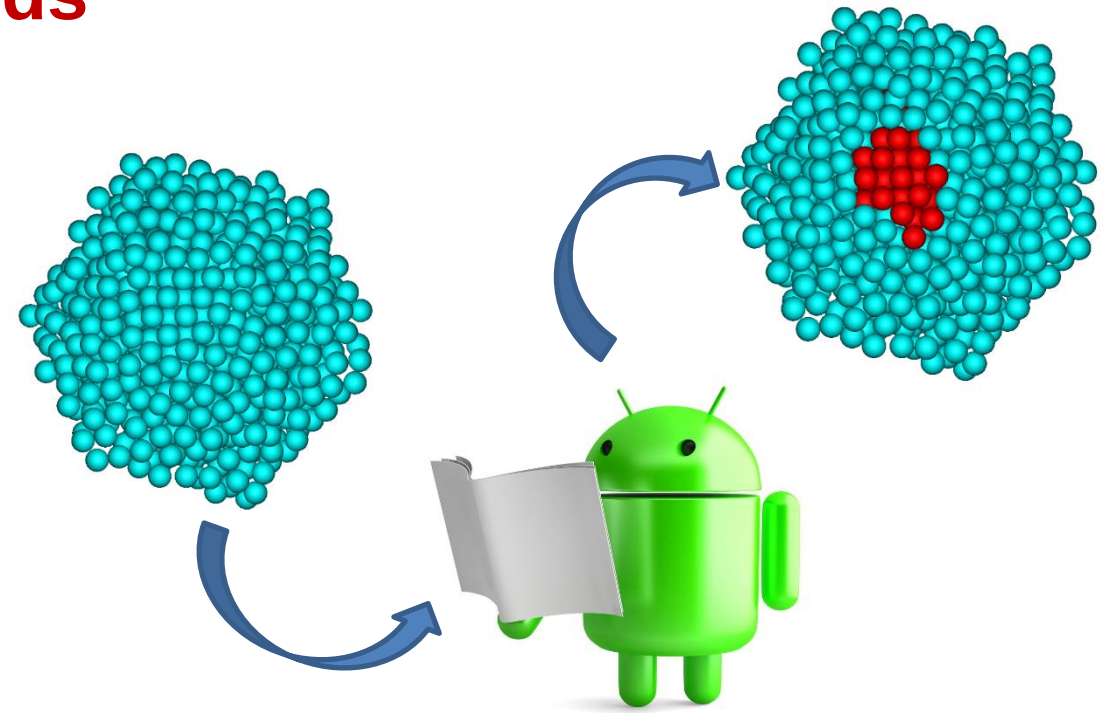
Utrecht University

Debye Institute for Nanomaterials Science

Autonomously revealing hidden local structures in supercooled liquids

Laura Filion

Soft Condensed Matter,
Debye Institute for
Nanomaterials Science,
Utrecht University



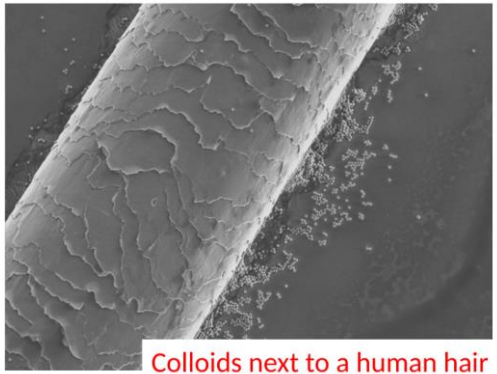
4TU.HTM & M2i Workshop on Artificial Intelligence in Materials Science

Colloidal Systems

Colloids

Nano- to micrometer particles
Suspended in a solvent

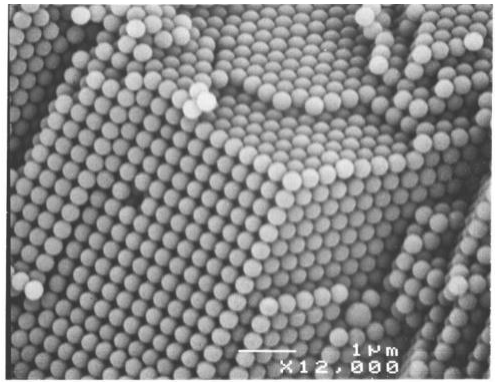
Brownian Motion

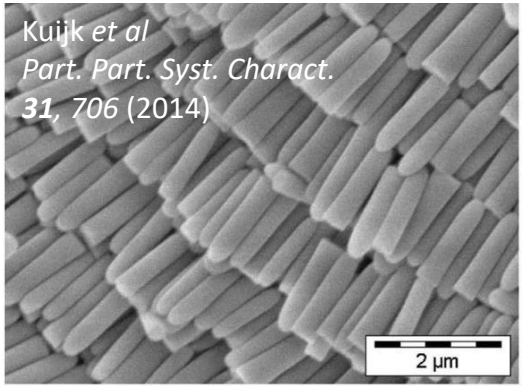


Colloids next to a human hair

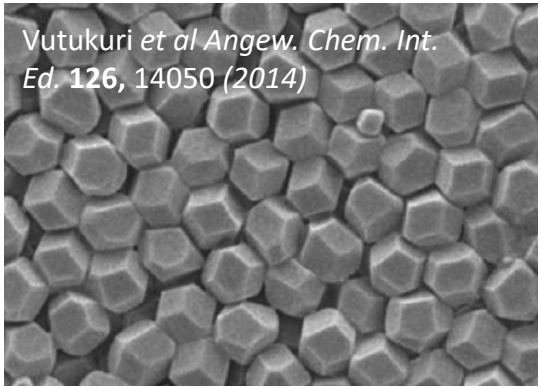
Self Assembly

Due to the Brownian motion, these particles can spontaneously order (self-assemble) into a wide variety of phases, similar to what is observed in atomic systems, e.g. **liquids, crystals, gases**, etc..





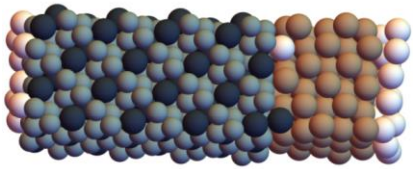
Kuijk *et al*
Part. Part. Syst. Charact.
31, *706* (2014)



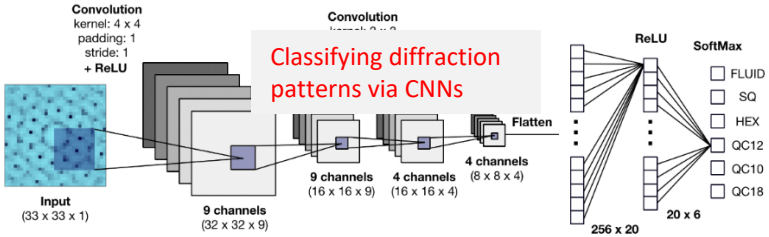
Vutukuri *et al Angew. Chem. Int. Ed.* **126,** 14050 *(2014)*

ML Research

Learning structure and dynamics via **supervised ML**



Neural-network-based order parameters for classification **of binary hard-sphere crystal structures**
E. Boattini, M. Ram, F. Smallenburg, **L. Filion**,
Molecular Physics **116**, 3066-3075 (2018)

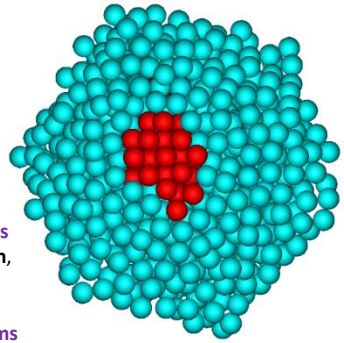


Classifying diffraction patterns via CNNs

Inverse design of soft materials via a deep-learning-based evolutionary strategy
G. Coli, E. Boattini, **L. Filion**, M. Dijkstra, submitted (2021)

Averaging local structure to predict the dynamic propensity in supercooled liquids
E. Boattini, F. Smallenburg, **L. Filion**, arXiv: 2105.05921 (2021)

Autonomously identifying distinctive particle environments via **unsupervised ML**

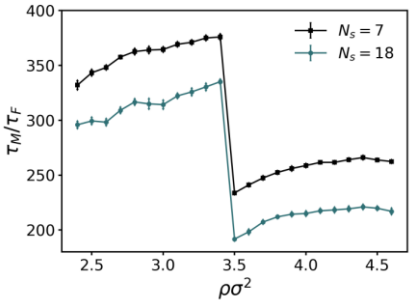


Autonomously revealing hidden local structures in **supercooled liquids**
E. Boattini, S. Marín-Aguilar, S. Mitra, G. Foffi, F. Smallenburg, **L. Filion**,
Nature Communications **11**, 5479 (2020)

Unsupervised learning for local **structure detection** in **colloidal systems**
E. Boattini, M. Dijkstra, **L. Filion**
The Journal of Chemical Physics **151**, 154901 (2019)

Fitting effective potentials with **symmetry functions** (up to 3-body)

Speed up (time for full model vs fitted model)

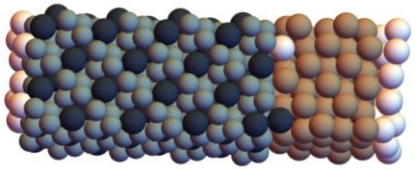


Modeling of many-body interactions between **elastic spheres** through symmetry functions
E. Boattini, N. Bezem, S.N. Punnathanam, F. Smallenburg, **L. Filion**,
The Journal of Chemical Physics **153**, 064902 (2020).

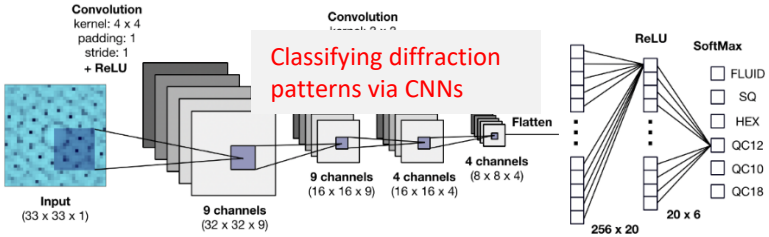
Modeling of effective interactions between ligand coated **nanoparticles** through symmetry functions
D. Chintha, S. Kumar Veesam, E. Boattini, **L. Filion**, S. N. Punnathanam, in preparation

ML Research

Learning structure and dynamics via **supervised ML**



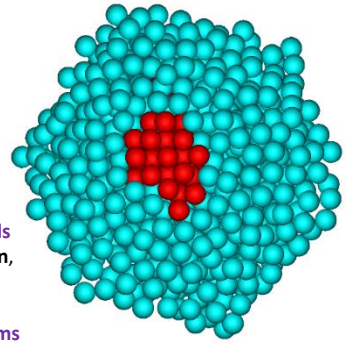
Neural-network-based order parameters for classification **of binary hard-sphere crystal structures**
E. Boattini, M. Ram, F. Smallenburg, **L. Filion**,
Molecular Physics **116**, 3066-3075 (2018)



Inverse design of soft materials via a deep-learning-based evolutionary strategy
G. Coli, E. Boattini, **L. Filion**, M. Dijkstra, submitted

Fitting the dynamics in glasses (3 projects in progress)
E. Boattini, D. Vos, R. Alkemade, F. Smallenburg, **L. Filion**

Autonomously identifying distinctive particle environments via **unsupervised ML**

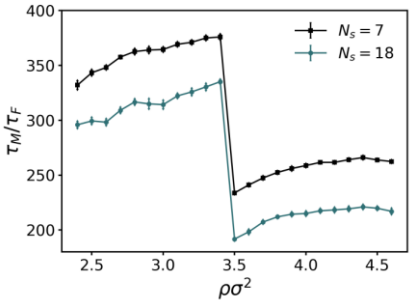


Autonomously revealing hidden local structures in **supercooled liquids**
E. Boattini, S. Marín-Aguilar, S. Mitra, G. Foffi, F. Smallenburg, **L. Filion**,
Nature Communications **11**, 5479 (2020)

Unsupervised learning for local **structure detection** in **colloidal systems**
E. Boattini, M. Dijkstra, **L. Filion**
The Journal of Chemical Physics **151**, 154901 (2019)

Fitting effective potentials with **symmetry functions** (up to 3-body)

Speed up (time for full model vs fitted model)



Modeling of many-body interactions between **elastic spheres** through symmetry functions
E. Boattini, N. Bezem, S.N. Punnathanam, F. Smallenburg, **L. Filion**,
The Journal of Chemical Physics **153**, 064902 (2020).

Modeling of effective interactions between ligand coated **nanoparticles** through symmetry functions
D. Chintha, S. Kumar Veesam, E. Boattini, **L. Filion**, S. N. Punnathanam, in preparation

Emanuele Boattini

Michel Ram

Susanna Marin-Aguila
CNRS, LPS, France

Saheli Mitra
CNRS, LPS, France

Frank Smallenburg
CNRS, LPS, France

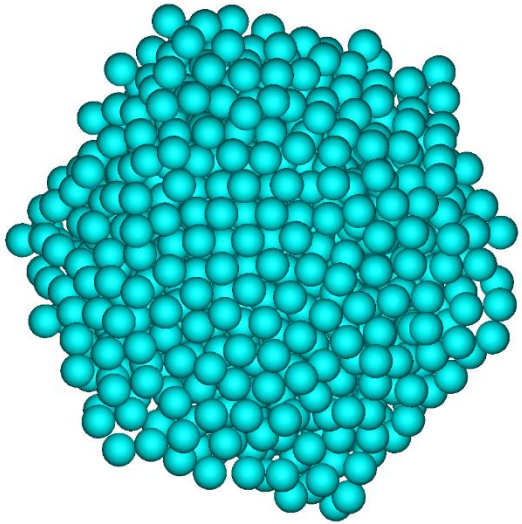
Giuseppe Foffi
CNRS, LPS, France

Marjolein Dijkstra

NWO
Netherlands Organisation
for Scientific Research

Common Problem

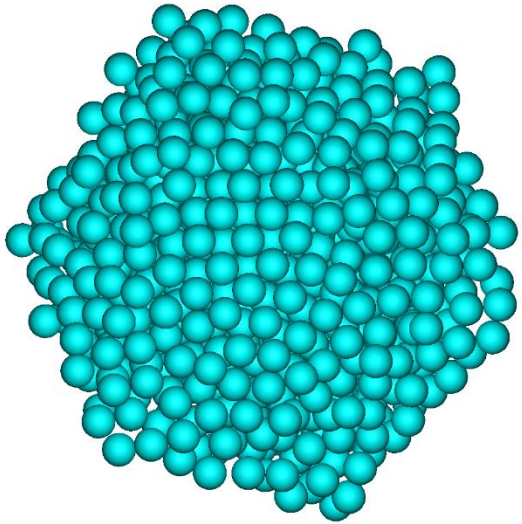
What is the local structure of my particles?

