



Machine Learning for Molecular Physics

WiSe 2021/22

L07 - Structure and property classification in bulk

1. Classification vs. regression
2. Local structure classification
3. Property prediction

Classification vs. regression

Classification vs. Regression

- artificial neural networks: '*...general framework for representing non-linear mappings between multi-dimensional spaces...*'
- regression
 - one or more continuous output values (e.g. an energy)
 - map input \mathbf{x} to an output y_k
- classification
 - set of discrete classes $\{C_k\}$, one output for each class
 - assign input \mathbf{x} to, e.g., class with largest output value y_k

Bishop, Neural networks: a pattern recognition perspective, Aston University, Birmingham, 1996.

Statistical pattern recognition

- posterior probability of class membership

$$P(\mathcal{C}_k | \mathbf{x}) = \frac{p(\mathbf{x} | \mathcal{C}_k) P(\mathcal{C}_k)}{p(\mathbf{x})}$$

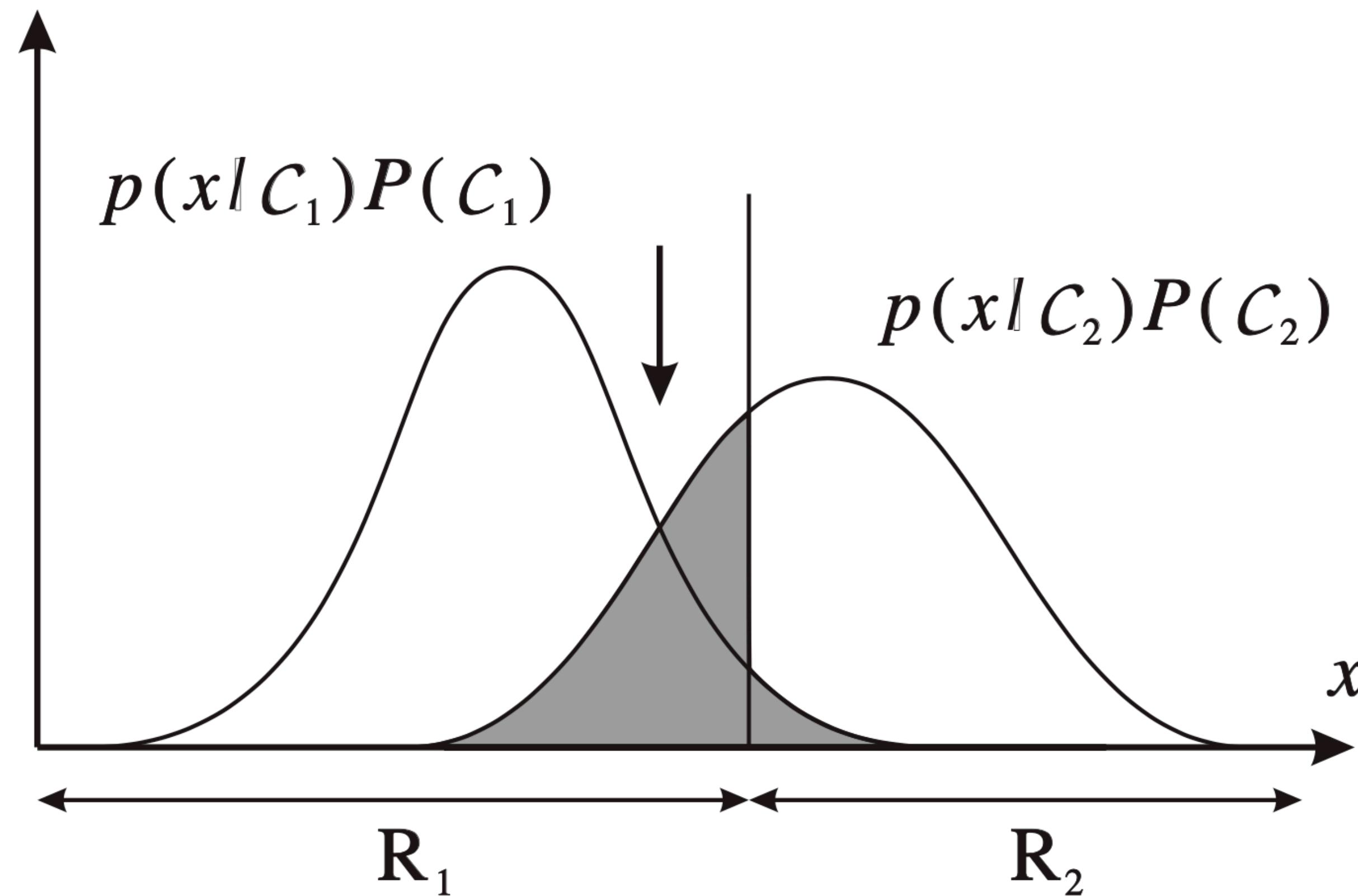
- assign \mathbf{x} to C_k if

$$P(\mathcal{C}_k | \mathbf{x}) > P(\mathcal{C}_j | \mathbf{x}) \quad \text{for all } j \neq k.$$

Bishop, Neural networks: a pattern recognition perspective, Aston University, Birmingham, 1996.

