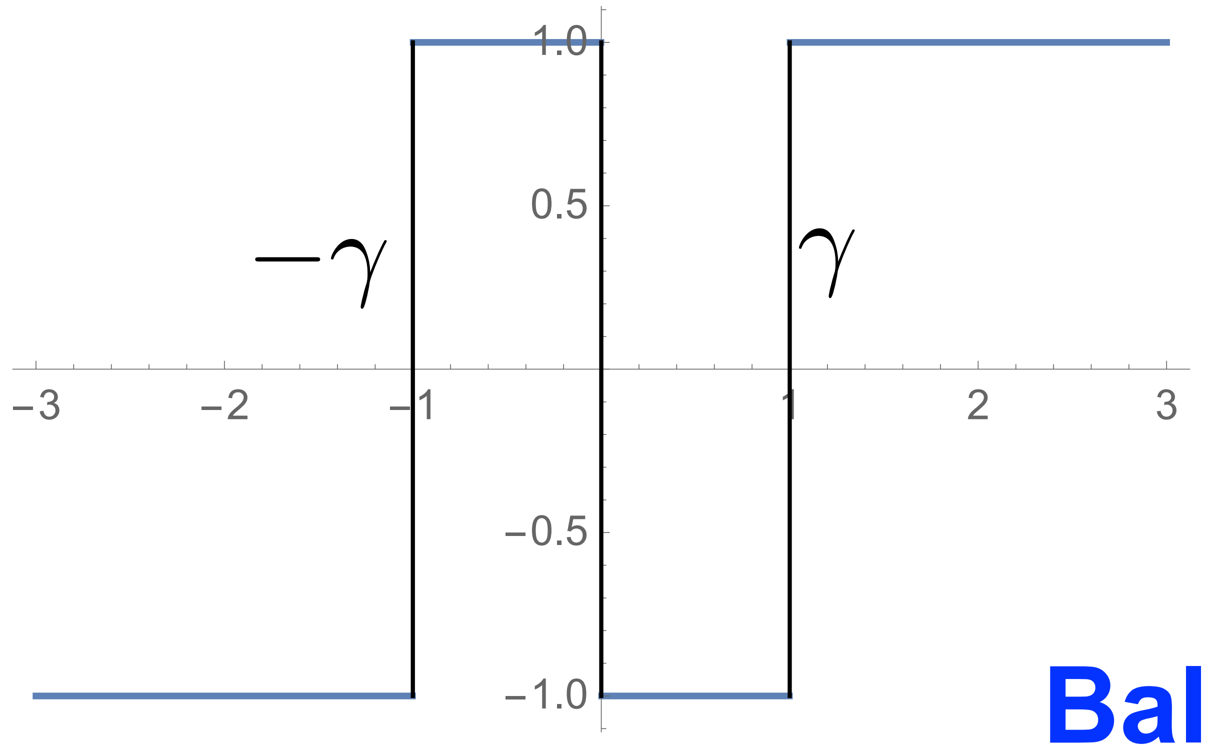
SAT : Convex space of solutions
UNSAT : Single Ground State

More interesting problems

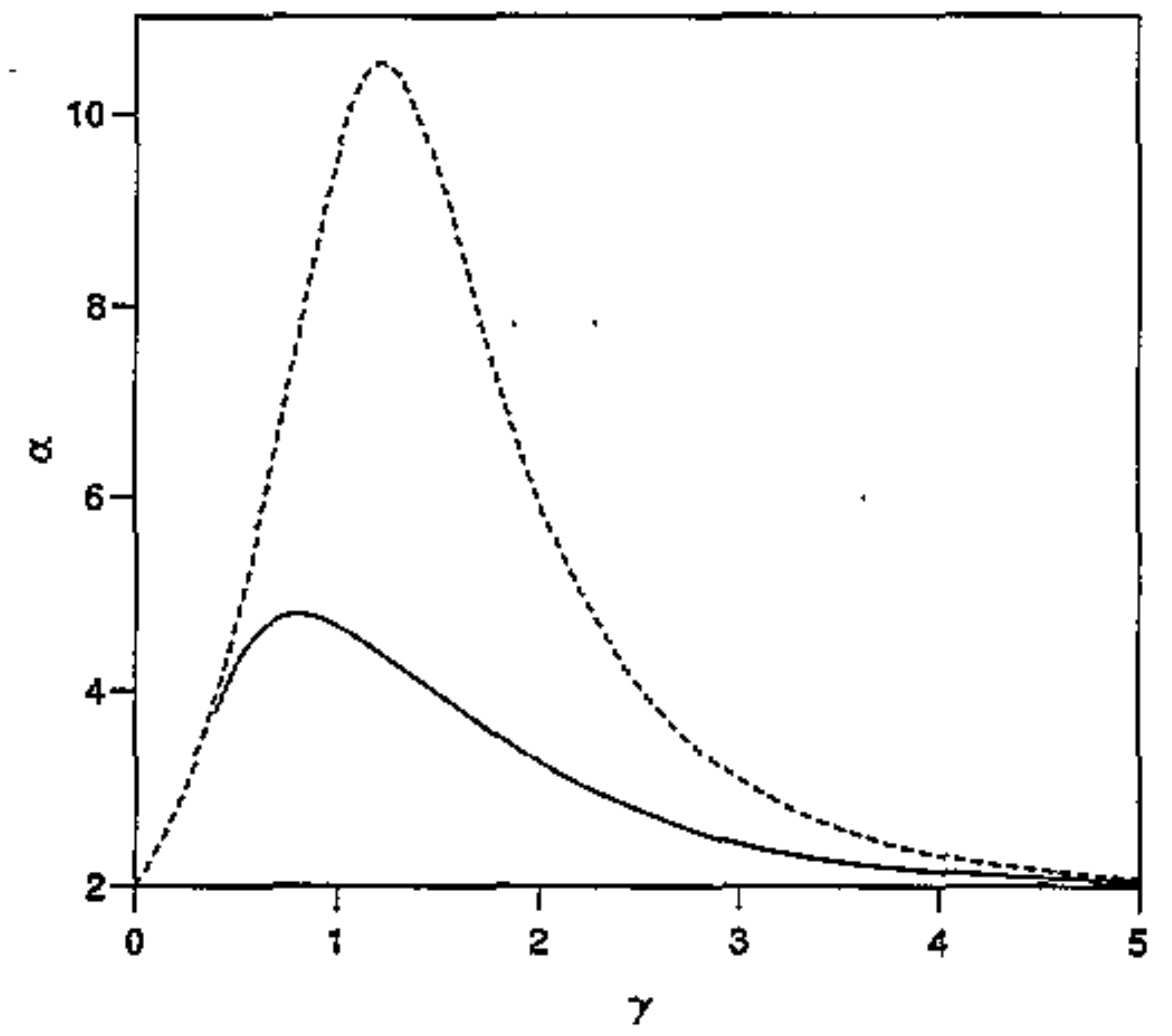
- **Multilayer — Non Convex problems**
- “Reverse Wedge Perceptron” : Perceptron with ‘internal representations’