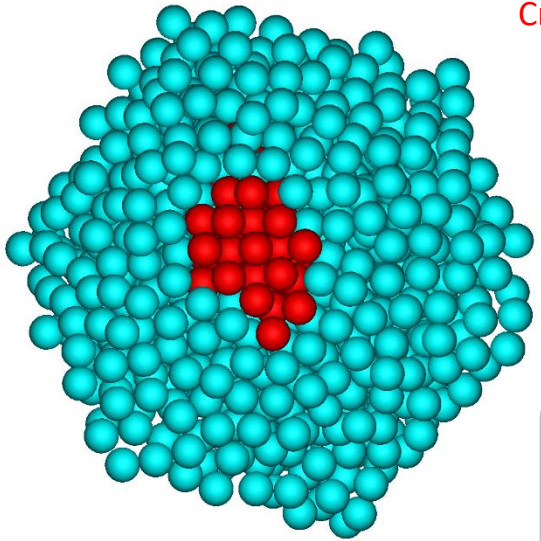


Which particles are crystalline, which are fluid?

Common Problem

What is the local structure of my particles?





Fluid (blue)
Crystal (red)

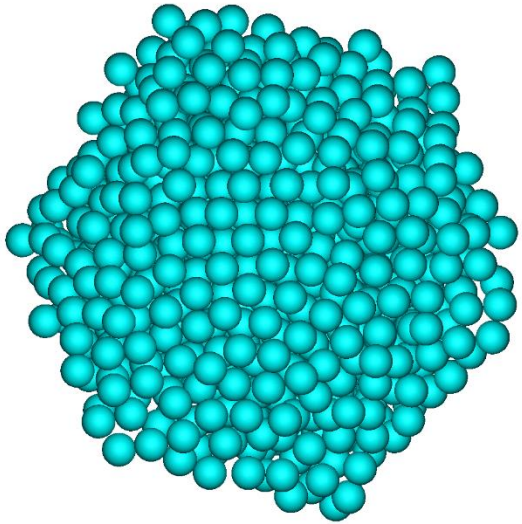
Which particles are crystalline, which are fluid?

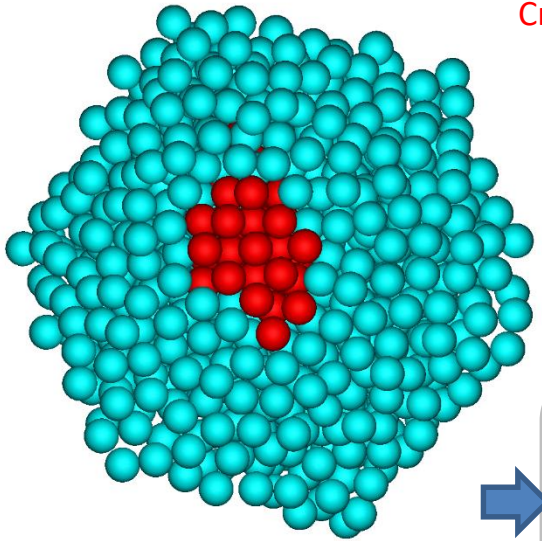
Three options:

- Human design
- Supervised ML
- Unsupervised ML

Common Problem

What is the local structure of my particles?





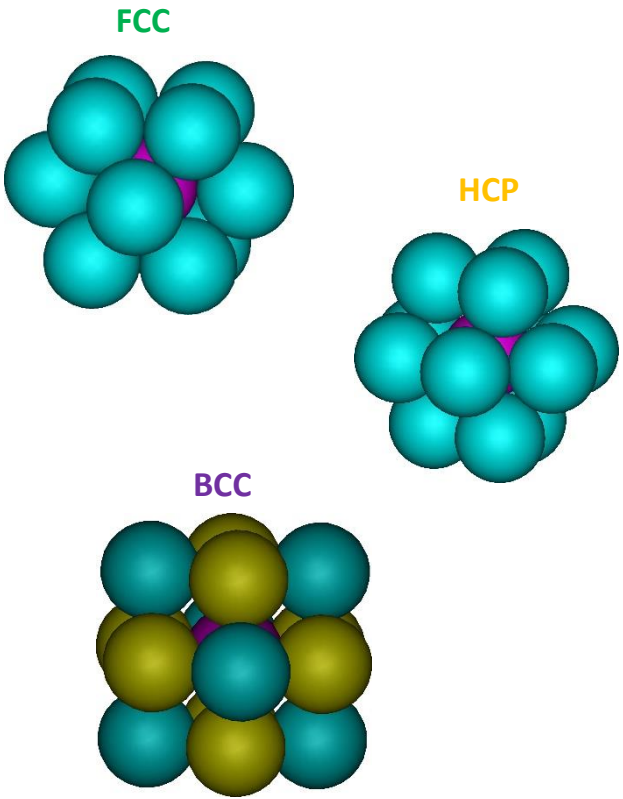
Fluid (blue)
Crystal (red)

Which particles are crystalline, which are fluid?

Three options:

- **Human design**
- Supervised ML
- Unsupervised ML

Local Ordered Environments

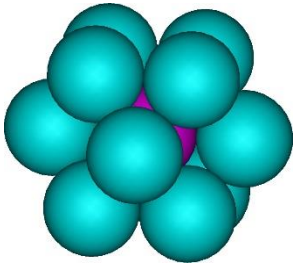


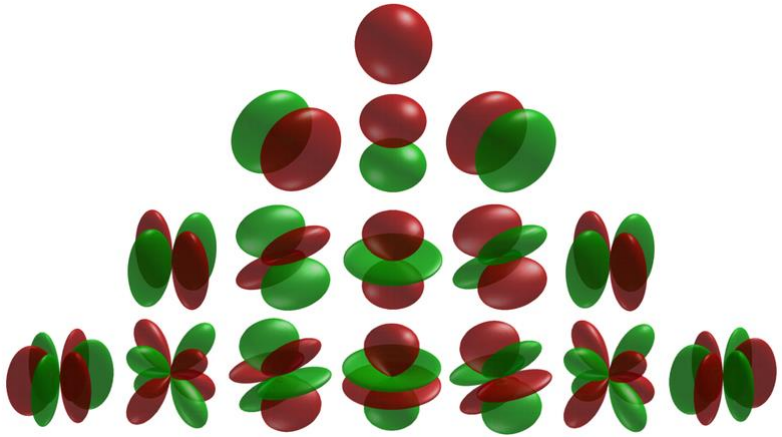
Goal: develop an algorithm that can distinguish on a single-particle basis which environment a particle is in (i.e. FCC, HCP, BCC, disordered)

Local Environment Descriptions:

- **Bond order parameters (BOP)**
- Common neighbour analysis (CNA)
- Templating
- ....

Bond order parameters





Sarxos, Wikimedia Commons (2007)

Want:

1) Description of local environment (density) that captures the symmetry

2) Rotationally invariant

Expand the local density in terms of ***spherical harmonics***

Develop ***rotationally invariant*** quantities from the spherical harmonics that depend **only on *l*:** q_l, w_l

Bond order parameters

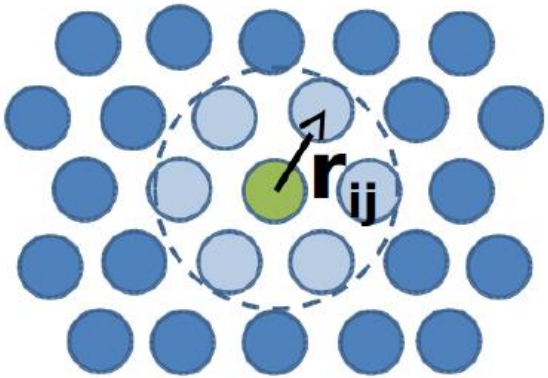
$$q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_l^m(\mathbf{r_{ij}})$$

Rotationally invariant quantities

$$q_l(i) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |q_{lm}(i)|^2}$$

Wigner 3j

$$w_l(i) = \frac{\sum_{m_1+m_2+m_3=0} \begin{pmatrix} l & l & l \\ m_1 & m_2 & m_3 \end{pmatrix} q_{lm_1}(i) q_{lm_2}(i) q_{lm_3}(i)}{\left(\sum_{m=-l}^{l} |q_{lm}(i)|^2\right)^{3/2}}$$

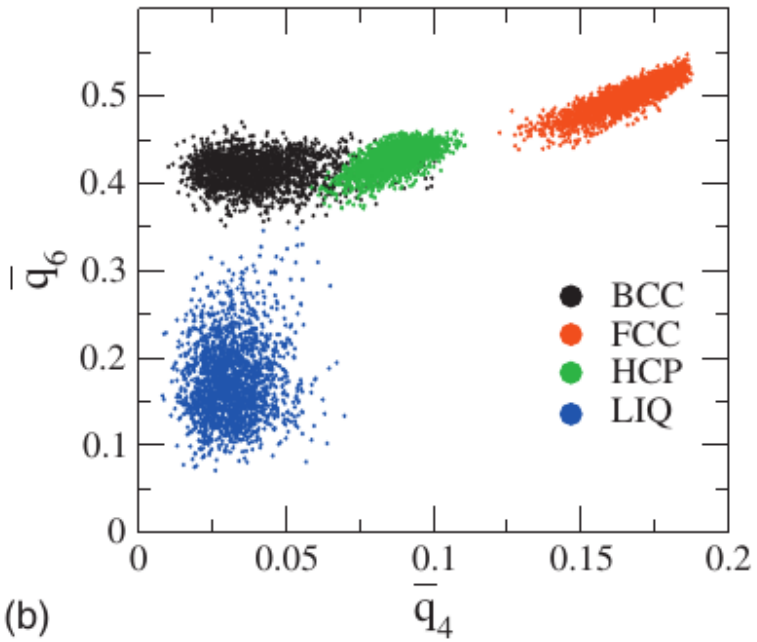


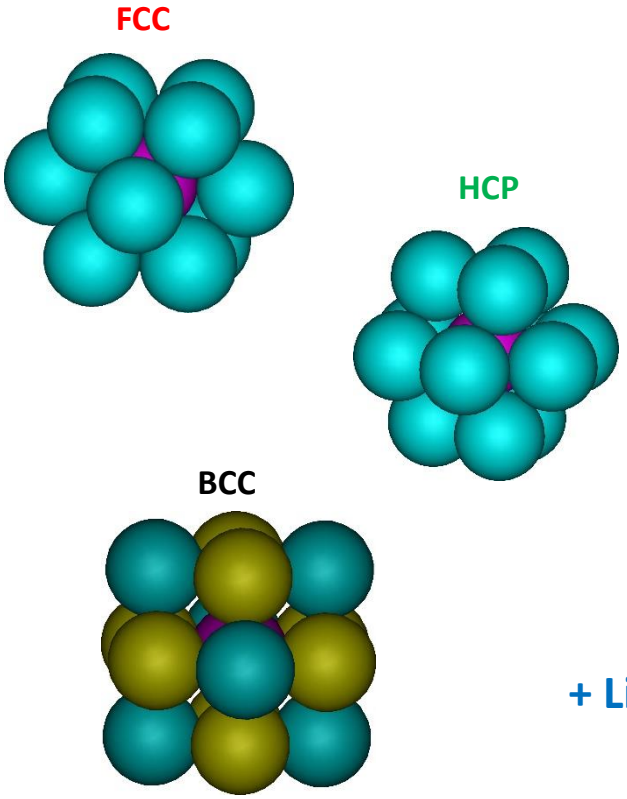
Local averages work better (*)

$$\bar{q}_l(i) = \frac{1}{N_b(i)+1} \left[q_l(i) + \sum_{k \in \mathcal{N}_b(i)} q_l(k) \right]$$

Steinhardt, Nelson, Ronchetti, Phys. Rev. B **28**, 784 (1983)
Lechner, Dellago, J. Chem. Phys **129**,114707 (2008) (*)