Statistical pattern recognition

- joint probability density $p(x, \mathcal{C}_k) = p(x|\mathcal{C}_k)P(\mathcal{C}_k)$ and decision boundary



Bishop, Neural networks: a pattern recognition perspective, Aston University, Birmingham, 1996.

Statistical pattern recognition

- reject criterion

$$\text{if } \max_k P(\mathcal{C}_k | \mathbf{x}) \begin{cases} \geq \theta, & \text{then classify } \mathbf{x} \\ < \theta, & \text{then reject } \mathbf{x} \end{cases}$$

Bishop, Neural networks: a pattern recognition perspective, Aston University, Birmingham, 1996.

Loss functions



- maximum likelihood, training data $\{\mathbf{x}^n, \mathbf{t}^n\}$

$$\mathcal{L} = \prod_n p(\mathbf{t}^n | \mathbf{x}^n)$$

$$E = -\ln \mathcal{L} = -\sum_n \ln p(\mathbf{t}^n | \mathbf{x}^n)$$

- independent distribution of target variables

$$p(\mathbf{t} | \mathbf{x}) = \prod_{k=1}^c p(t_k | \mathbf{x}).$$

Bishop, Neural networks: a pattern recognition perspective, Aston University, Birmingham, 1996.

Loss function regression

- target variable with Gaussian noise

$$t_k = h_k(\mathbf{x}) + \epsilon_k.$$

$$p(\epsilon_k) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{\epsilon_k^2}{2\sigma^2}\right)$$

- distribution of target values

$$p(t_k | \mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{\{y_k(\mathbf{x}; \mathbf{w}) - t_k\}^2}{2\sigma^2}\right)$$

- loss function

$$E = \frac{1}{2\sigma^2} \sum_{n=1}^N \sum_{k=1}^c \{y_k(\mathbf{x}^n; \mathbf{w}) - t_k^n\}^2 + Nc \ln \sigma + \frac{Nc}{2} \ln(2\pi).$$

Bishop, Neural networks: a pattern recognition perspective, Aston University, Birmingham, 1996.

Loss function classification

- two classes C_1 and C_2 with $P(C_1|\mathbf{x}) = y$ and $P(C_2|\mathbf{x}) = 1 - y$
- target value $t = 1$ if $\mathbf{x} \in C_1$ and $t = 0$ if $\mathbf{x} \in C_2$