$$\tau = F_\gamma[\vec{w} \cdot \vec{\xi}]$$



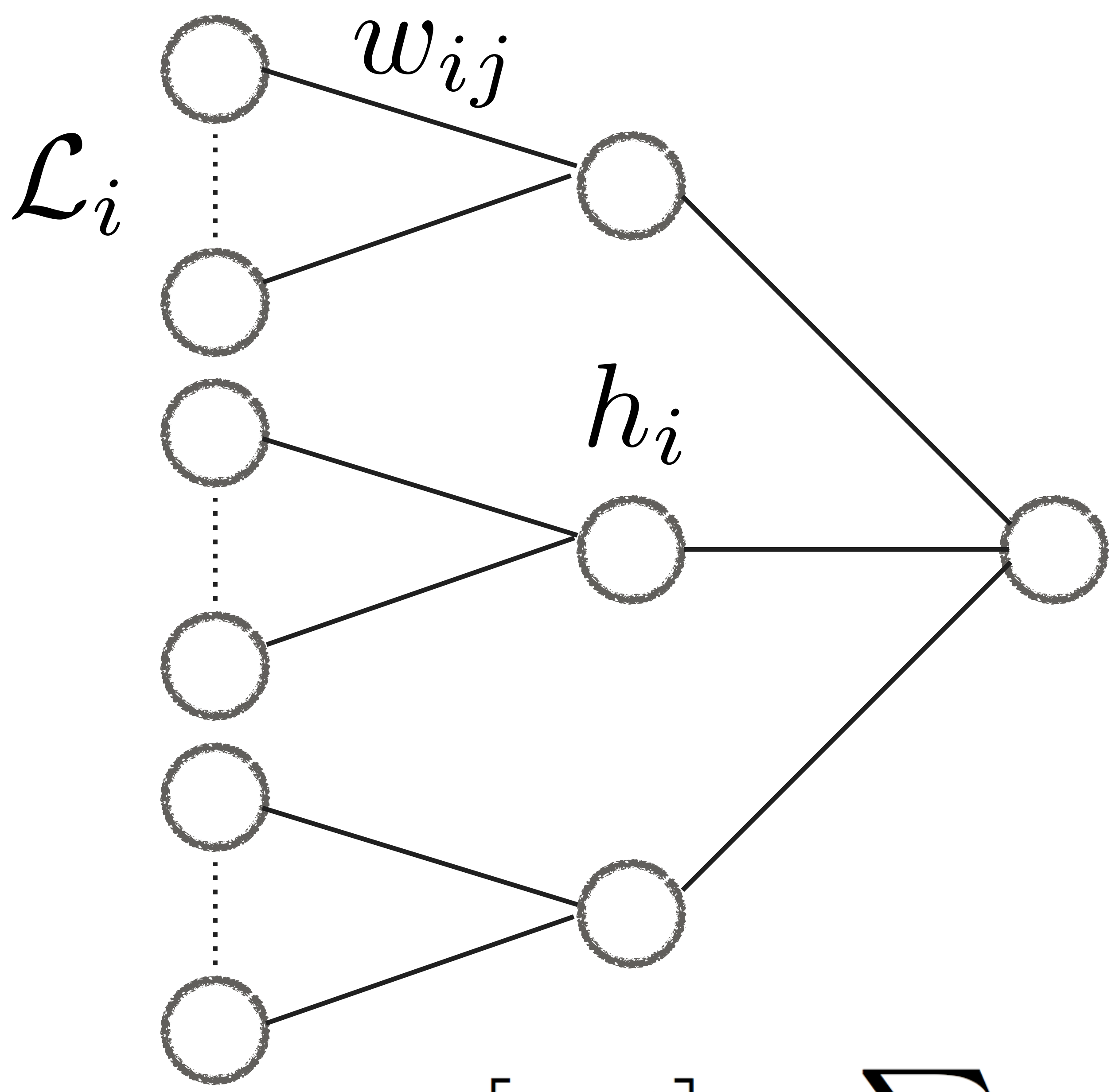
Jamming in a Glassy Phase



Baldassi Monasson Zecchina 1993

Multilayer neural networks

SF, Hwang, Urbani, 2018



Given random input-output associations $\{\xi^\mu, \tau_\mu\}$

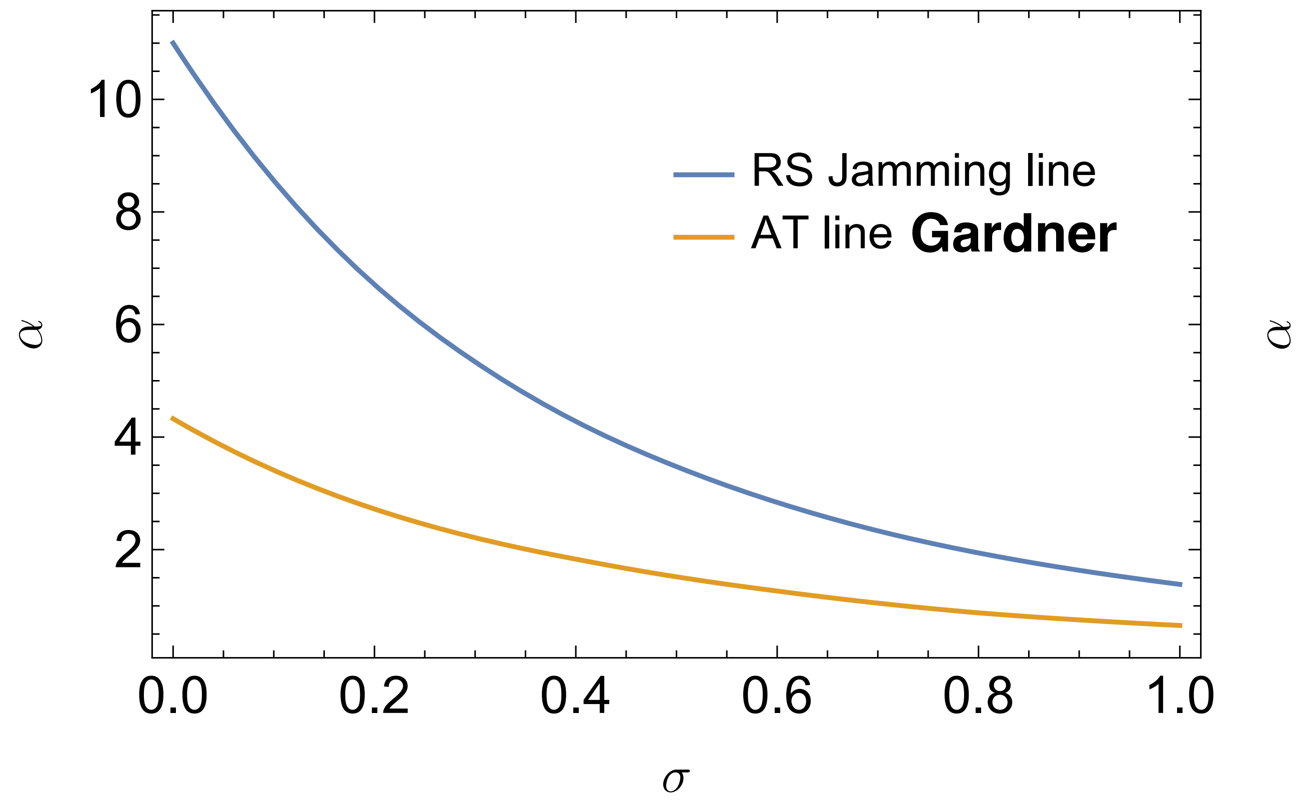
$$\text{Find } \underline{w} \text{ such that } \Delta_\mu > 0 \quad \forall \mu$$

$$\Delta_\mu = \Delta(h_1^\mu, \dots, h_K^\mu)$$

Contraints forbid K dimensional domains in the space of the h

$$h_i\left[\underline{w}, \xi\right] = \sum_{j \in \mathcal{L}_i} w_{i,j} \xi_j$$

$$\mathcal{F}[\underline{h}] = \begin{cases} \mathrm{sgn}\left[\prod_{i=1}^{K} h_i\right] & \textbf{parity} \\ \mathrm{sgn}\left[\sum_{i=1}^{K} \mathrm{erf} h_i\right] & \textbf{soft committee} \\ \mathrm{sgn}\left[\frac{1}{K}\sum_{i=1}^{K} \rho_{\mathrm{ReLU}}(h_i, \sigma) - \sigma\right] & \textbf{ReLU 2-layer} \end{cases}$$

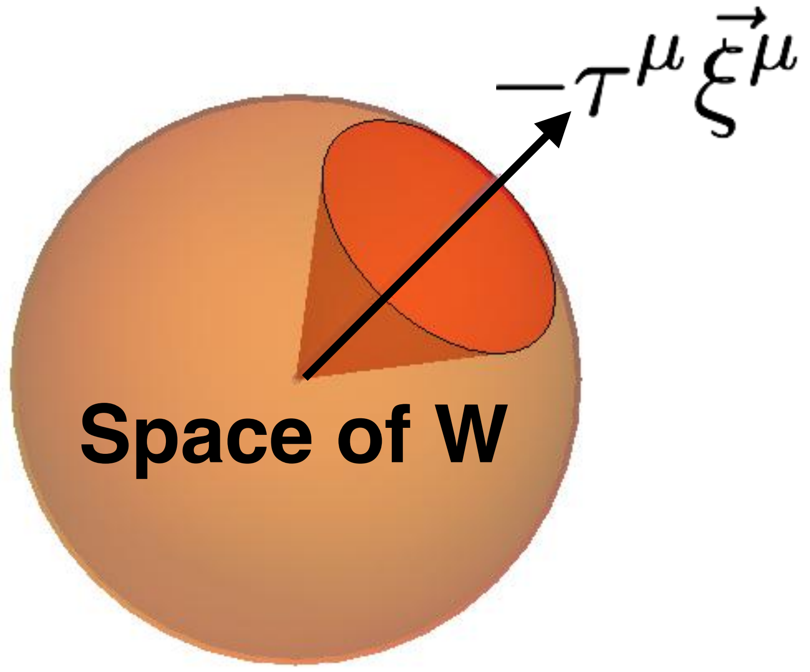


Monasson, Zecchina, Barkai, Hansel, Kanter, Engel, Sompolinsky, Saad, Solla '80-'90