Result

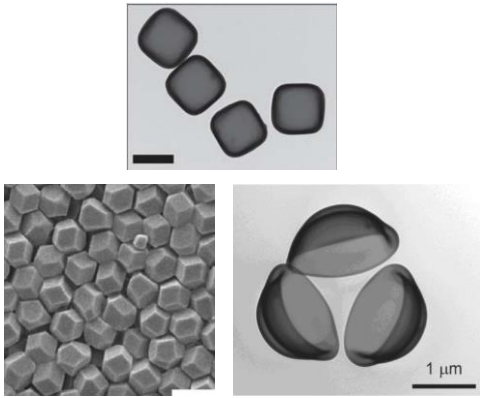




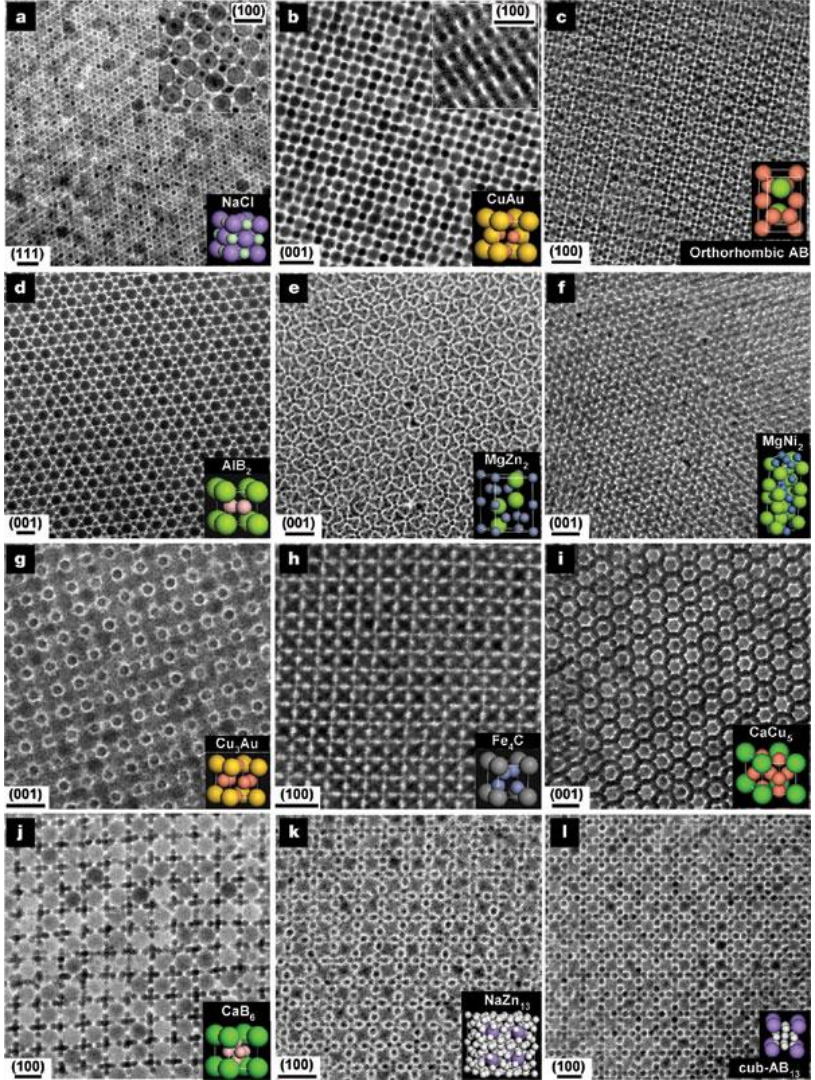
All four clearly identifiable!

Lechner, Dellago, J. Chem Phys. **129**, 114707 (2008)

Extension?



In principle possible, but each time must determine (by **trial and error**) the "important" rotationally invariant quantities



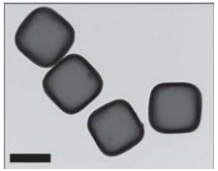
Binary:
Shevchenko *et al.*, *Nature* **439**, 55 (2006)

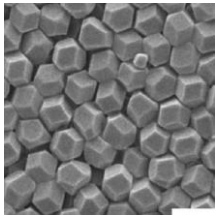
Cubes:
Rossi *et al.*
Soft Matter **7**, 4139 *(2011)*

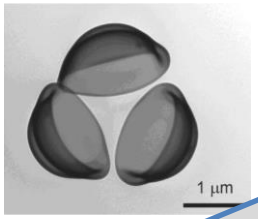
Polyhedra:
Vutukuri et al., Angew. Chem. Int. Ed. **126,** 14050 *(2014)*

Bowls:
Utrecht SCM group

Extension?

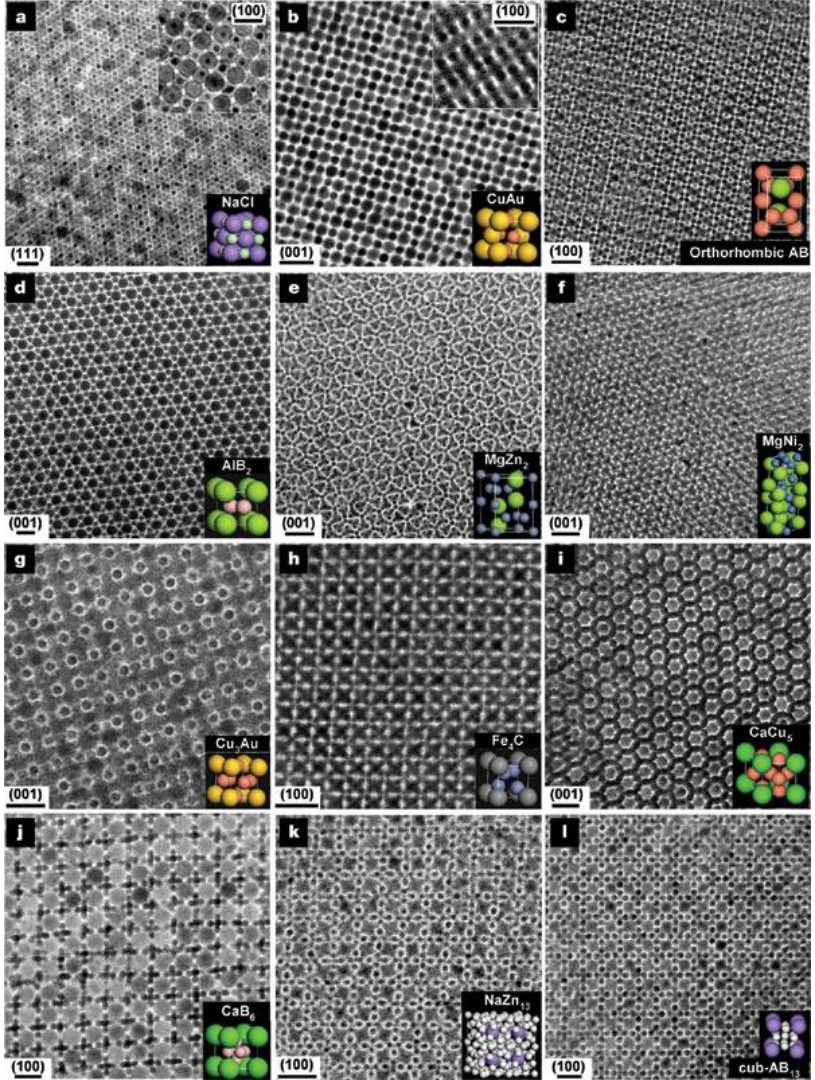






In principle possi
time mus
ar
quantities

Question: Is there an ML solution?



Binary:
Shevchenko *et al.*, *Nature* **439**, 55 (2006)

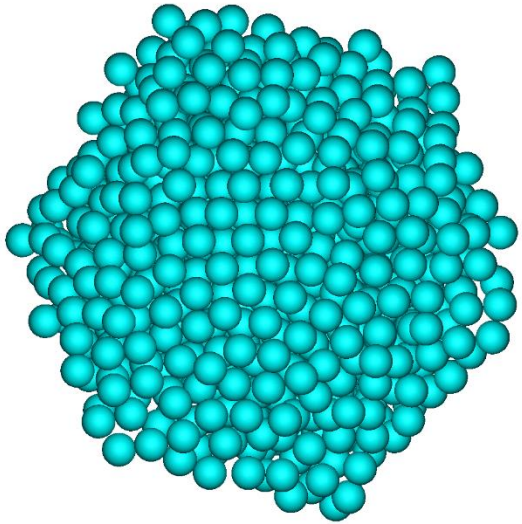
Cubes:
Rossi *et al.*
Soft Matter **7**, 4139 *(2011)*

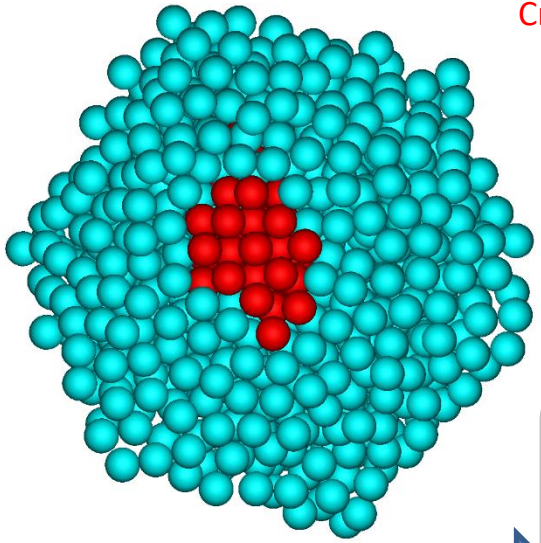
Polyhedra:
Vutukuri et al., Angew. Chem. Int. Ed. **126,** 14050 *(2014)*

Bowls:
Utrecht SCM group

Common Problem

What is the local structure of my particles?





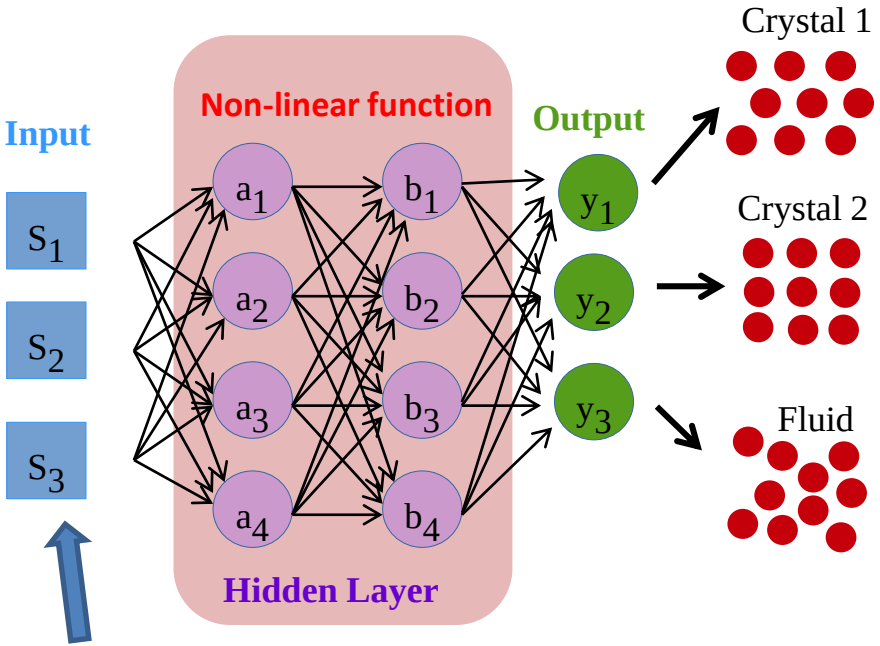
Fluid (blue)
Crystal (red)

Which particles are crystalline, which are fluid?

Three options:

- Human design
- **Supervised ML**
- Unsupervised ML

Neural Network Order Parameter



Description of the local environment

$$\{\mathbf{Q}(i) = (\bar{q}_1(i), \bar{q}_2(i), \bar{q}_3(i), \ldots, \bar{q}_d(i))\}$$

1) Collect **training data** from **simulations**:
   Simulate **known structures**, and collect e.g. q_l and w_l for many particles, at different pressures, temperatures, etc.

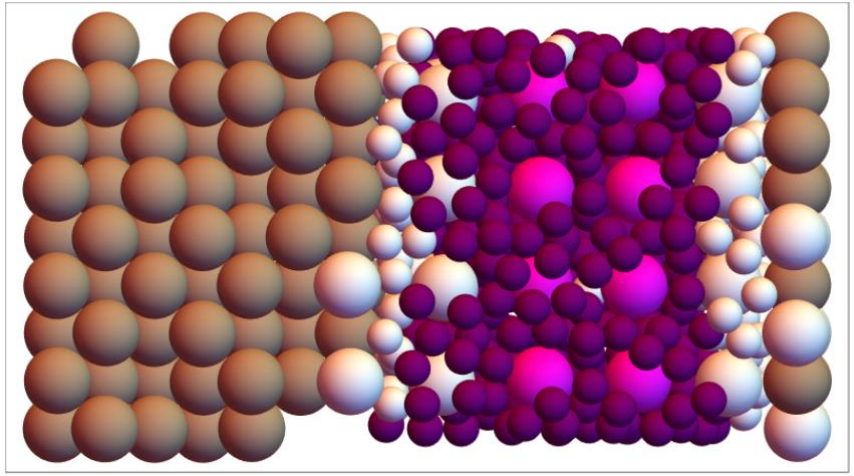
2) Optimize **(fit)** nonlinear function **(neural network)** until output of neural network matches true structures.

3) **Test** trained network on an independent dataset.