$$p(t|\mathbf{x}) = y^t (1 - y)^{1-t}$$

- likelihood of observing training data

$$\prod_n (y^n)^{t^n} (1 - y^n)^{1-t^n}.$$

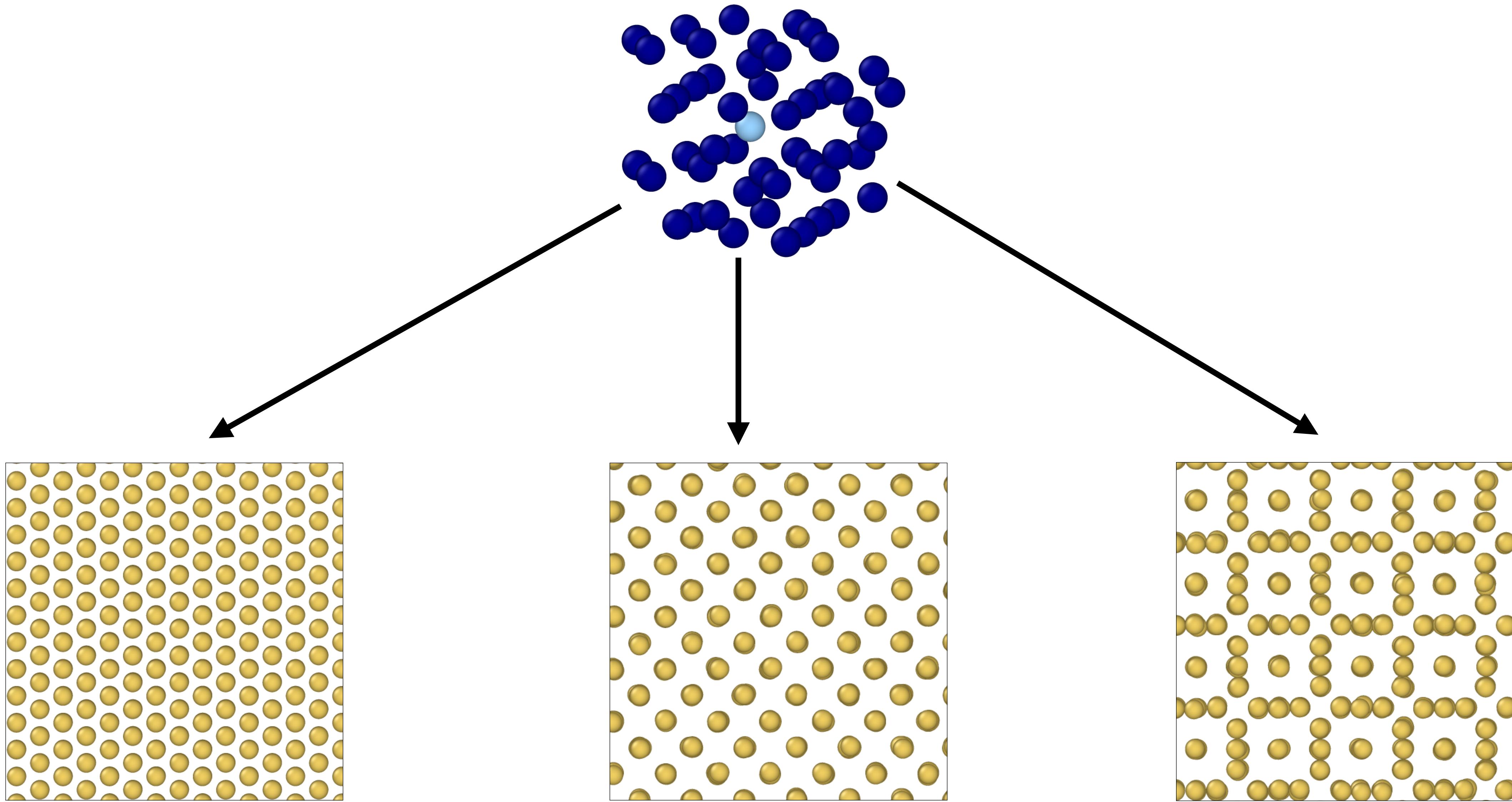
- loss function → cross-entropy

$$E = - \sum_n \{ t^n \ln y^n + (1 - t^n) \ln(1 - y^n) \}$$

Bishop, Neural networks: a pattern recognition perspective, Aston University, Birmingham, 1996.

Local structure classification

Local structure identification



Local structure identification

- common ‘classical’ approaches
 - common neighbour analysis (CNA)
 - centro-symmetric deviation theory
 - bond-angle distribution
 - Steinhardt bond-order parameters