Negative Perceptron

Each Constraint Excludes a Convex region of the sphere

Exactly Solvable



Red region Excluded

$$\tau_\mu \frac{1}{\sqrt{N}} \underline{\xi}^\mu \cdot \underline{w} > \sigma$$

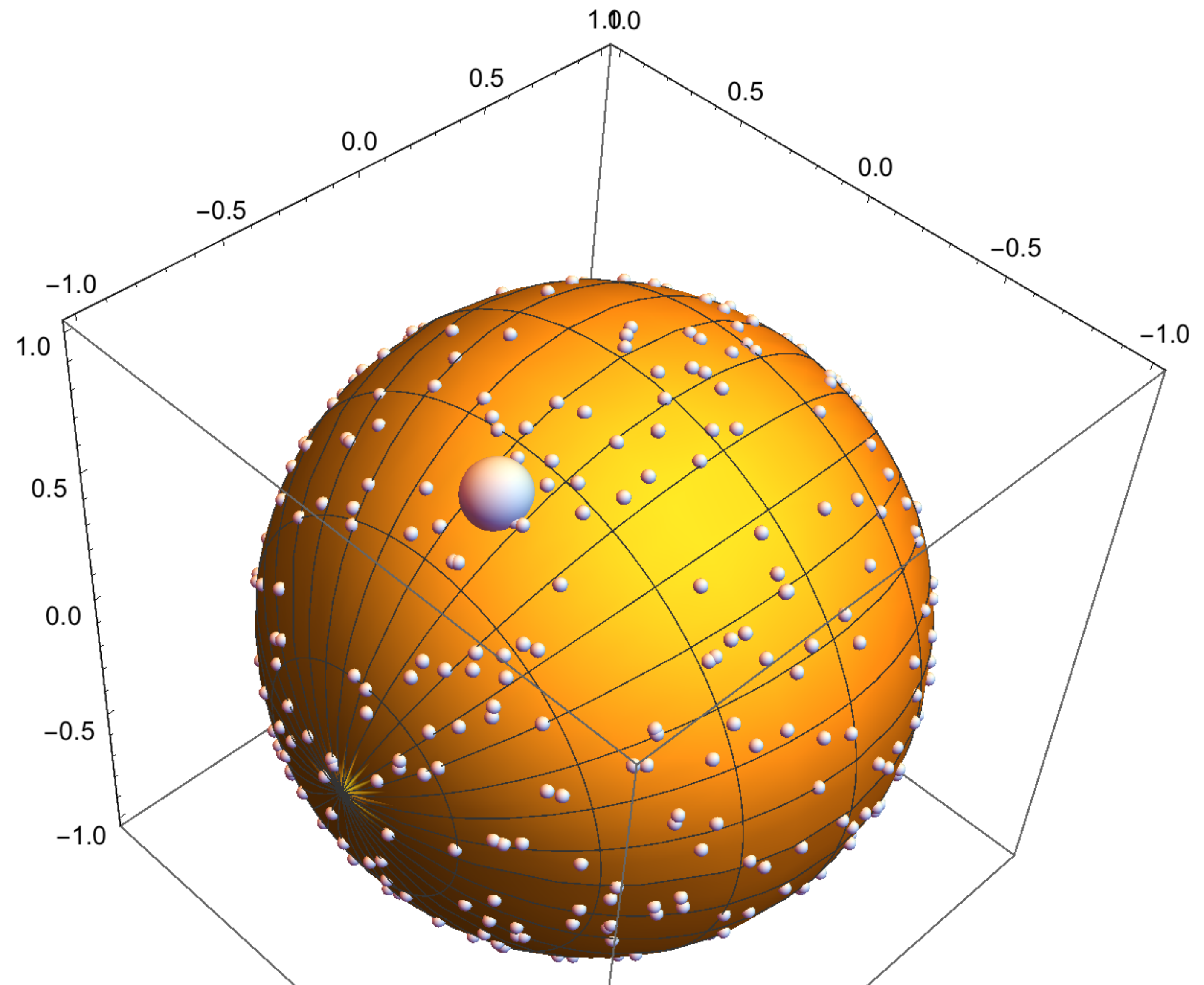
$$\sigma < 0$$

$$\alpha = M/N$$

mean field limit
Exact solution

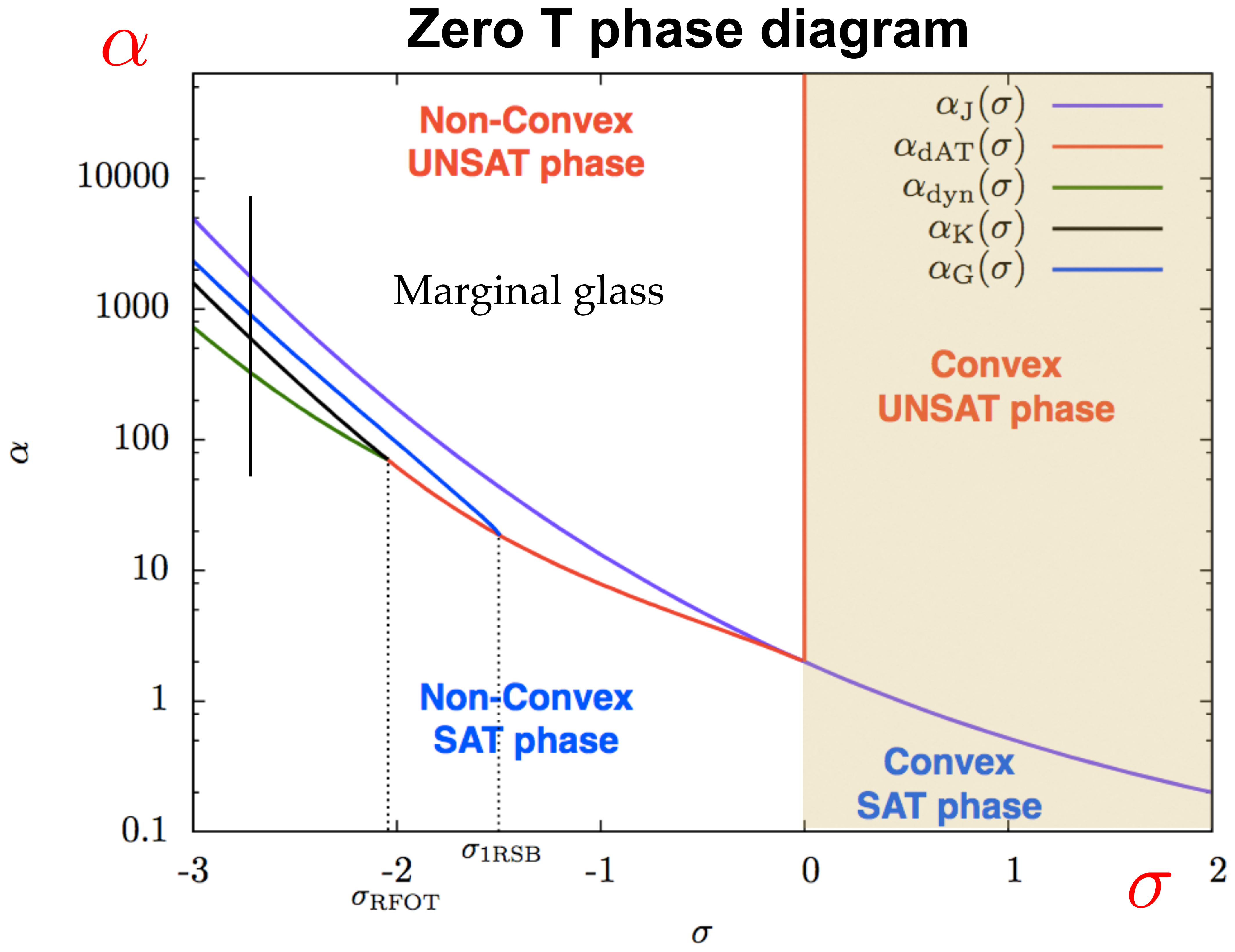
Lorentz Gas: Single Particle on the N Sphere + M obstacles

Non convex problem
Space of solutions can be disconnected
Multiple minima Glassy phases possible — RSB