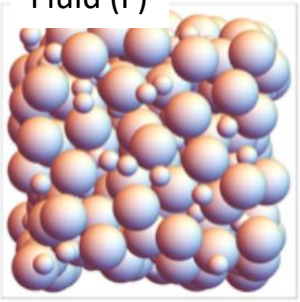
Binary hard spheres (a=0.58)

Accuracy > 99% for all 6 environments

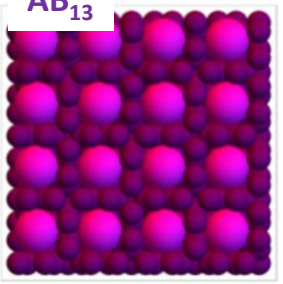
Example 1: FCC-AB_{13} coexistence



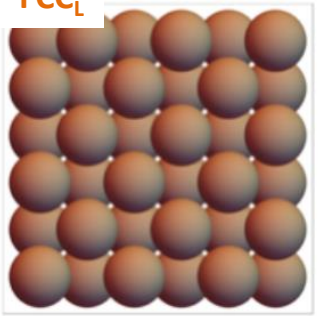
Fluid (F)



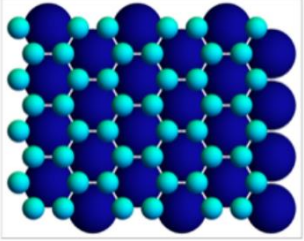
AB_{13}



FCC_L



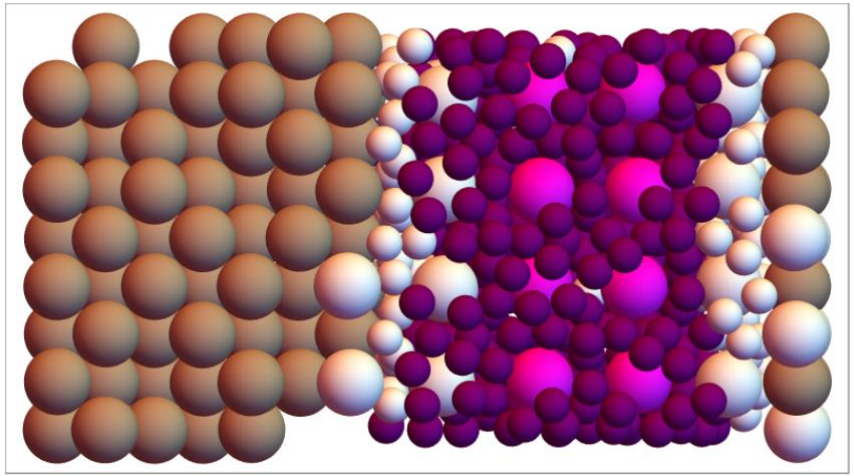
AB_2



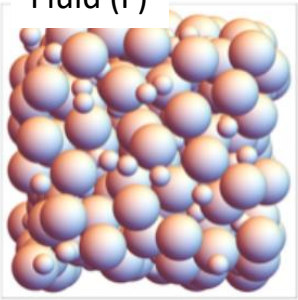
Binary hard spheres (a=0.58)

Accuracy > 99% for all 6 environments

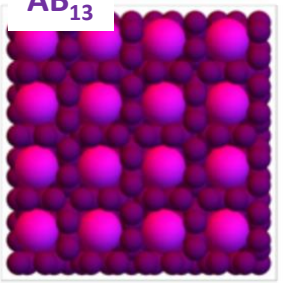
Example 1: FCC-AB_{13} coexistence



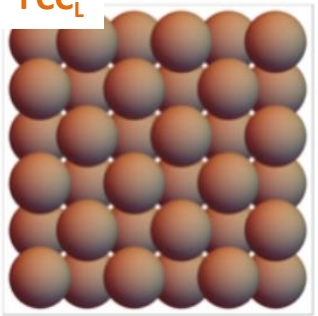
Fluid (F)



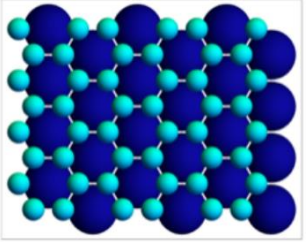
AB_{13}



FCC_L



AB_2



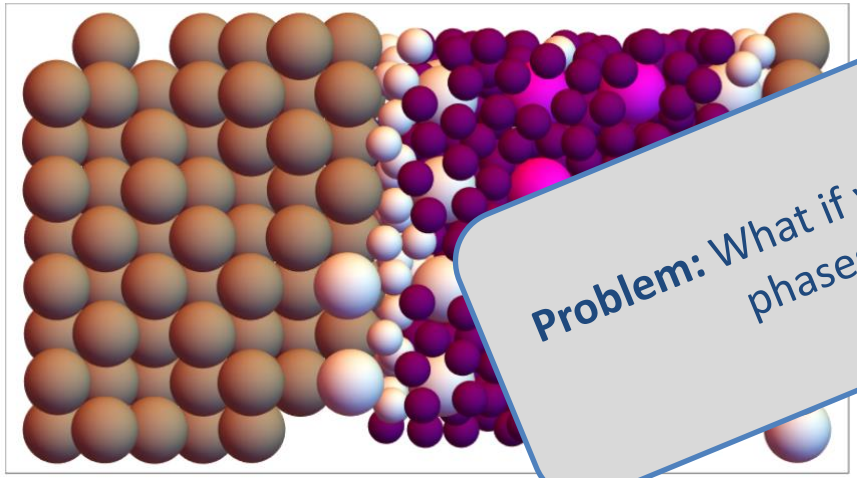
Works similarly for all known binary hard-sphere crystals (~99% accuracy)

E. Boattini, M. Ram, F. Smallenburg, and L. Filion, Mol. Phys. **116**, 3066 (2018)

Binary hard spheres (a=0.58)

Accuracy > 99% for all 6 environments

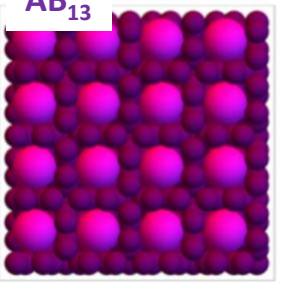
Example 1: FCC-AB_{13} coexistence



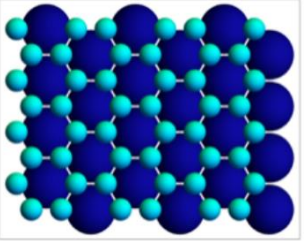
Fluid (F)

Problem: What if you do not know what phases to look for?

AB_{13}



AB_2

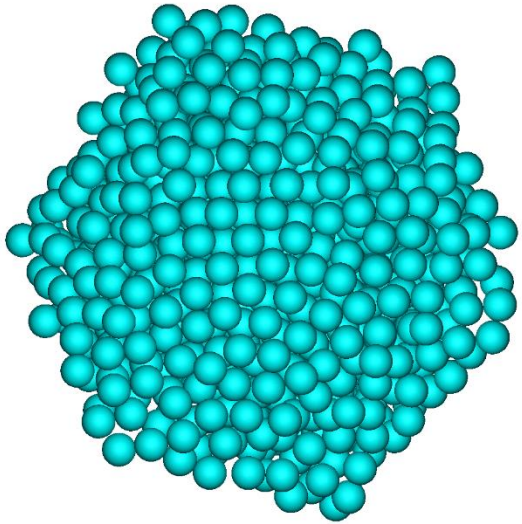


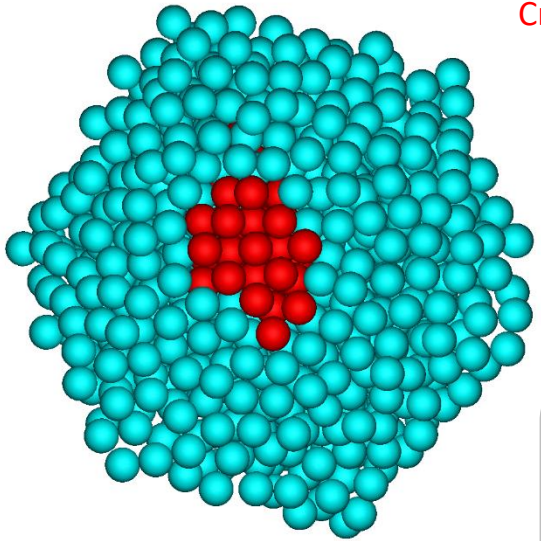
Works similarly for all known binary hard-sphere crystals (~99% accuracy)

E. Boattini, M. Ram, F. Smallenburg, and L. Filion, Mol. Phys. **116**, 3066 (2018)

Common Problem

What is the local structure of my particles?





Fluid (blue)
Crystal (red)

Which particles are crystalline, which are fluid?

Three options:

- Human design
- Supervised ML
- **Unsupervised ML**

Our Goal

Design a **simple** and **fast** algorithm to:

- **autonomously identify** the **distinct local environments** present in the system without prior knowledge of the possibilities.
- **robust** (works for many different systems)

Our Goal

Design a **simple** and **fast** algorithm to:

- **autonomously identify** the **distinct local environments** present in the system without prior knowledge of the possibilities.
- **robust** (works for many different systems).

What we need

1. A general **description of local environments** useful for distinguishing **liquid** and **crystal**

2. A way of **understanding the main features distinguishing** the **local environments** in the system.

3. A way of **grouping together similar environments.**

Our Goal

Design a **simple** and **fast** algorithm to:

- **autonomously identify** the **distinct local environments** present in the system without prior knowledge of the possibilities.
- **robust** (works for many different systems)

What we need

✓ 1. A general **description of local environments** useful for distinguishing **liquid** and **crystal** (**bond order parameters – vector q_l from earlier**).

2. A way of **understanding the main features distinguishing** the **local environments** in the system.

3. A way of **grouping together similar environments**.

Our Goal

Design a **simple** and **fast** algorithm to:

- **autonomously identify** the **distinct local environments** present in the system without prior knowledge of the possibilities.
- **robust** (works for many different systems)

What we need

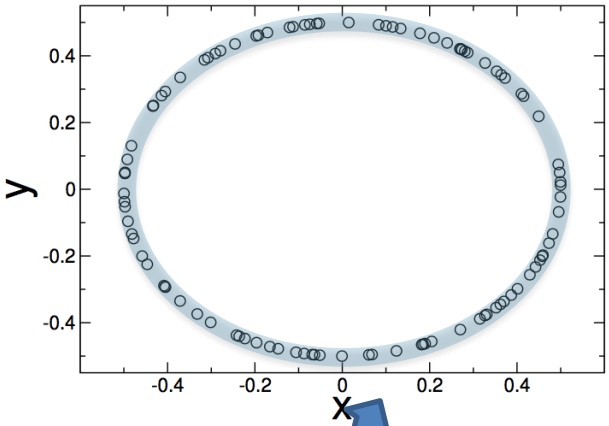
✓ 1. A general **description of local environments** useful for distinguishing **liquid** and **crystal** (**bond order parameters – vector q_l from earlier**).

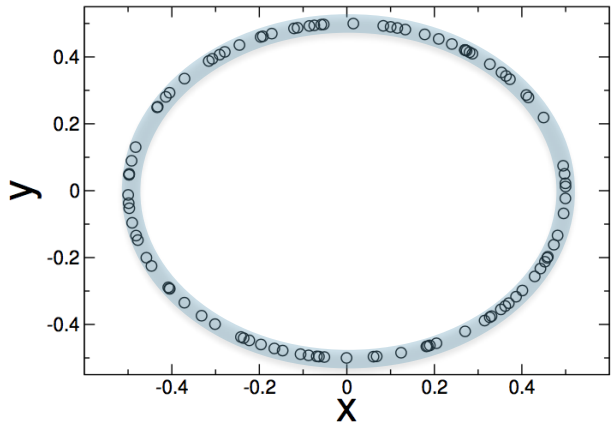
2. A way of **understanding the main features distinguishing** the **local environments** in the system (**dimensionality reduction**).

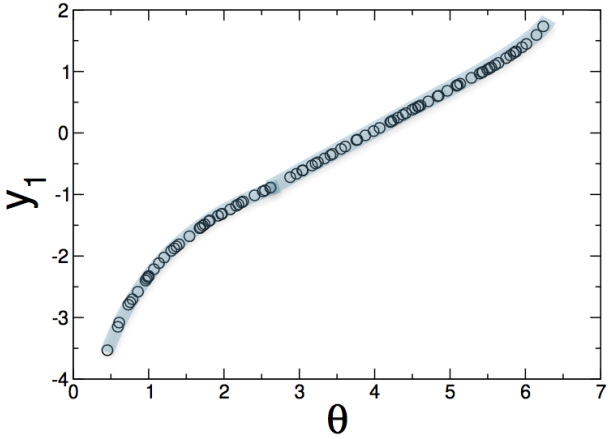
3. A way of **grouping together similar environments** .

Non-linear Dimensionality Reduction:

Nonlinear **can** find a 1d projection of original data



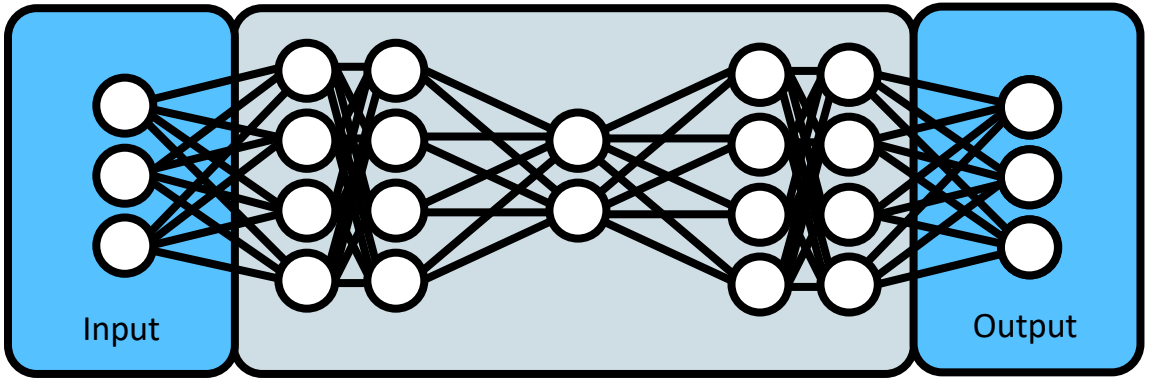




Linear cannot find reduce dimensionality

Dimensionality Reduction: Neural-Network-based Autoencoder (AE)

Unsupervised learning technique for **nonlinear dimensionality reduction**.

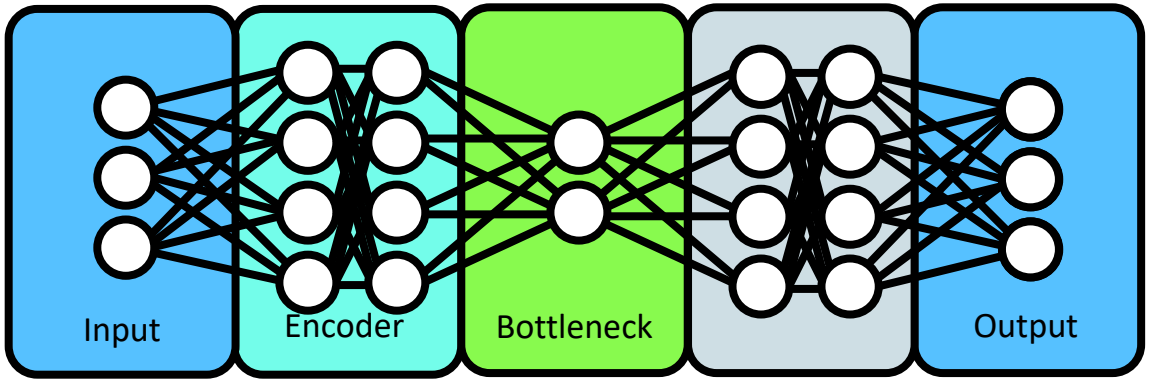


$\mathbf{Q}(i) \in \mathbb{R}^d$ — **Input copy** → $\hat{\mathbf{Q}}(i) \in \mathbb{R}^d$

The AE is **trained** in order to **reproduce** as **accurately** as possible the **input vectors as its output**.