Steinhardt bond-order parameters

- based on spherical harmonics

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

$$q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(\mathbf{r}_{ij})$$

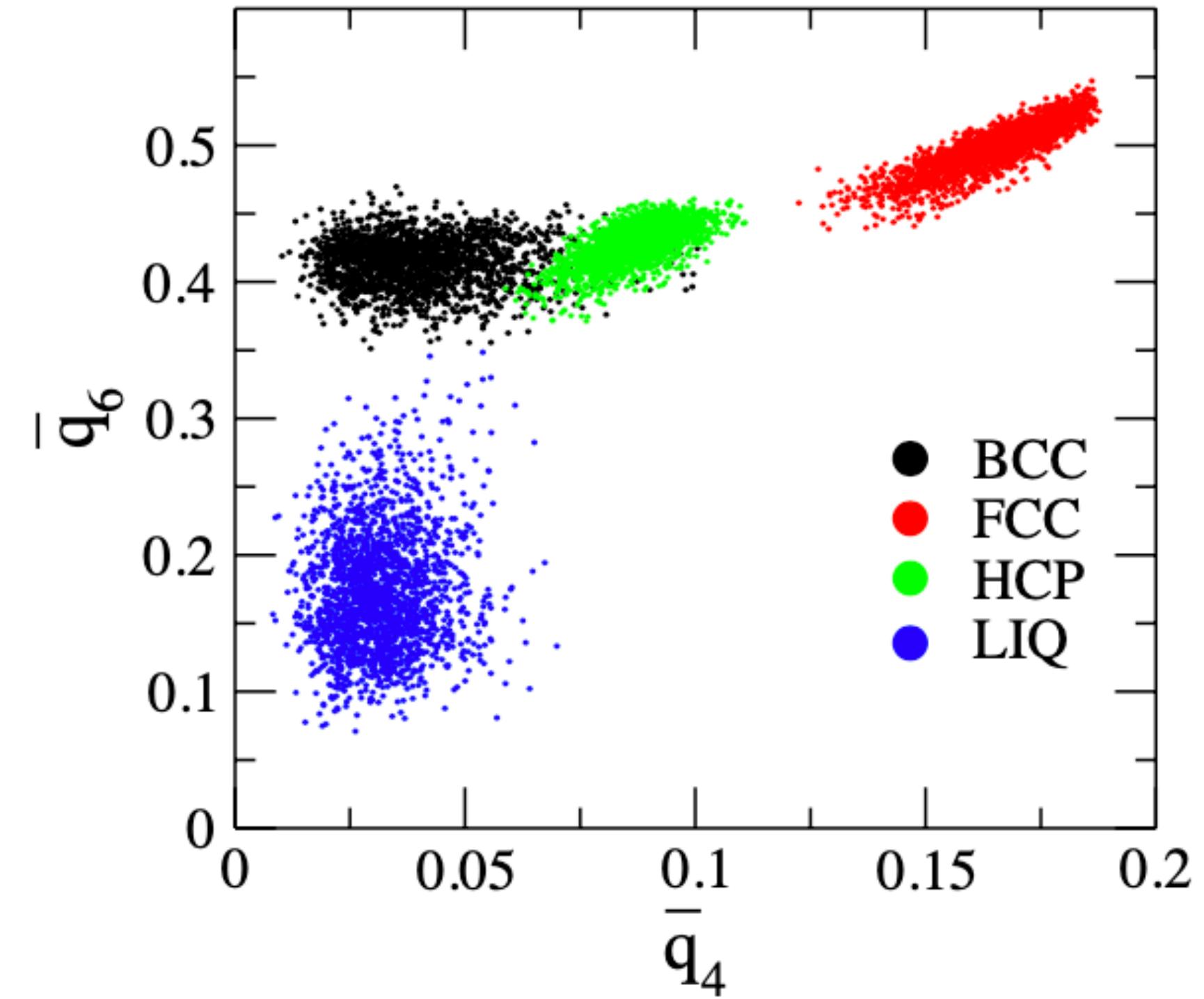
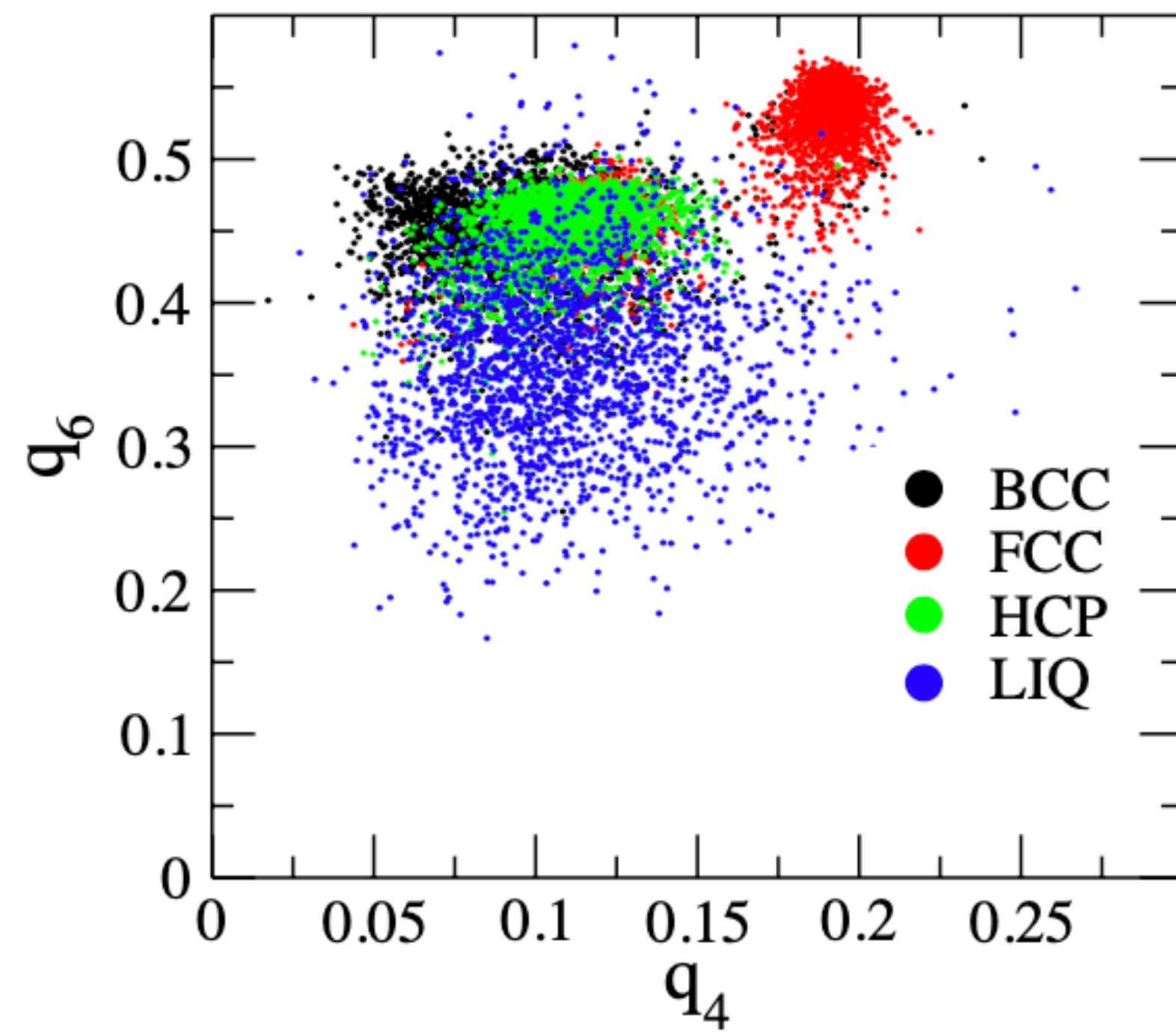
- averaged version

$$\bar{q}_l(i) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^l |\bar{q}_{lm}(i)|^2}$$

$$\bar{q}_{lm}(i) = \frac{1}{\tilde{N}_b(i)} \sum_{k=0}^{\tilde{N}_b(i)} q_{lm}(k)$$