SF, Parisi JPA 2015, SF, Parisi, Sevelev, Urbani, Zamponi Scipost 2017

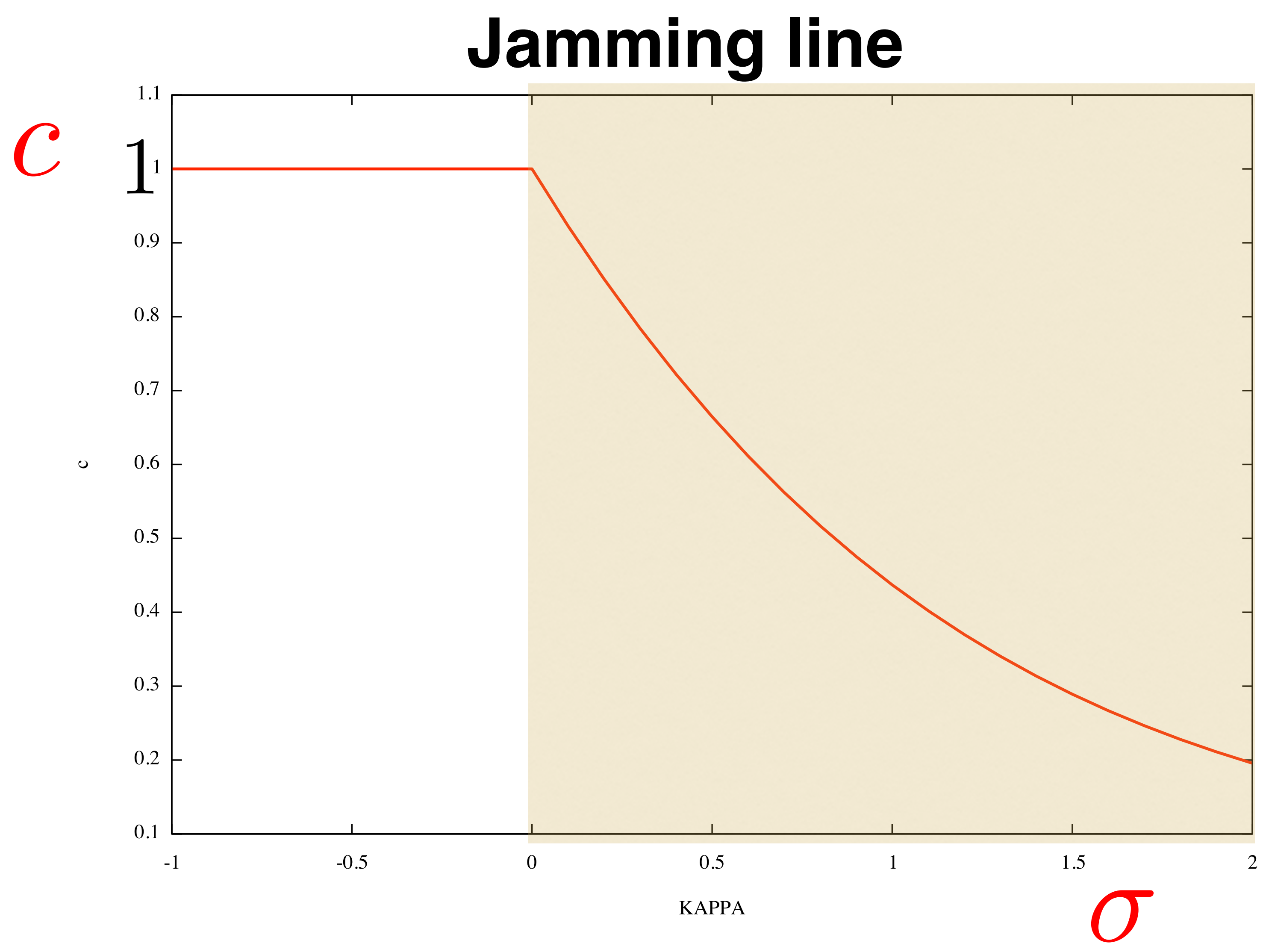
Negative Perceptron



Number of contacts

$c < 1$ hypostatic

$c = 1$ isostatic

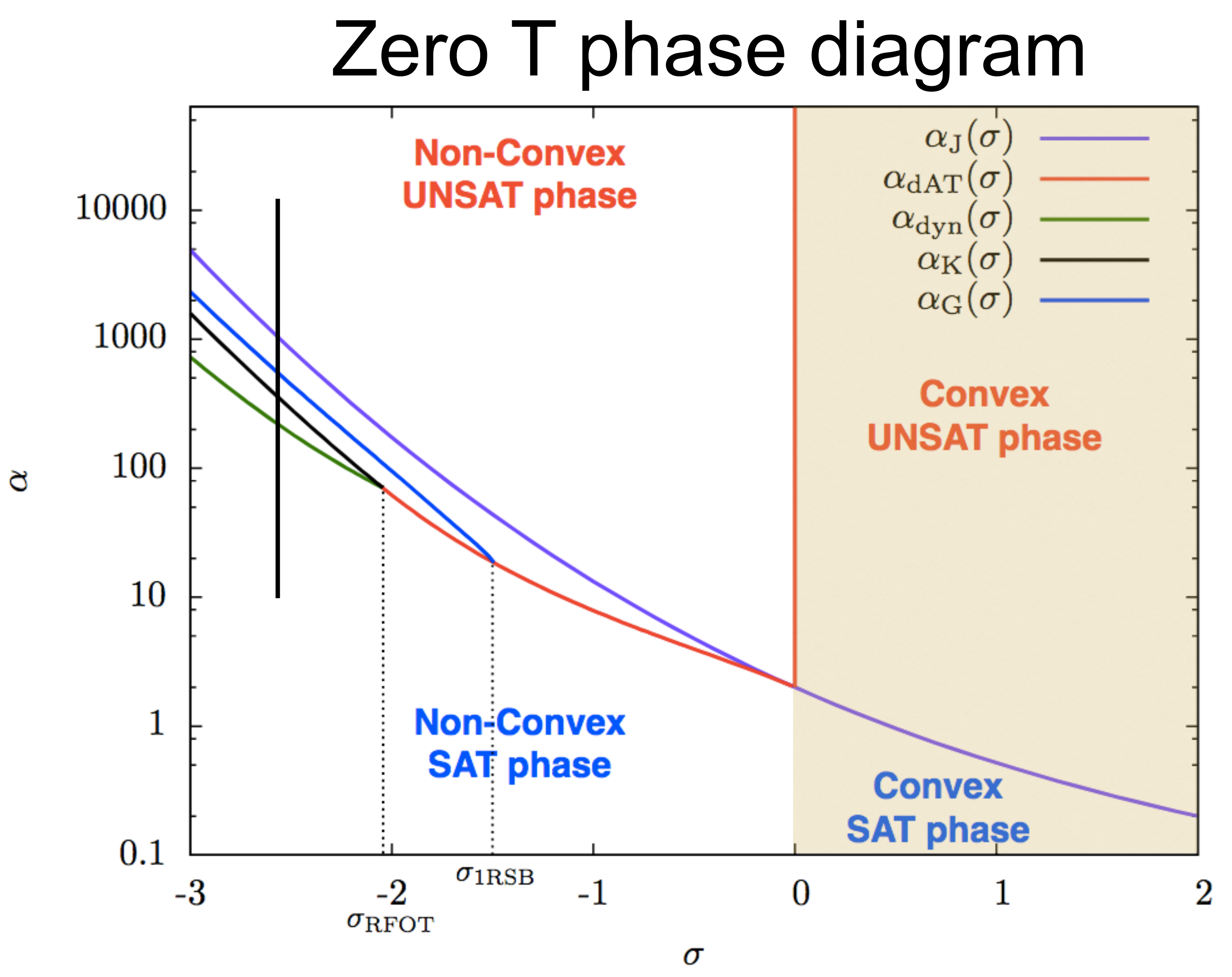


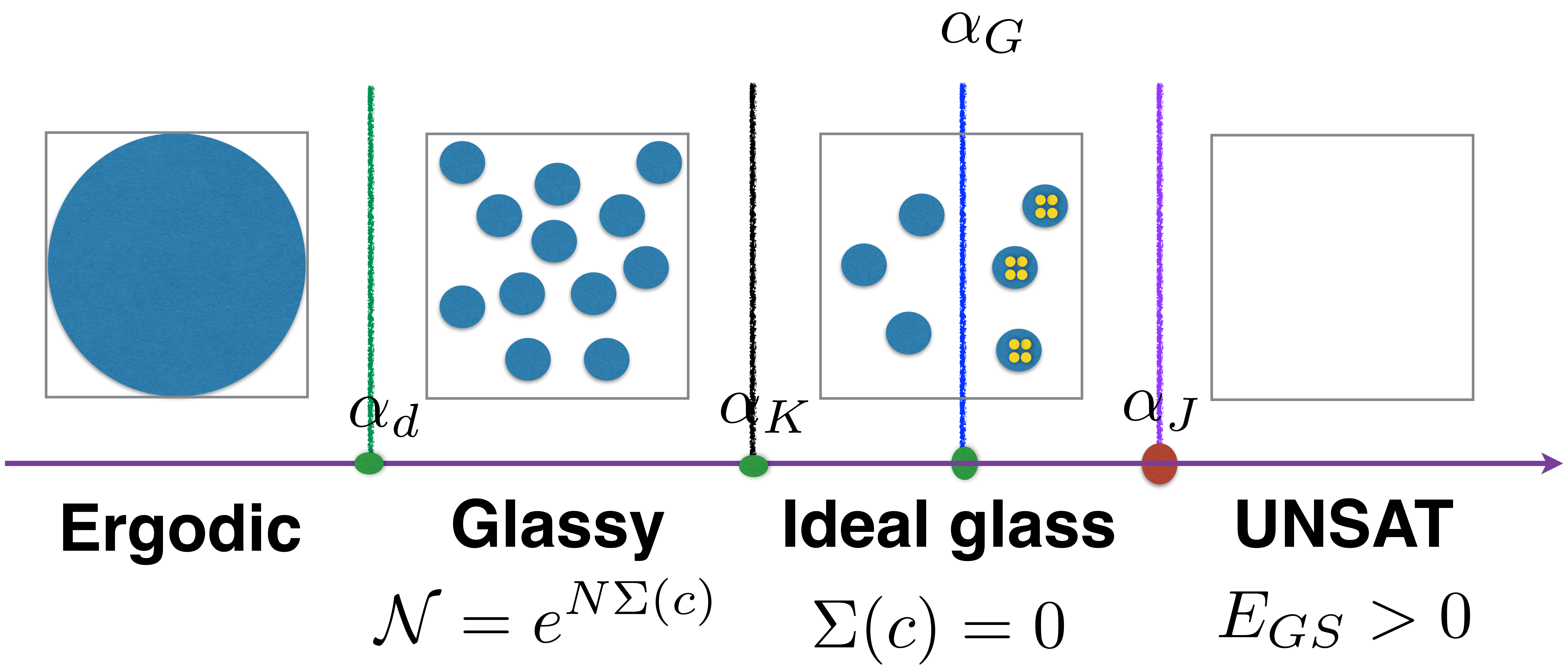
Universal Pattern : Liquid —> Stable Glass —> Marginal Glass —> Jamming