Dimensionality Reduction: Neural-Network-based Autoencoder (AE)

Unsupervised learning technique for **nonlinear dimensionality reduction**.

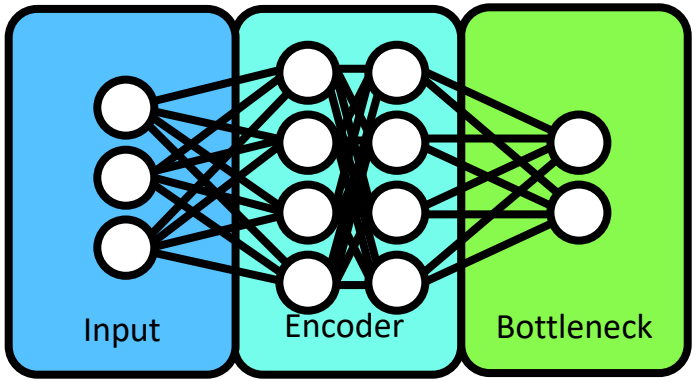


$\mathbf{Q}(i) \in \mathbb{R}^d$ —Projection→ $\mathbf{Y}(i) \in \mathbb{R}^c$, $c < d!$ (Dimension reduction) → $\hat{\mathbf{Q}}(i) \in \mathbb{R}^d$

First, the **encoder** part of the AE performs a **nonlinear projection** of the **original input** onto a **smaller dimensional space** (bottleneck).

Dimensionality Reduction: Neural-Network-based Autoencoder (AE)

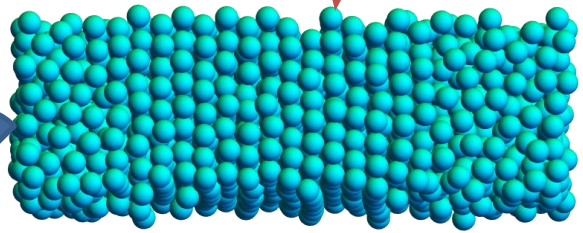
Unsupervised learning technique for **nonlinear dimensionality reduction**.



$$\mathbf{Q}(i) \in \mathbb{R}^d \xrightarrow{\text{Projection}} \mathbf{Y}(i) \in \mathbb{R}^c \quad c < d!$$

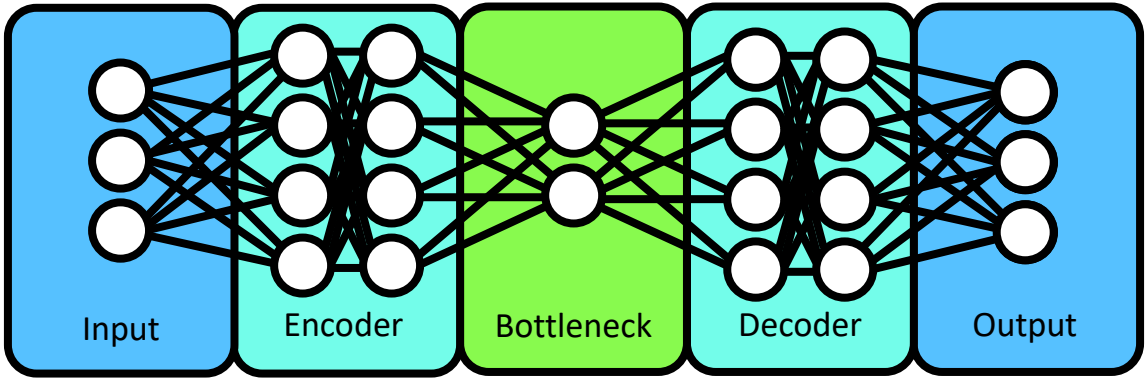
Once trained, **only the encoder part is retained** in order to project the original input vectors onto the low-dimensional space defined by the bottleneck.

Coexistence between the **fluid**, **hexagonal-closed packed (HCP)** and **face-centered cubic (FCC)** phases, in a hard-sphere system.

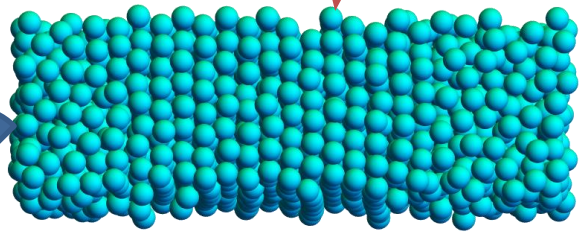


Send in each particle's environment

Optimize NN until agrees well with particle environments

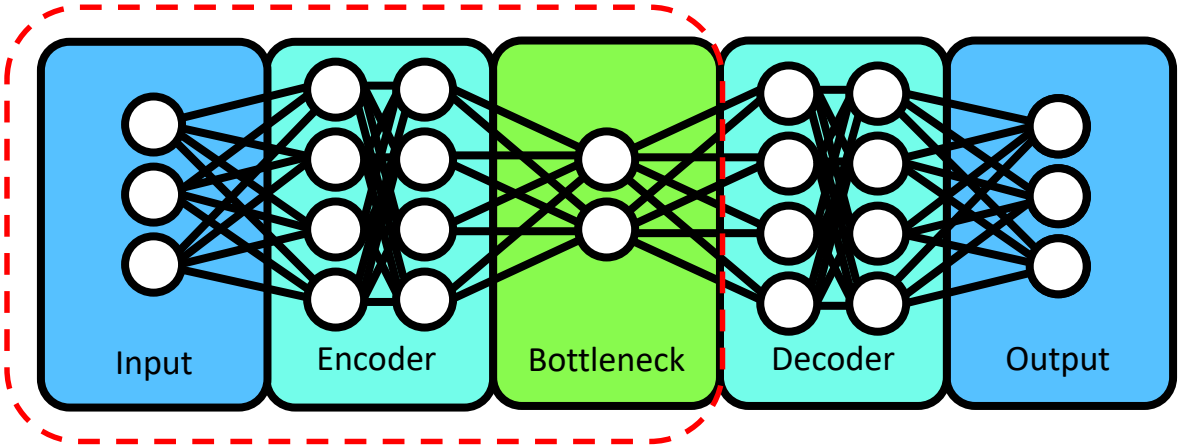


Coexistence between the **fluid**, **hexagonal-closed packed (HCP)** and **face-centered cubic (FCC)** phases, in a hard-sphere system.

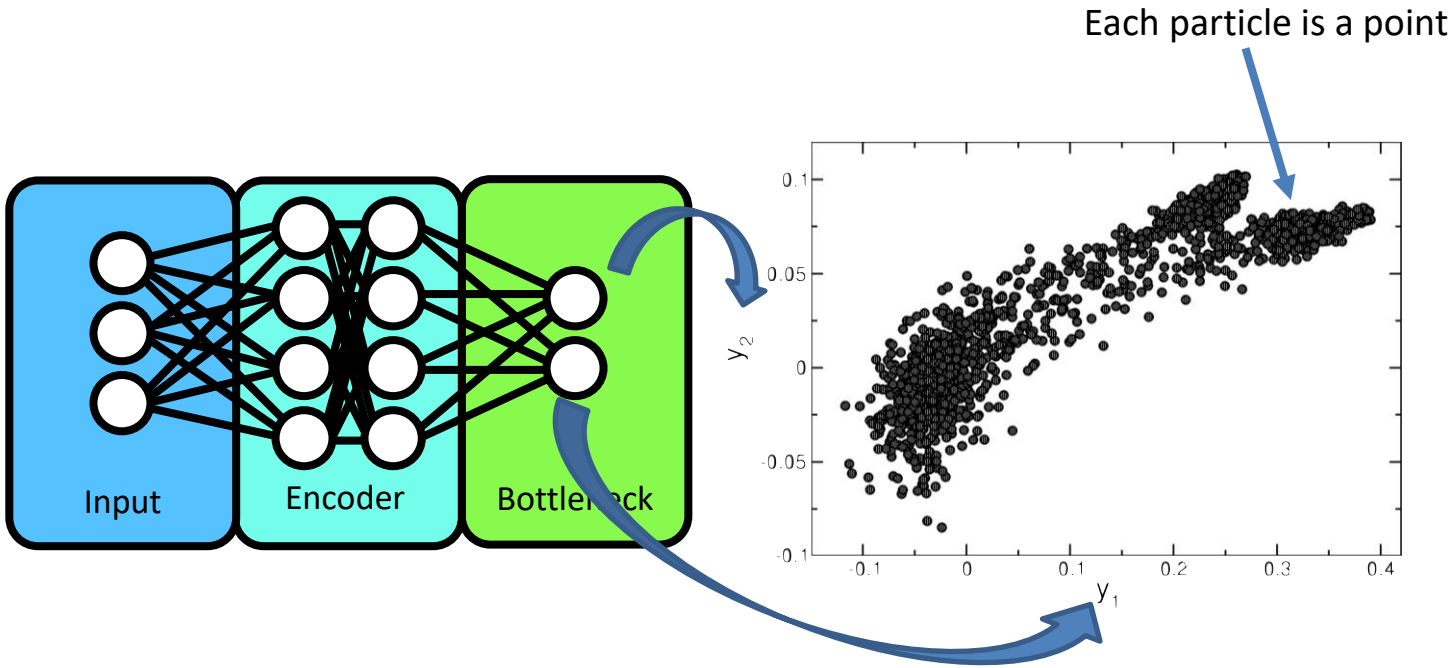


Send in each particle's environment

Optimize NN until agrees well with particle environments



Neural-Network-based Autoencoder (AE)



2 dimensional projection found by the encoder

Multiple ways to determine "**best**" number of dimensions

Our Goal

Design a **simple** and **fast** algorithm to:

- **autonomously identify** the **distinct local environments** present in the system without prior knowledge of the possibilities.
- **robust** (works for many different systems)

What we need

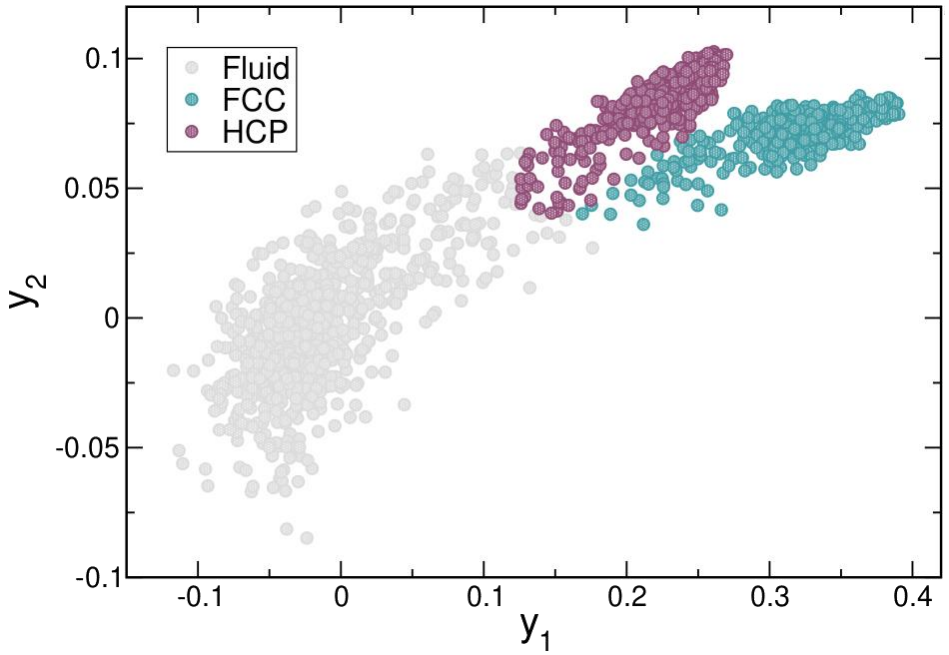
✓ 1. A general **description of local environments** (bond order parameters – vector q_l from earlier)**.**

✓ 2. A way of **understanding the main features distinguishing** the **local environments** in the system (dimensionality reduction).