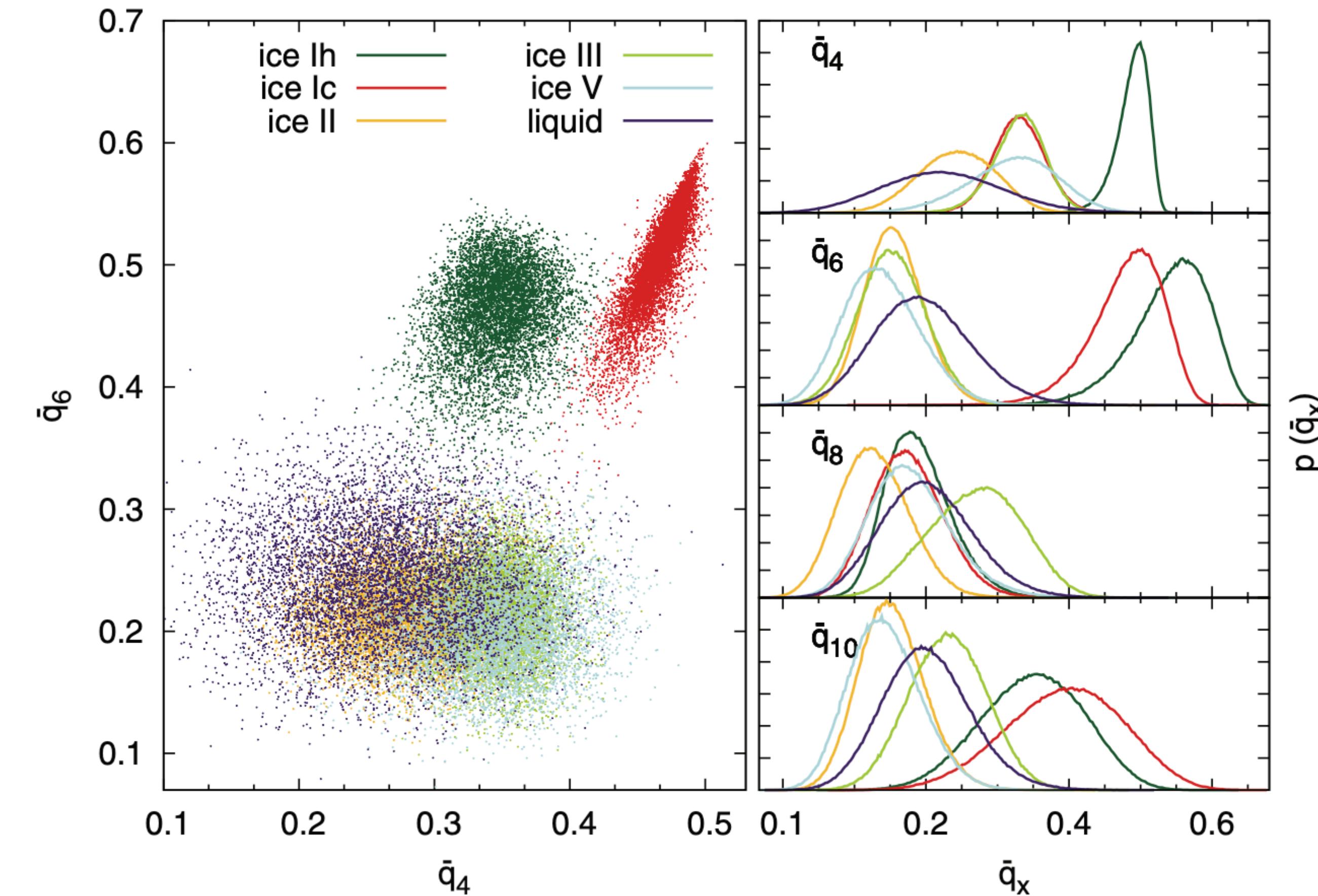
Structure identification maps

- Lennard-Jones system



Lechner, Dellago, JCP 129, 114707 (2008)