The jamming line lies in a Marginal (f-RSB) glassy

SF, Parisi JPA 2015, SF, Parisi, Sevelev, Urbani, Zamponi Scipost 2017

Negative Perceptron





Universal Pattern : Liquid —> Stable Glass —> Marginal Glass —> Jamming —> UNSAT

The jamming line lies in a Marginal (f-RSB) glassy phase

SF, Parisi JPA 2015, SF, Parisi, Sevelev, Urbani, Zamponi Scipost 2017

Isostatic Jamming

A scaling regime
close to jamming emerges
Same universality class of
hard spheres

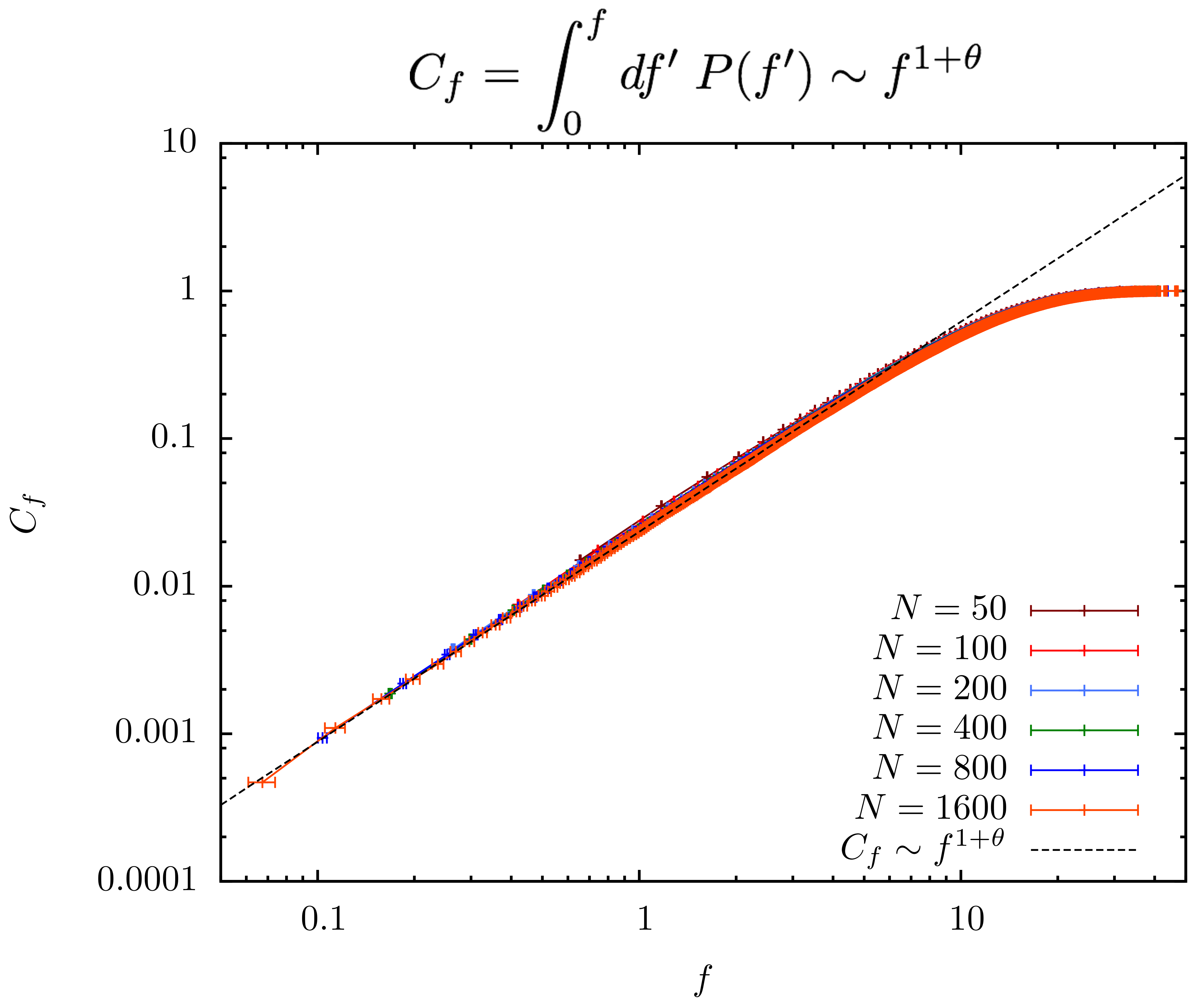
Pseudo-gap in force distribution

$$P(f) \sim f^{\theta} \qquad \theta = 0.42311$$

Singular Power Law in gap distribution

$$P(h) \sim h^{-\gamma} \quad \gamma = 0.41269$$

Same Exponents as Spheres !



SF, Parisi JPA 2015

Multilayer neural networks

Given random input-output associations $\{\xi^\mu, \tau_\mu\}$

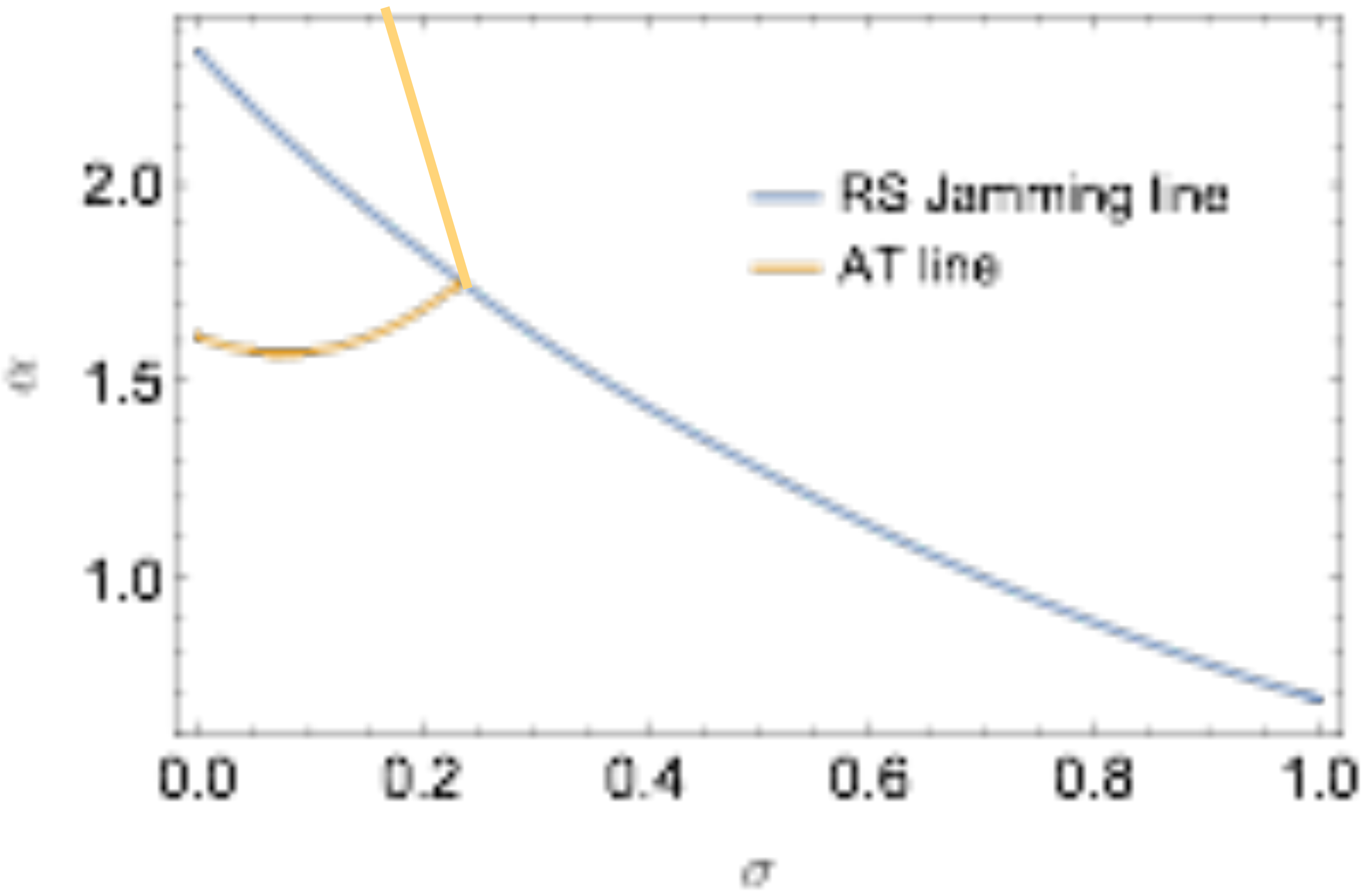
SF, Hwang, Urbani, 2018

Find $\underline{w}$ such that $\Delta_\mu > 0 \quad \forall \mu$

Gap variables

Parity $\Delta^\mu = \tau^\mu \prod_{i=1}^{K} h_i[\underline{w}, \underline{\xi}^\mu] - \sigma$