3. A way of **grouping together similar environments** (**clustering**).

Clustering via Gaussian Mixture Models (GMMs)

Results of the clustering

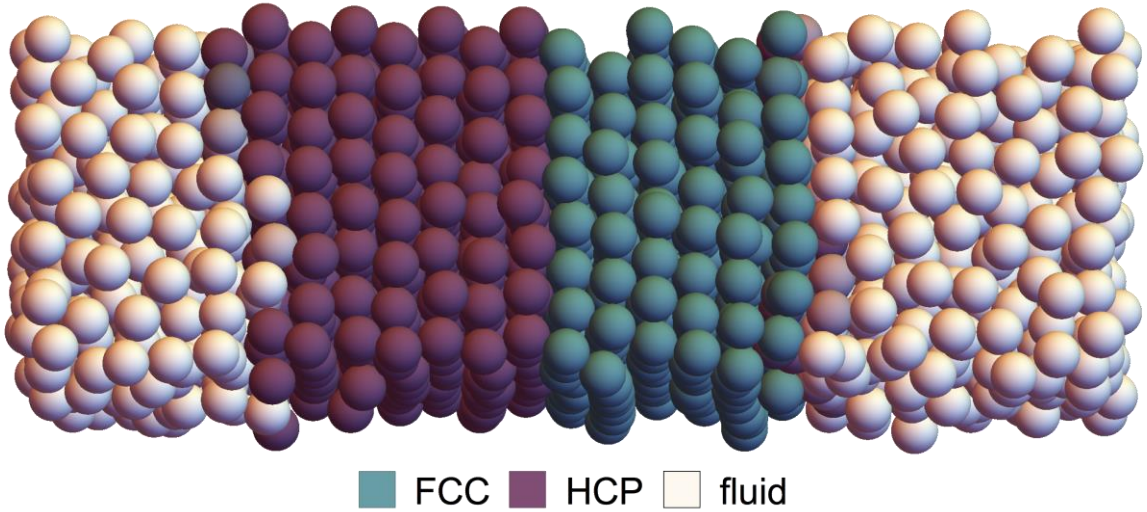


Fits the density of the data with Gaussians, & cluster the Gaussians

The clustering **autonomously identifies 3 distinct environments** corresponding to the **fluid, FCC and HCP** phases

Clustering via Gaussian Mixture Models (GMMs)

Final Classification



E. Boattini, M. Dijkstra, L. Filion, J. Chem. Phys. **151**, 154901 (2019)

Our Goal

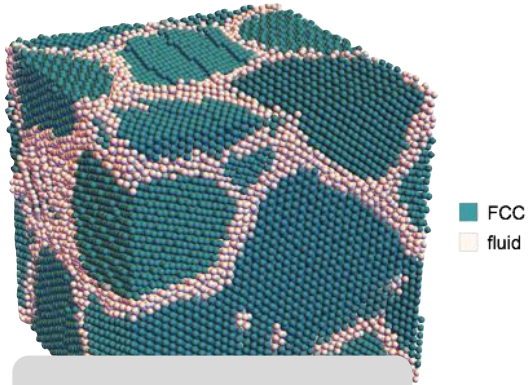
Design a **simple** and **fast** algorithm to:

- ✓ **autonomously identify** the **distinct local environments** present in the system without prior knowledge of the possibilities.
- **robust** (works for many different systems)

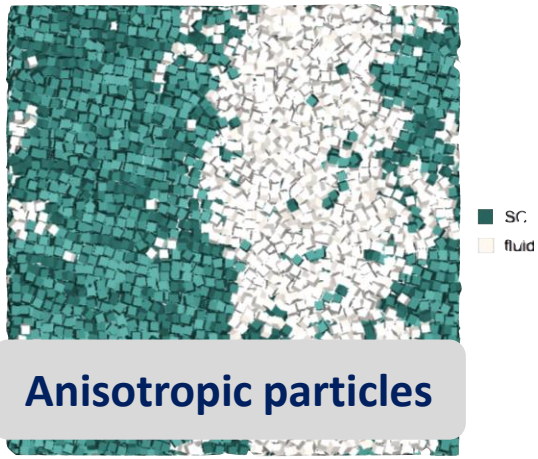
What we need

1. ✓ A general **description of local environments** (bond order parameters – vector q_l from earlier)**.**
2. ✓ A way of **understanding the main features distinguishing** the **local environments** in the system (dimensionality reduction).
3. ✓ A way of **grouping together similar environments** (clustering).

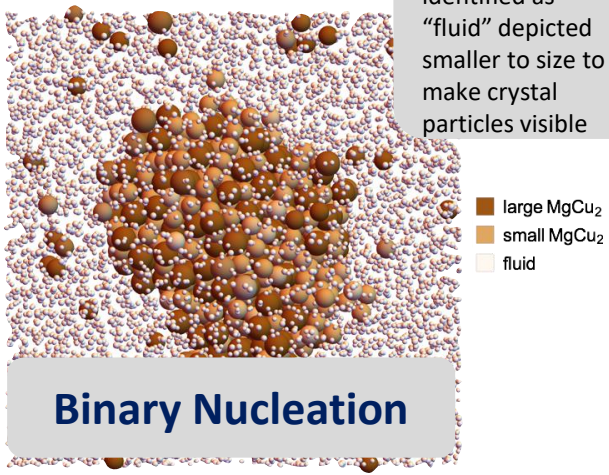
Additional applications



Grain boundaries

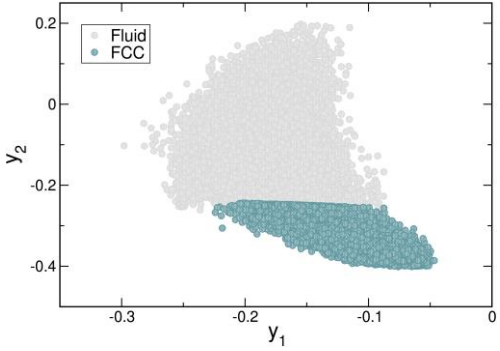


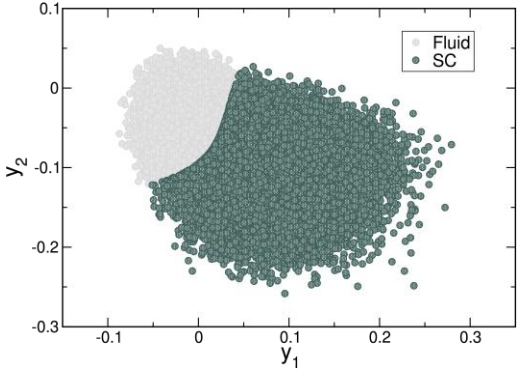
Anisotropic particles

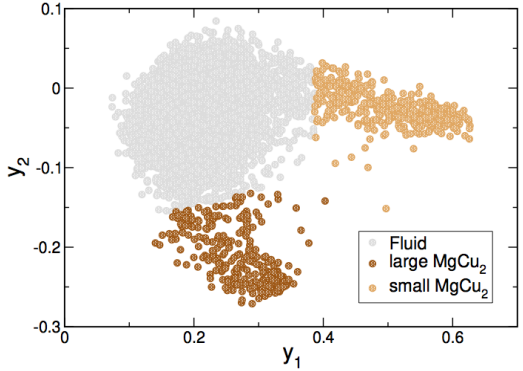


Binary Nucleation

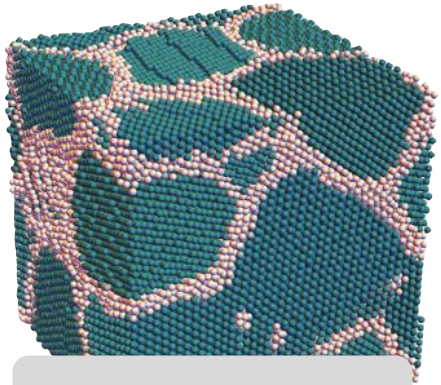
Note: particles identified as "fluid" depicted smaller to size to make crystal particles visible



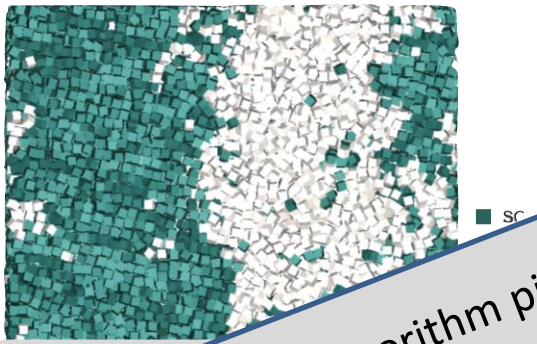




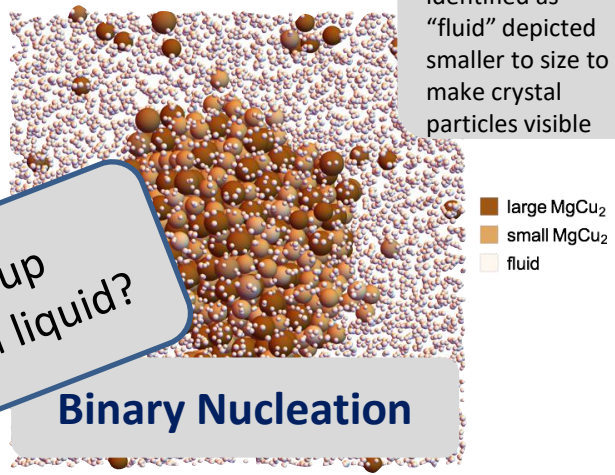
Additional applications



Grain boundaries



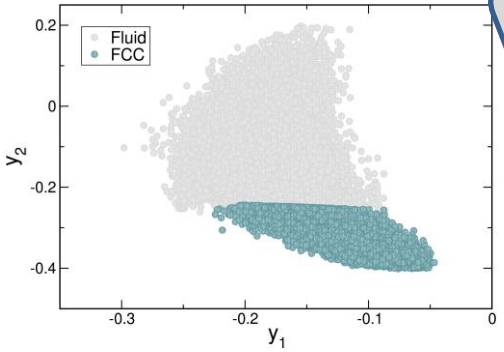
Api

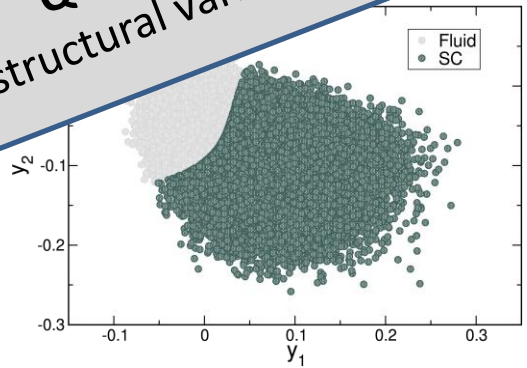


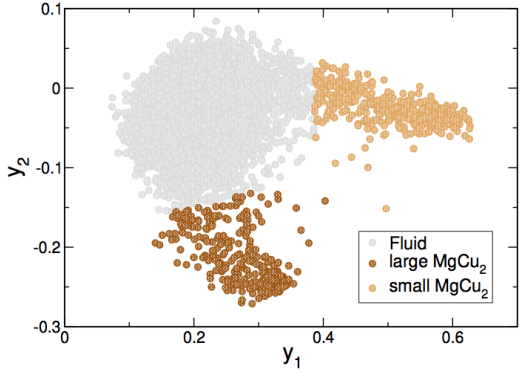
Binary Nucleation

Note: particles identified as "fluid" depicted smaller to size to make crystal particles visible

Question: Can this algorithm pick up structural variations in a supercooled liquid?



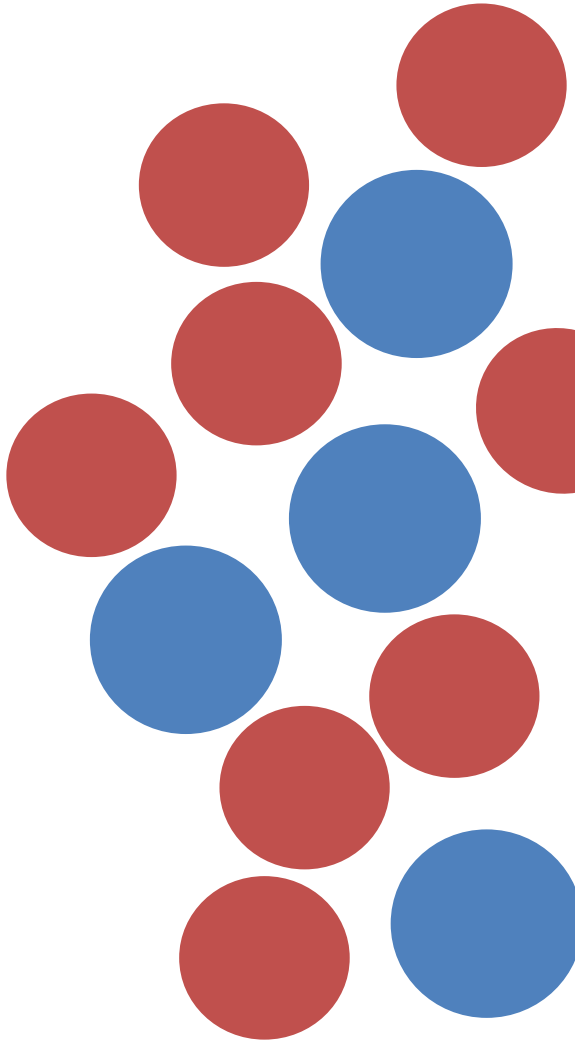




Structure in disorder: a glass?

Three glass formers:

- **Binary hard spheres (HS)**
  - Size ratio: 0.85
  - Composition: 30% large, 70% small
- **Wahnström (W)**
  - Binary mixture of repulsive Lennard-Jones particles
  - Size ratio: 0.833
  - Composition: 50% large, 50% small
- **Kob-Andersen (KA)**
  - Non-additive binary mixture of Lennard-Jones particles
  - $\sigma_{BB} = 0.88\ \sigma_{AA}$, $\sigma_{AB} = 0.8\ \sigma_{AA}$, $\epsilon_{BB} = 0.5\ \epsilon_{AA}$, and $\epsilon_{AB} = 1.5\ \epsilon_{AA}$.
  - Composition: 80% large (A), 20% small (B)

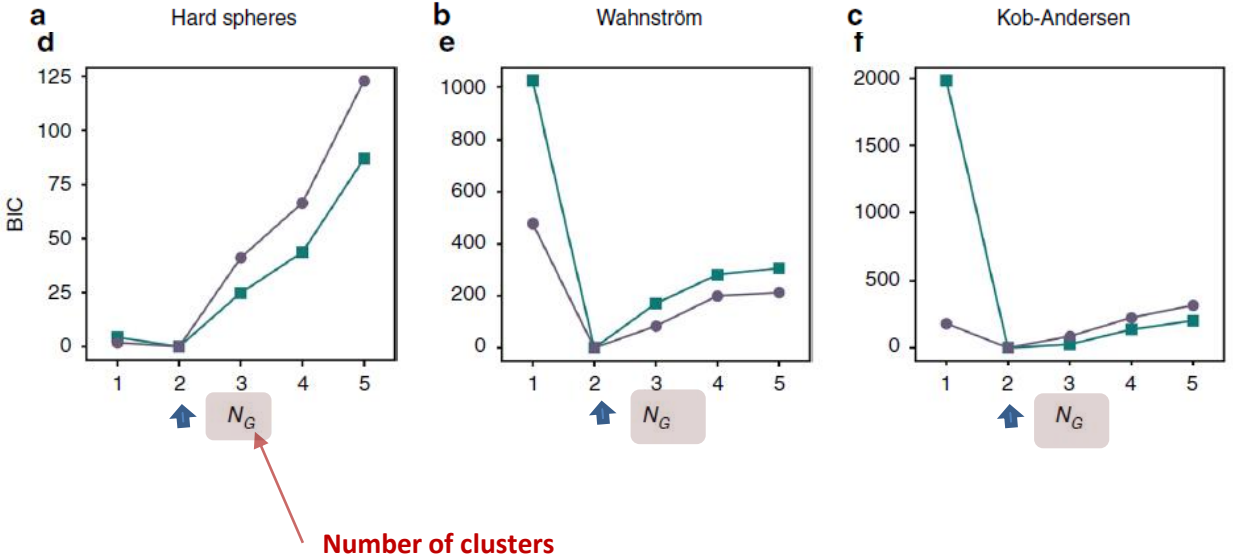


Algorithm Analysis:

- dimensionality for autoencoder:
  HS (2), W(2), KA(4)

- Algorithm identifies **two clusters** of local environments in **all models**
  (min in BIC always at 2)

Bayesian Information Criterion (BIC)



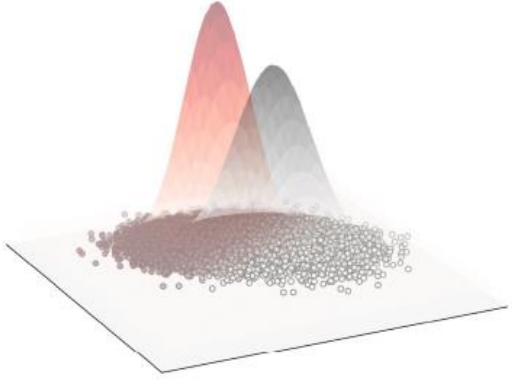
One dimensional order parameter:

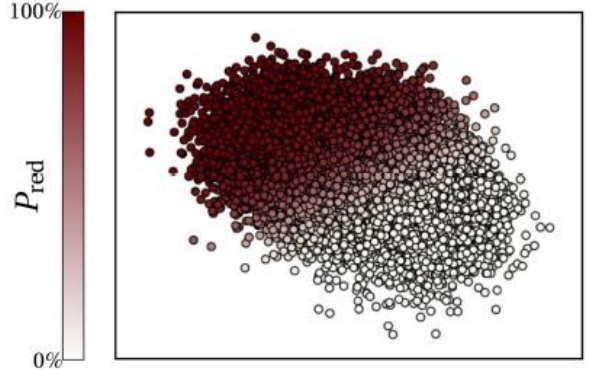
Since there are two clusters, we defined a scalar order parameter

$$P_{\text{red}} = \frac{g_{\text{red}}}{g_{\text{white}} + g_{\text{red}}}$$

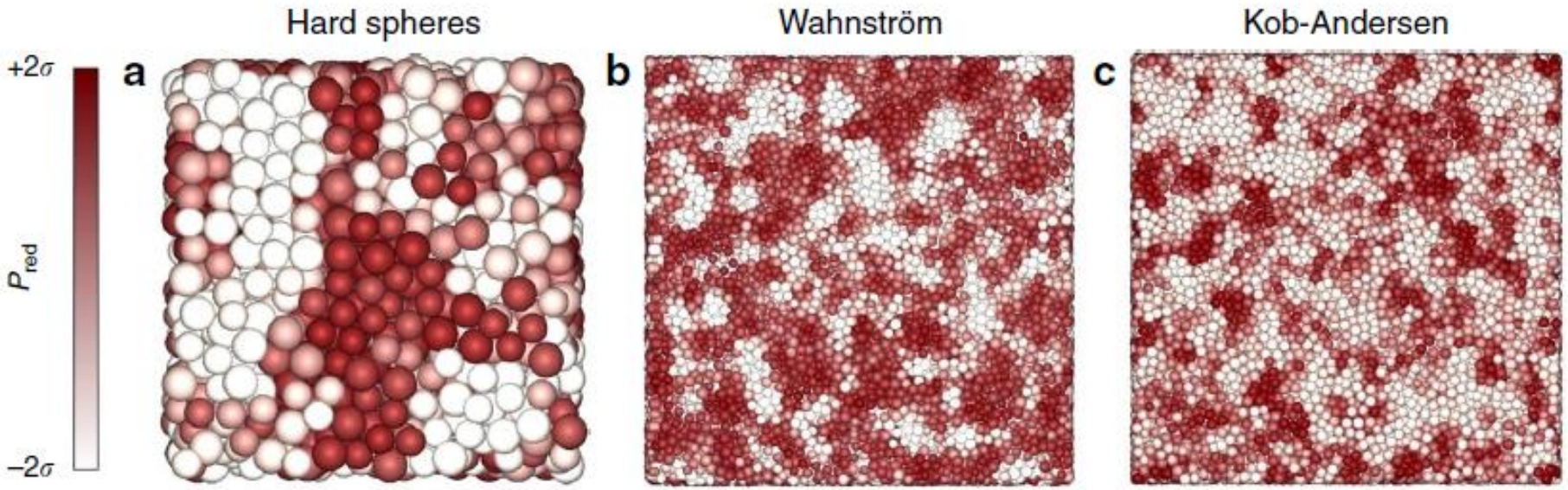
$P_{\text{red}} \in [0,1]$

where $g_{\text{red(white)}}$ are the values of the fitted Gaussians

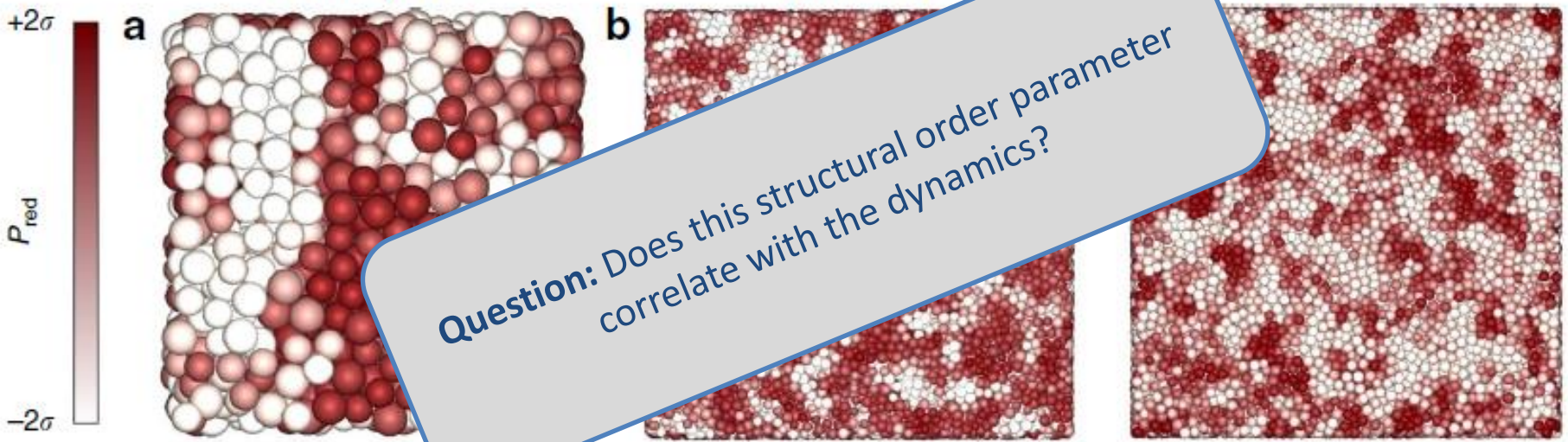




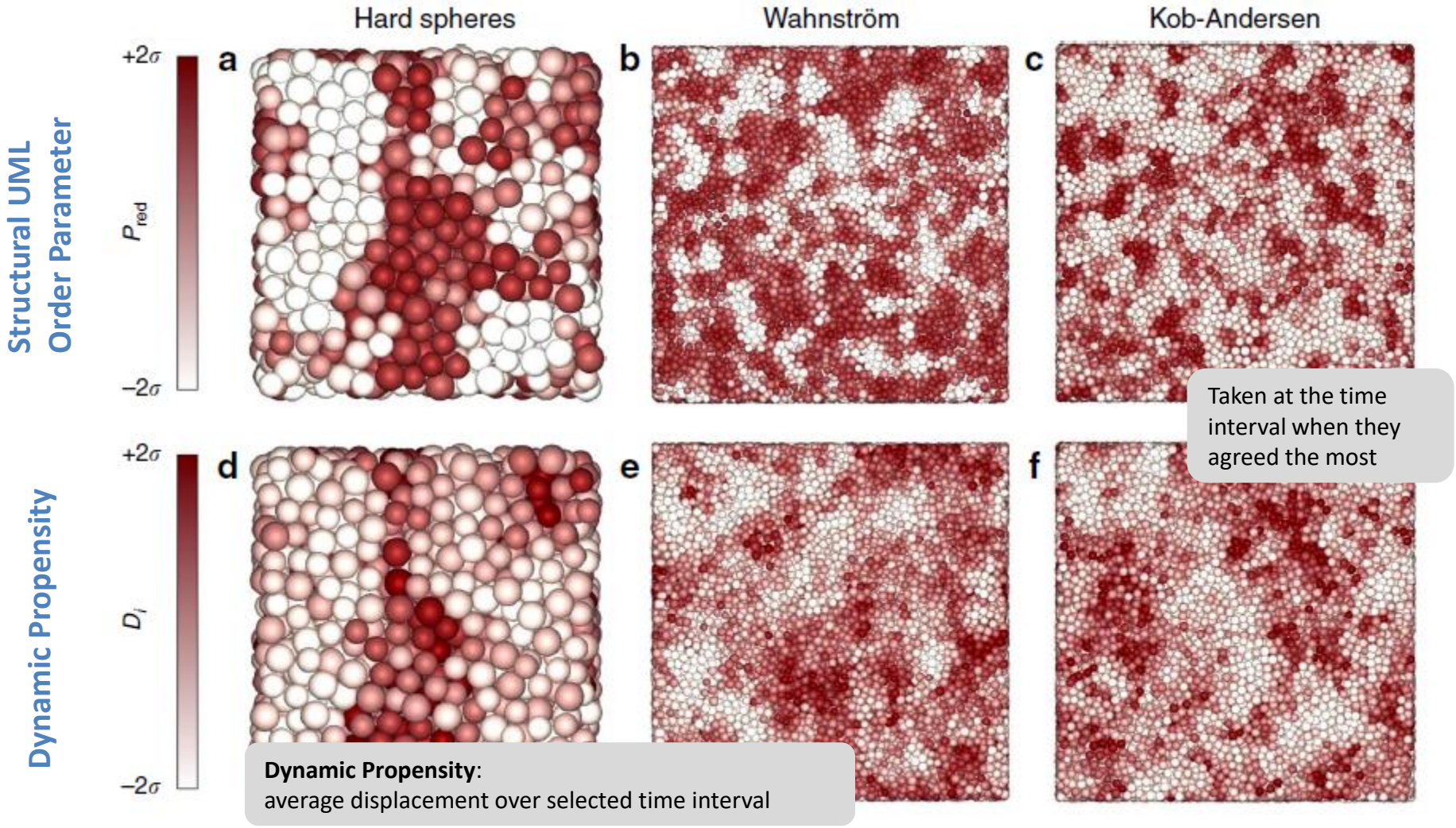
Structural order parameter



Structural order parameter



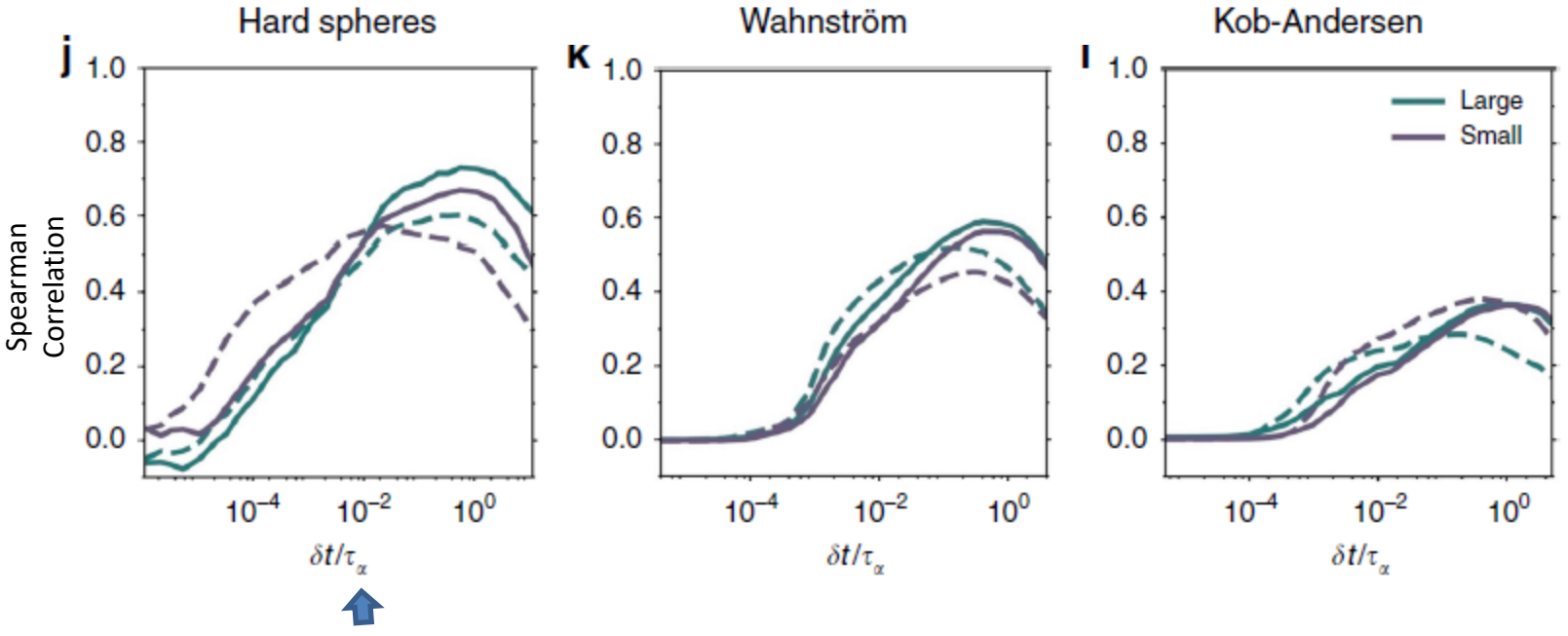
Question: Does this structural order parameter correlate with the dynamics?