Classifying phases in ice

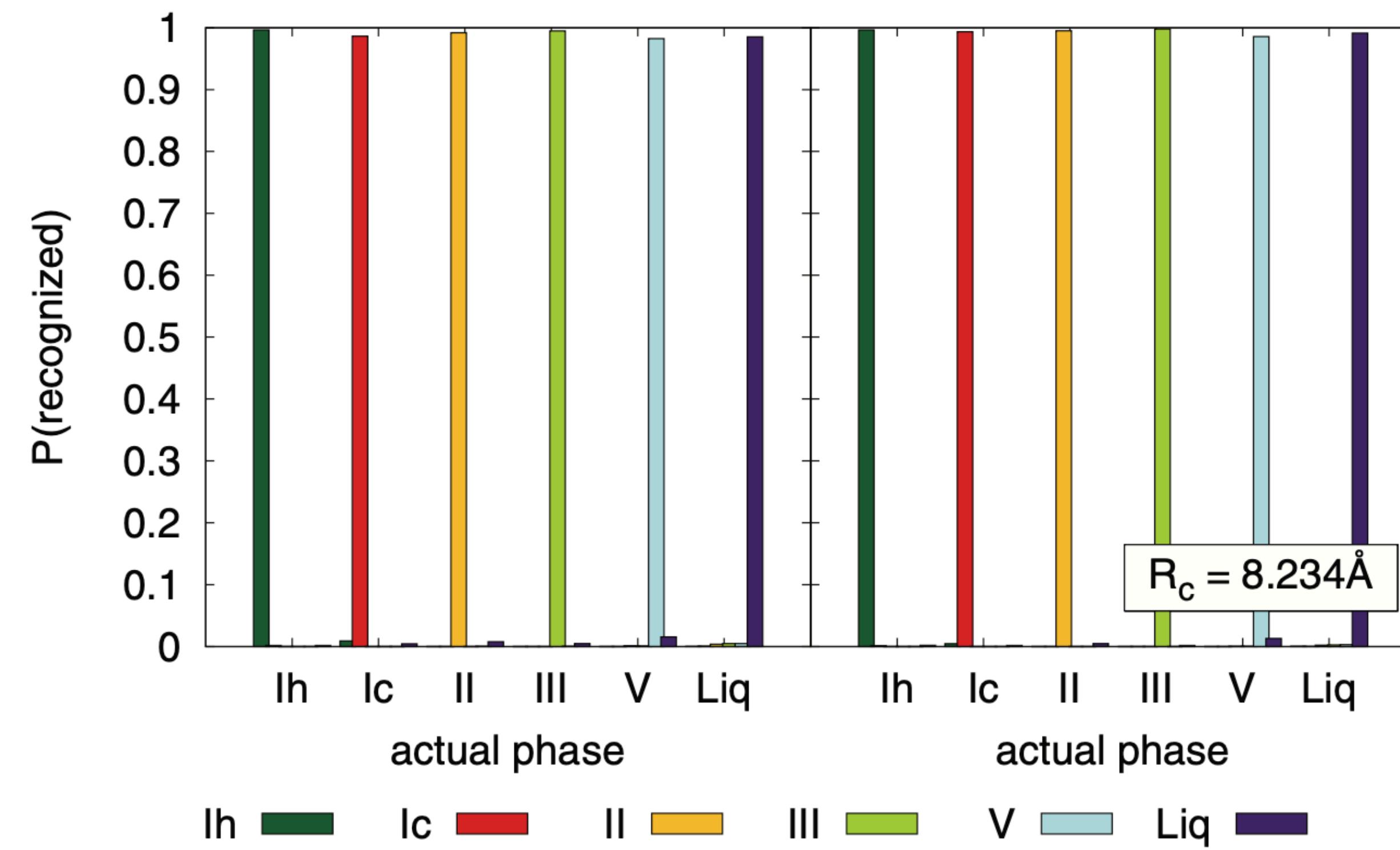


- Steinhardt parameters not sufficient

Geiger, Dellago, JCP 139, 164105 (2013)

Classification neural network

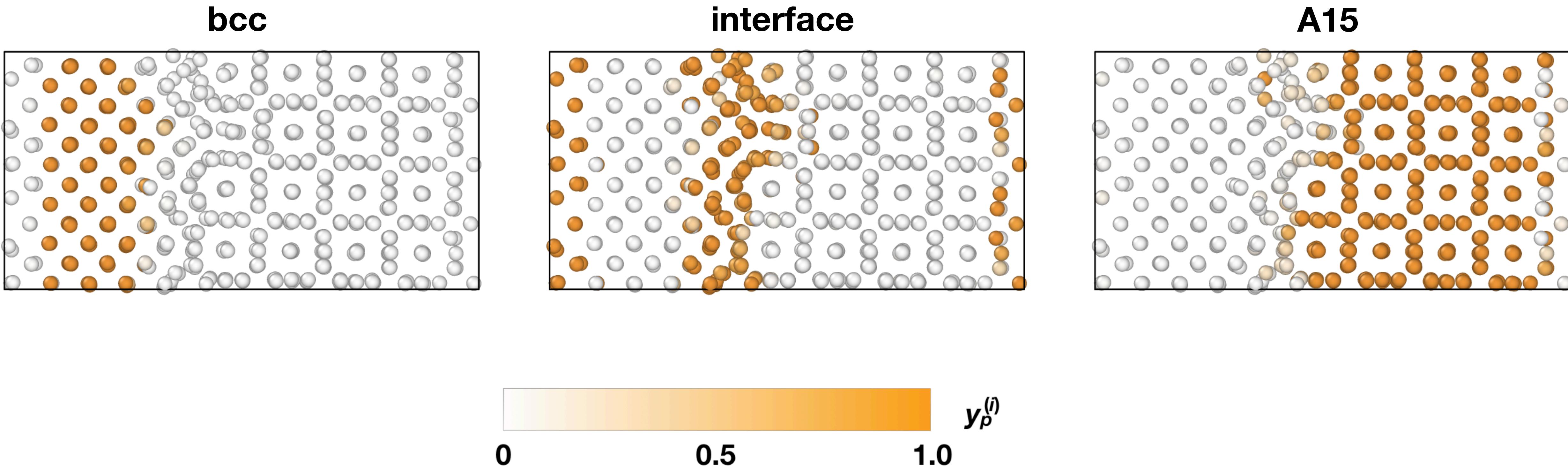
- input descriptors: Behler-Parrinello symmetry functions + additions, only for oxygen, approx. 50 functions (radial and angular)
- feed-forward NN, training set 10^5 configurations