K dimensional domains in the space of the $\{h_1, ..., h_K\}$



K=3 Committee

Jamming from the ergodic phase
Hypostatic & Noncritical

Jamming from marginal glass
Isostatic & Critical

Multilayer neural networks

Given random input-output associations $\{\xi^\mu, \tau_\mu\}$

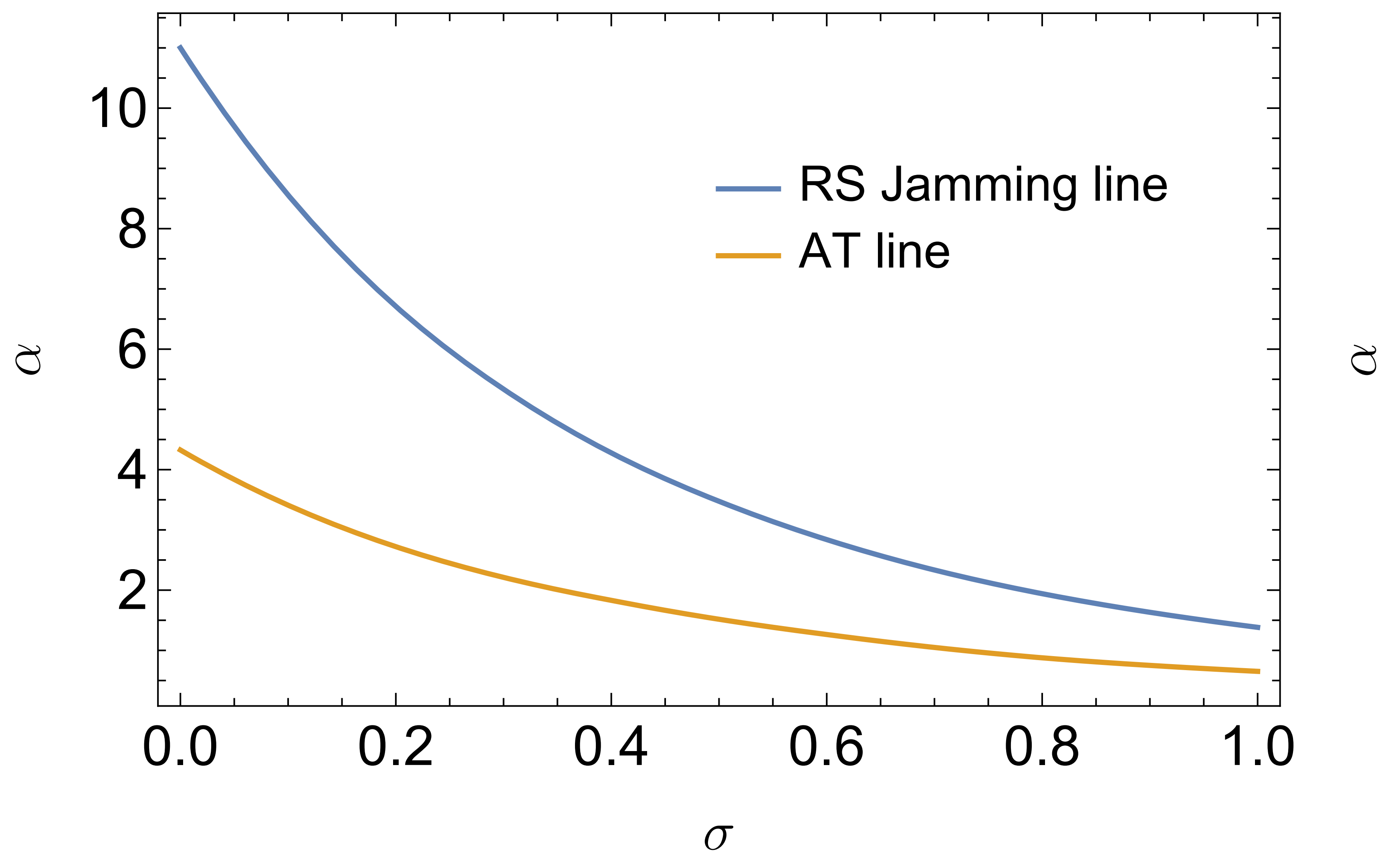
SF, Hwang, Urbani, 2018

Gap variables

Find $\underline{w}$ such that $\Delta_\mu > 0 \quad \forall \mu$

Parity $\quad \Delta^\mu = \tau^\mu \prod_{i=1}^{K} h_i[\underline{\underline{w}}, \underline{\xi}^\mu] - \sigma$

K dimensional domains in the space of the $\{h_1, ..., h_K\}$

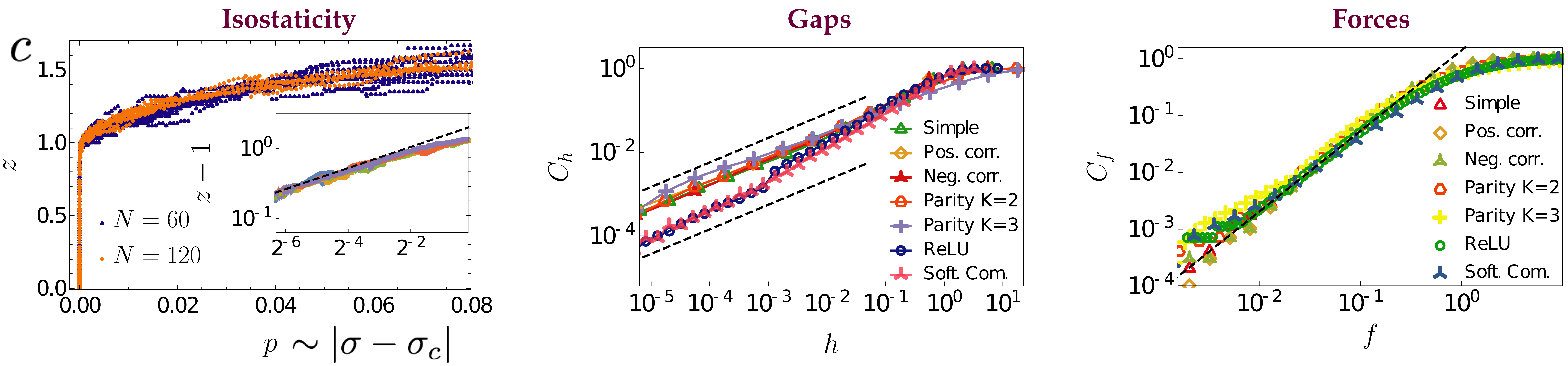


Jamming from the ergodic phase
Hypostatic & Noncritical

Jamming from marginal glass
Isostatic & Critical

Results at critical jamming

SF, Hwang, Urbani, 2018



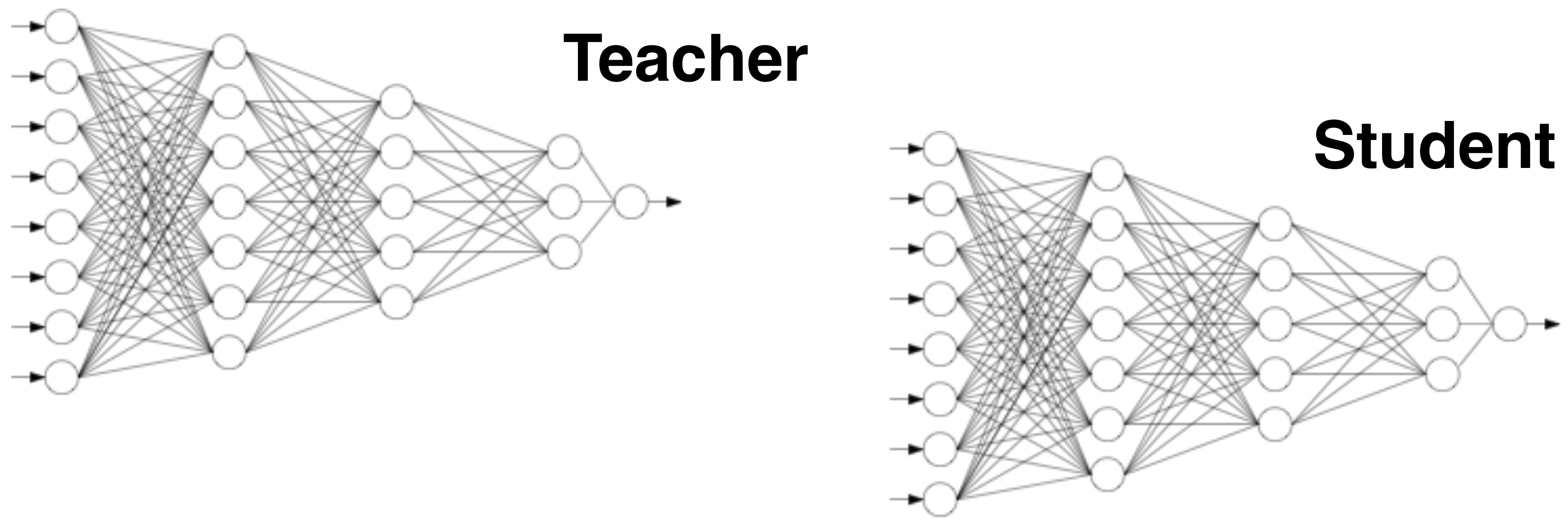
The hard spheres universality class is recovered