Taken at the time interval when they agreed the most

Dynamic Propensity:
average displacement over selected time interval

Correlation between structure and dynamics



Dashed lines: basic P_{red}
Solid Lines: P_{red} averaged over a local region

Conclusions

- We have introduced a **simple, fast,** and **easy to implement** **unsupervised learning algorithm** for **autonomously recognizing local structural motifs.**
- Algorithm works extremely well in all situations we have explored so far
- In all glassy three model systems (HS, W, KA), two local environments are identified
- Structural variations correlate well with dynamics in HS and W, less well in KA

E. Boattini, M. Ram, F. Smallenburg, and L. Filion, Mol. Phys. 116, 3066 (2018)
E. Boattini, M. Dijkstra, L. Filion, J. Chem. Phys. **151**, 154901 (2019)
E. Boattini, S. Marin-Aguilar, S. Mitra , G. Foffi, F. Smallenburg, L. Filion, Nat. Commun. **11**, 5479 (2020)

Conclusions

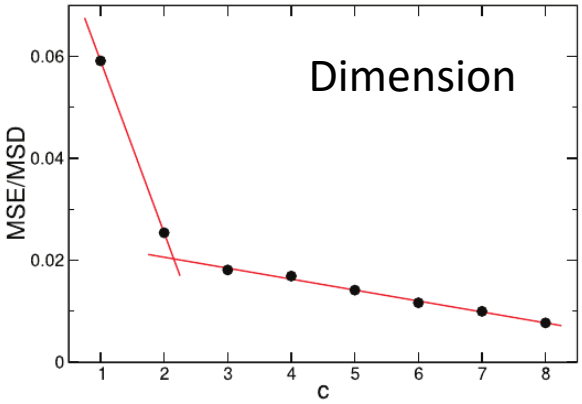
- We have introduced a **simple, fast,** and **easy to implement** **unsupervised learning algorithm** for **autonomously recognizing local structural motifs.**
- Algorithm works extremely well in all situations we have explored so far
- In all glassy three model systems (HS, W, KA), two local environments are identified
- Structural variations correlate well with dynamics in HS and W, less well in KA

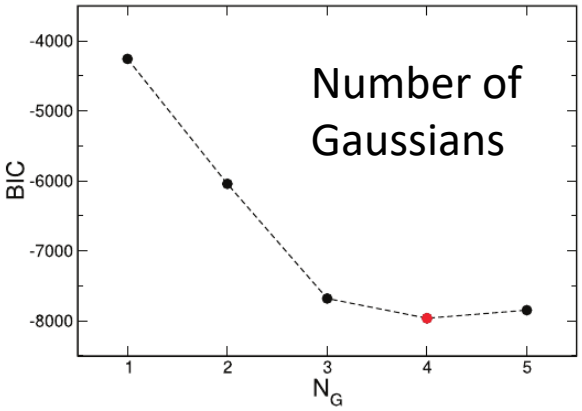
Missing ingredients?

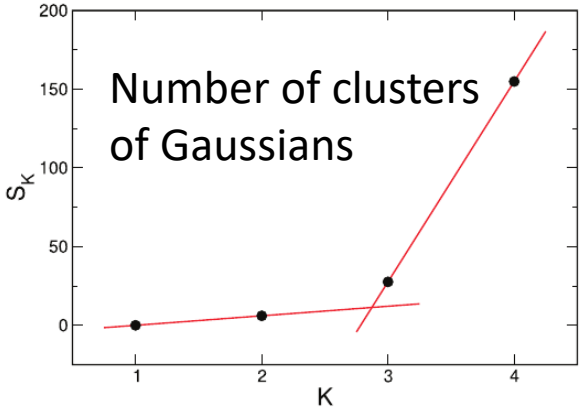
- **No local density** information included in the input – perhaps important for improving structural order parameter
- Purely **small local** environmental descriptors

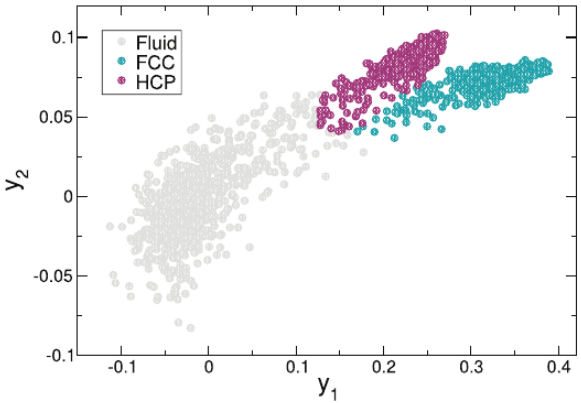
E. Boattini, M. Ram, F. Smallenburg, and L. Filion, Mol. Phys. 116, 3066 (2018)
E. Boattini, M. Dijkstra, L. Filion, J. Chem. Phys. **151**, 154901 (2019)
E. Boattini, S. Marin-Aguilar, S. Mitra , G. Foffi, F. Smallenburg, L. Filion, Nat. Commun. **11**, 5479 (2020)

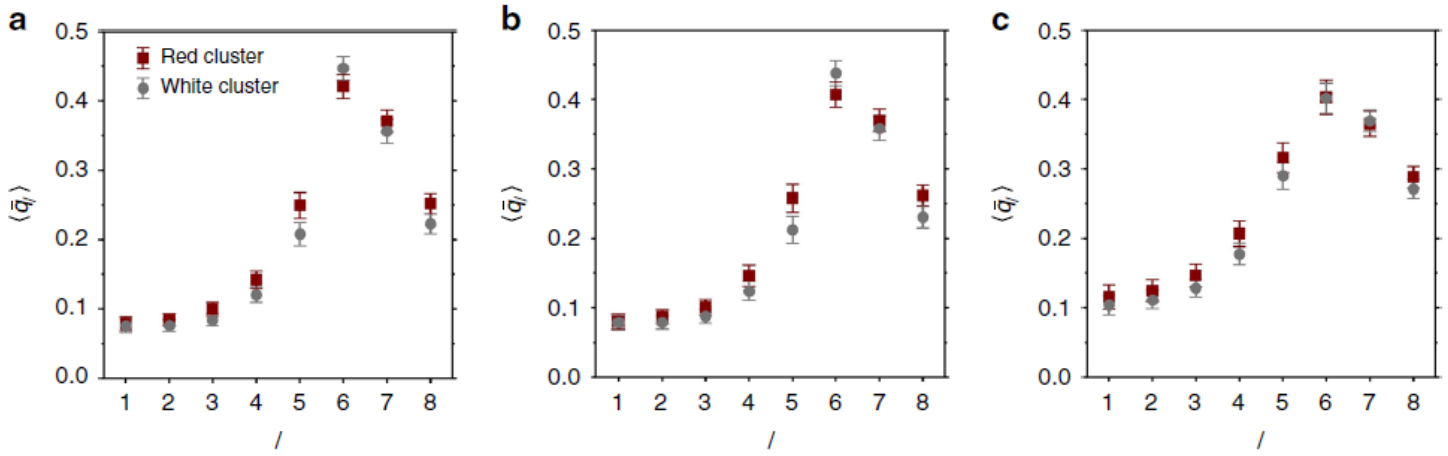
Questions?


Dimension
MSE/MSD
0.06
0.04
0.02
0
1
2
3
4
5
6
7
8
c
(b)

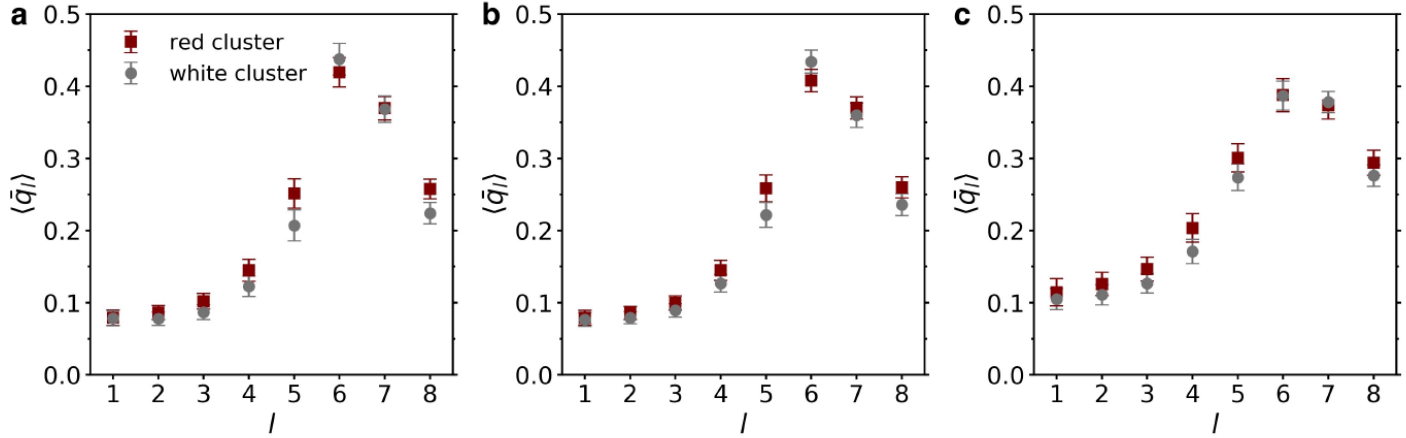

Number of Gaussians
BIC
-4000
-5000
-6000
-7000
-8000
1
2
3
4
5
N_G
(c)


Number of clusters of Gaussians
S_K
200
150
100
50
0
1
2
3
4
K
(d)

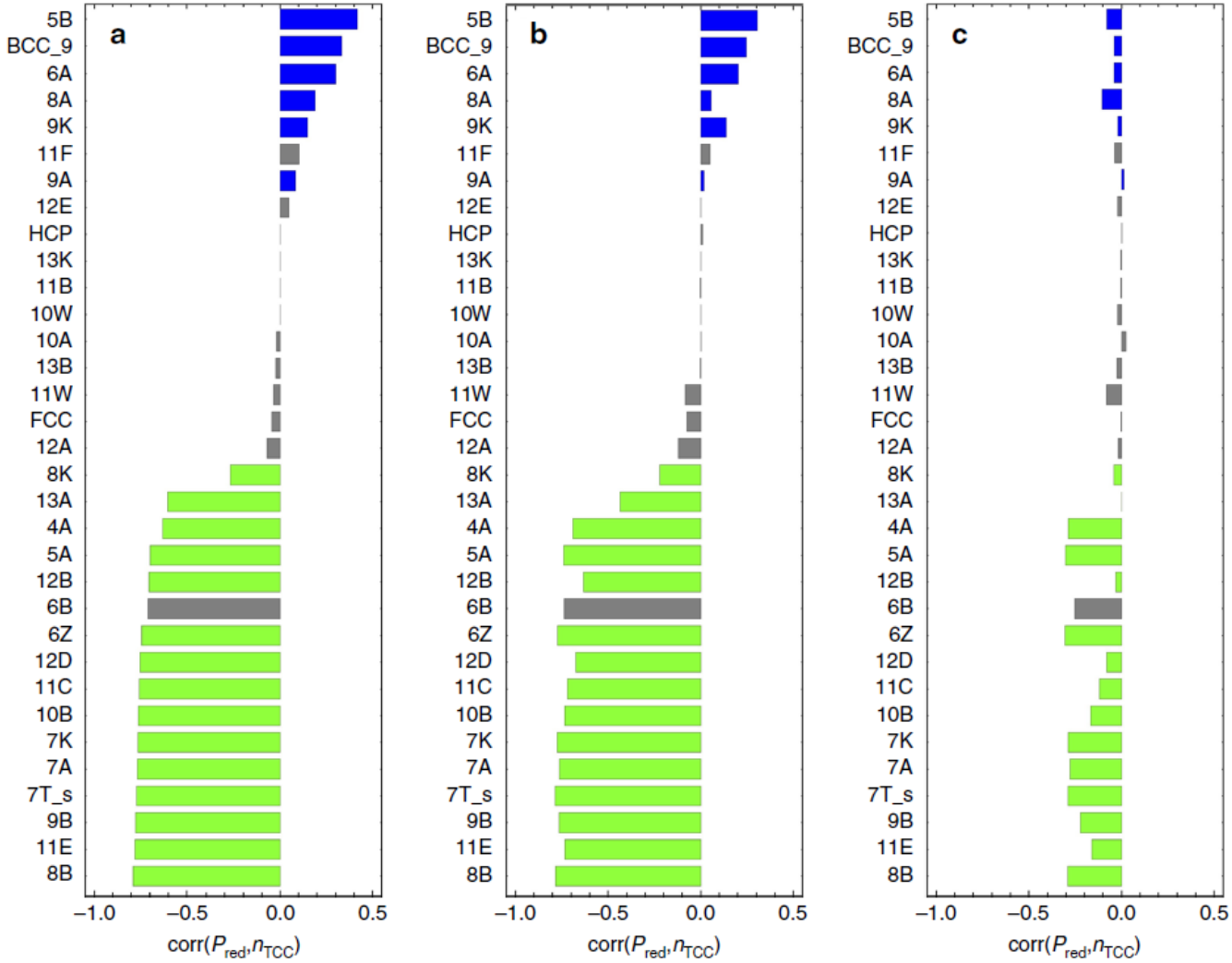

Fluid
FCC
HCP
y_2
0.1
0.05
0
-0.05
-0.1
-0.1
0
0.1
0.2
0.3
0.4
y_1
(e)



Large Particles



Small Particles



Green:
Polytetrahedral

Blue:
Square Pyramids