Geiger, Dellago, JCP 139, 164105 (2013)

Input descriptors

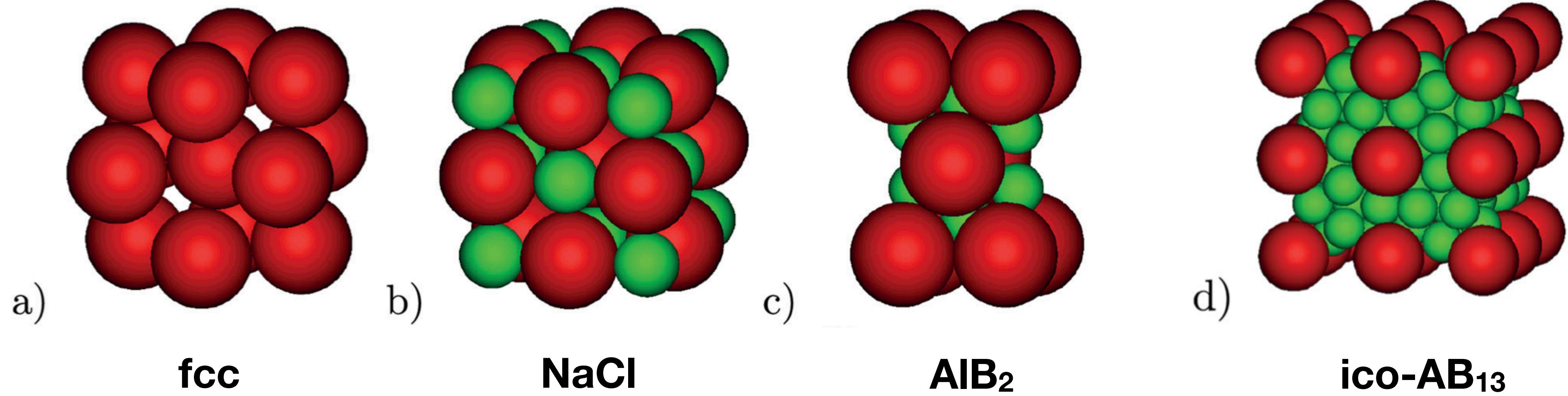
- topologically closed-packed phases and interfaces
- 11 radial symmetry functions + 3 Steinhard parameters ($l = 6, 7, 8$)



Rogal et al., PRL 123, 245701 (2019)

Input descriptors

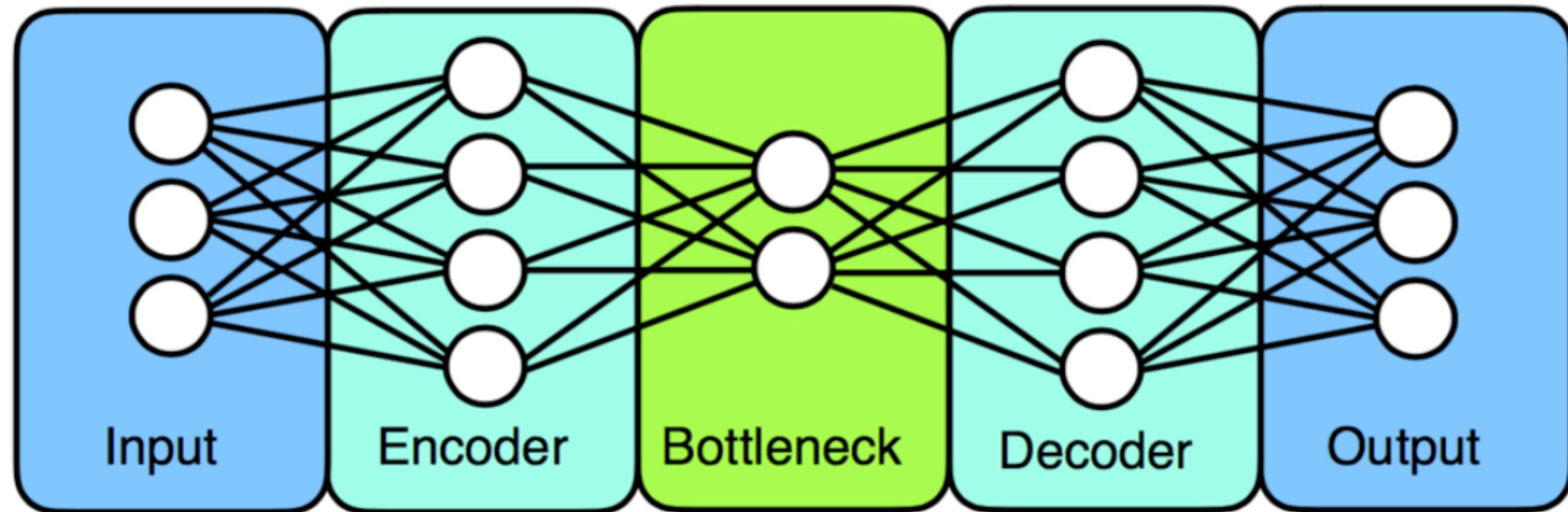
- binary hard spheres (two different sizes)
- 24 bond-order parameters $\mathbf{Q}(i) = (\{\bar{q}_l(i)\}, \{\bar{w}_{l'}(i)\}, \{\bar{q}_l^{LL(ss)}(i)\}, \{\bar{w}_{l'}^{LL(ss)}(i)\})$
- output labels for large/small particles