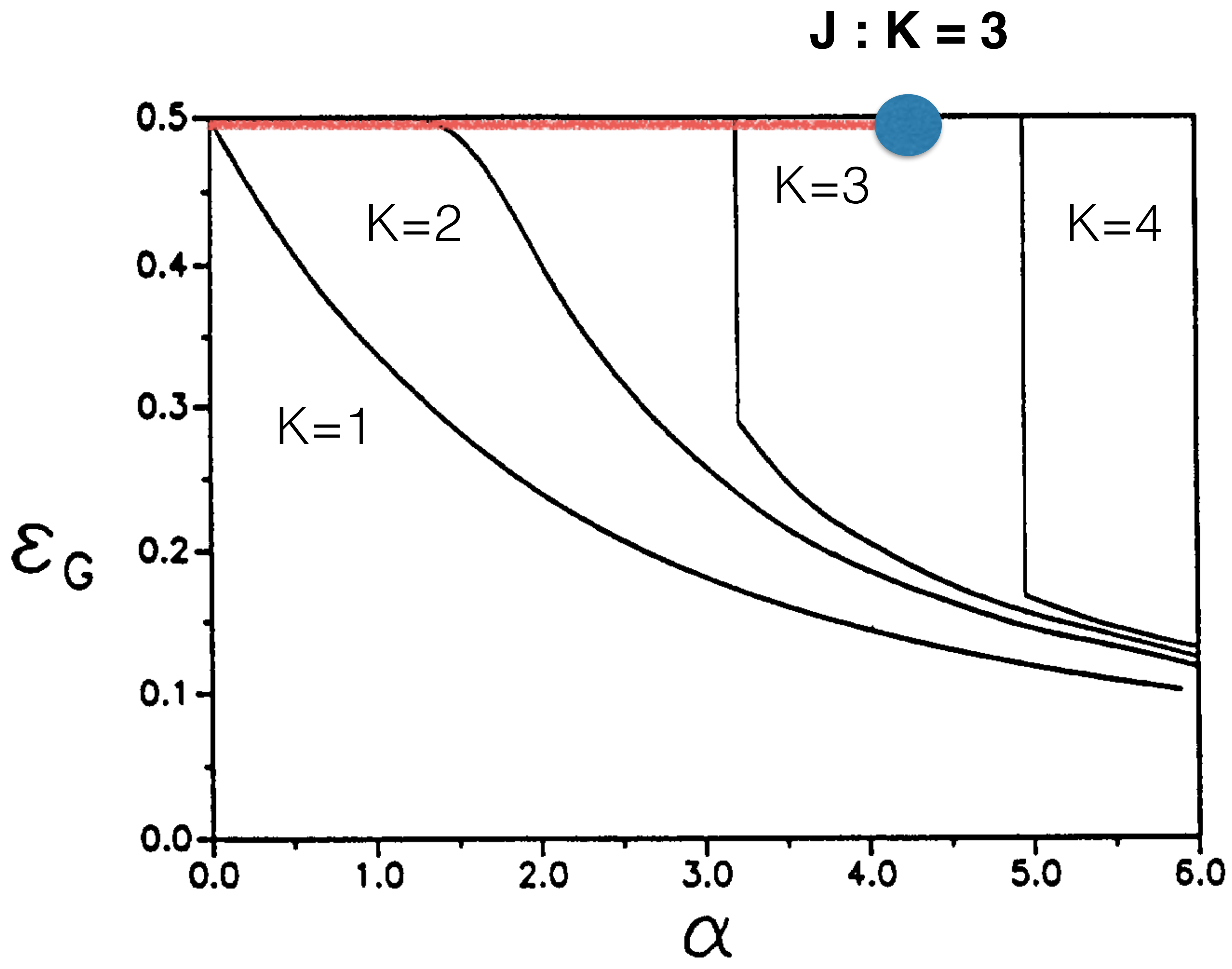
This happens even if the scaling equations that describe the jamming point are fullRSB equations in higher dimensions with respect to Perceptron.

We proved that a *dimensional reduction* mechanism takes place.

Consequences for Generalisation

Structured Data





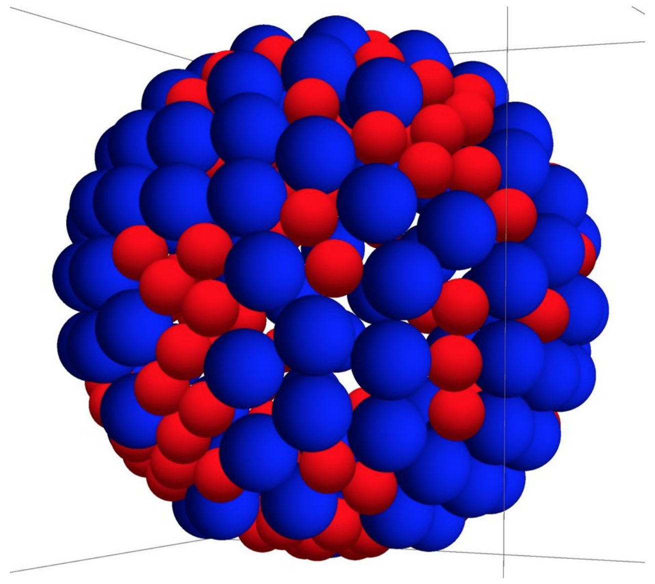
Finding the Teacher weights transition to crystal.
Possible glassy phases
No generalisation
Confusion phase above J

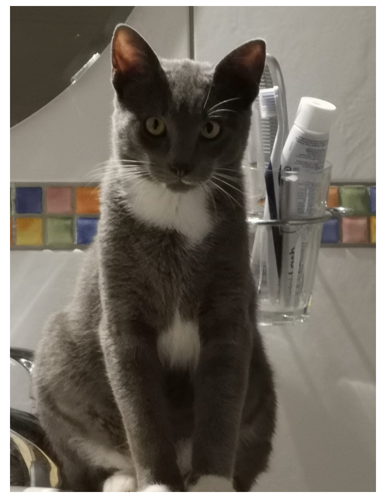
SF, Hwang, Urbani, 2018

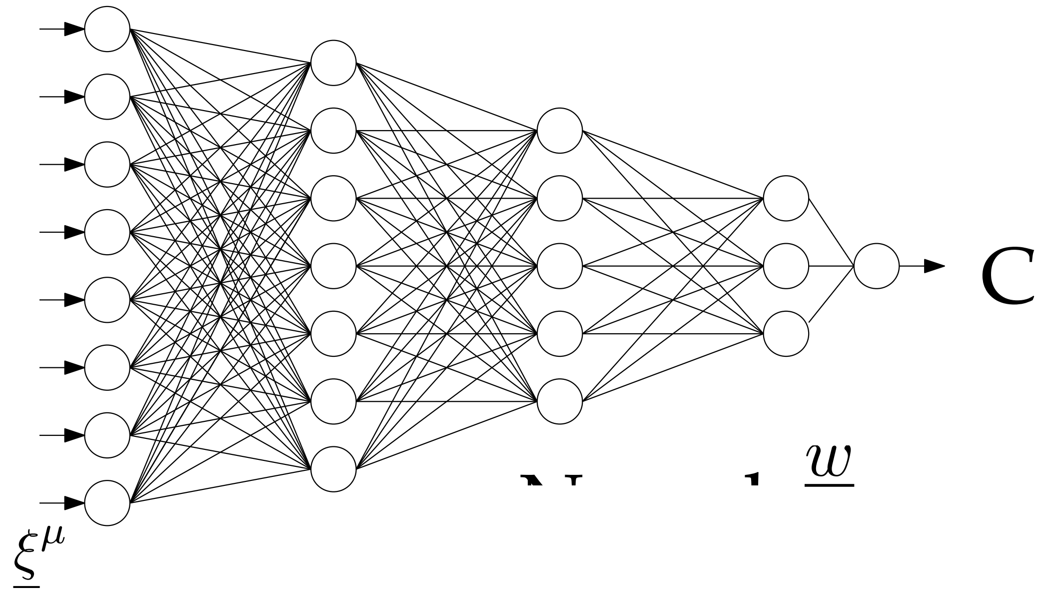
Super-Universal Hypothesis

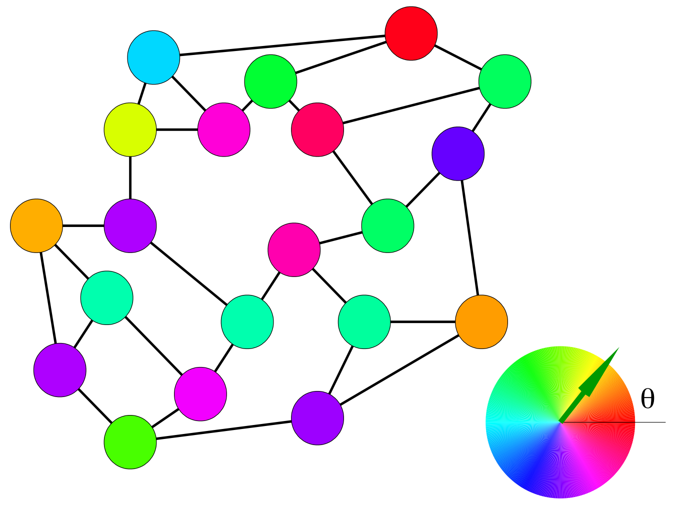
- Non-critical jamming in convex problems — ergodicity
- Unique universality class for Jamming criticality in random non-convex CSP isostatic points — Jamming in a critical glass.

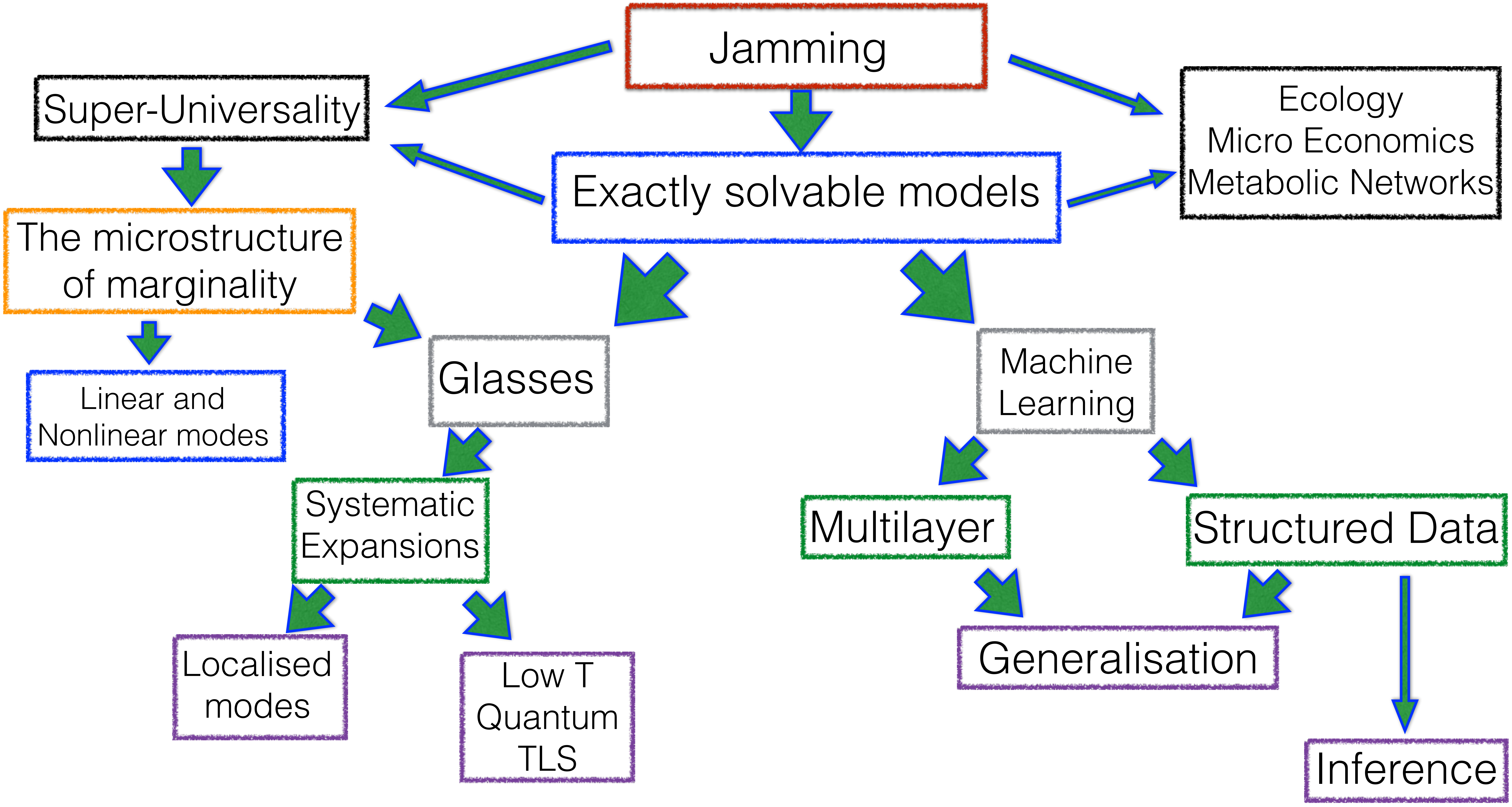
SF, Parisi 2015







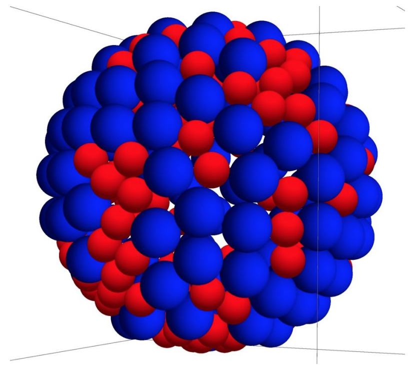



Jamming
Super-Universality
Exactly solvable models
Ecology
Micro Economics
Metabolic Networks
The microstructure of marginality
Linear and Nonlinear modes
Glasses
Machine Learning
Systematic Expansions
Multilayer
Structured Data
Localised modes
Low T Quantum TLS
Generalisation
Inference

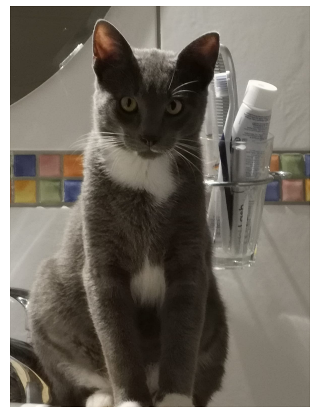
Continuous CSP

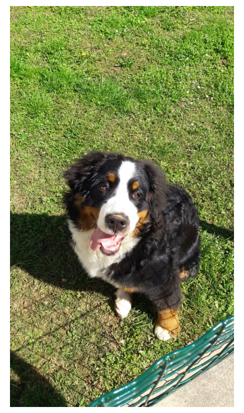
- CSP with continuous variables

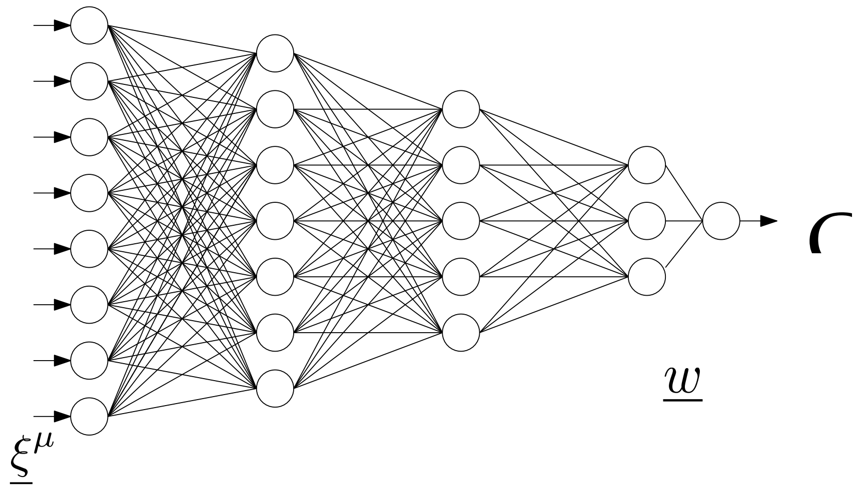
Sphere Packing
pays + math



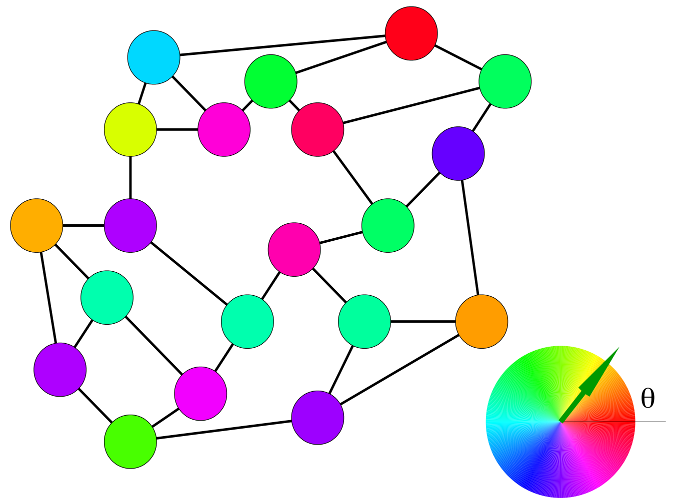
Supervised Learning
artificial intelligence







Continuous Coloring (scheduling)
compute science



Ecology
Micro Economics
Metabolic Networks

- The Space of Solutions is **continuous**, it shrinks to a point at the SAT-UNSAT threshold. Scaling laws appear close to the **Jamming Transition**.