Boattini et al., Mol. Phys. 116, 3066 (2018)

Unsupervised learning

- autoencoder: learn low-dimensional, non-linear projections that preserves most relevant features
- input: $\mathbf{Q}(i) = (\{\bar{q}_l(i)\}, \{\bar{q}_l^{ss}(i)\})$
- optimal bottleneck dimension → ‘elbow’ plot



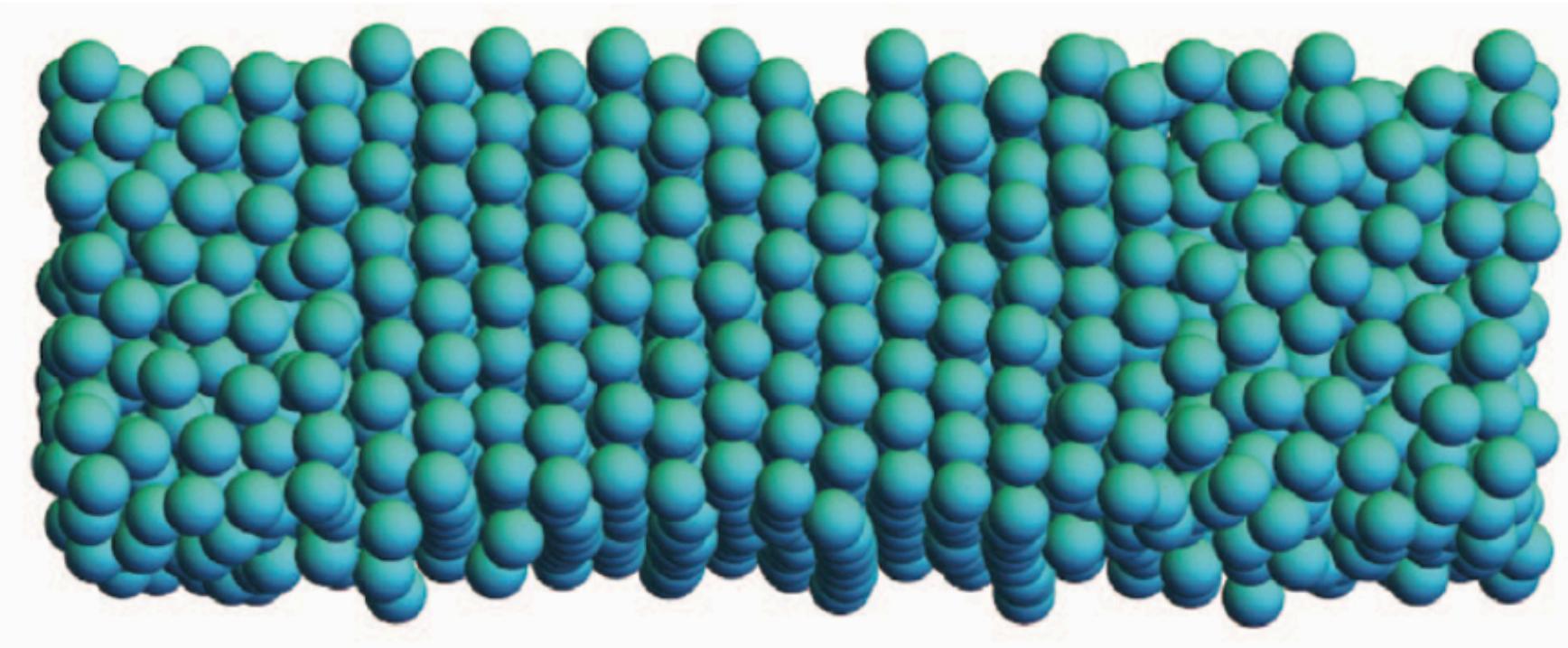
$$E(\mathbf{W}, \mathbf{B}; \{\mathbf{Q}(i)\}) = \frac{1}{N} \sum_{i=1}^N \|\mathbf{Q}(i) - \hat{\mathbf{Q}}(i)\|^2 + \lambda \sum_{j=1}^M w_j^2$$

Boattini et al., JCP 151, 154901 (2019)

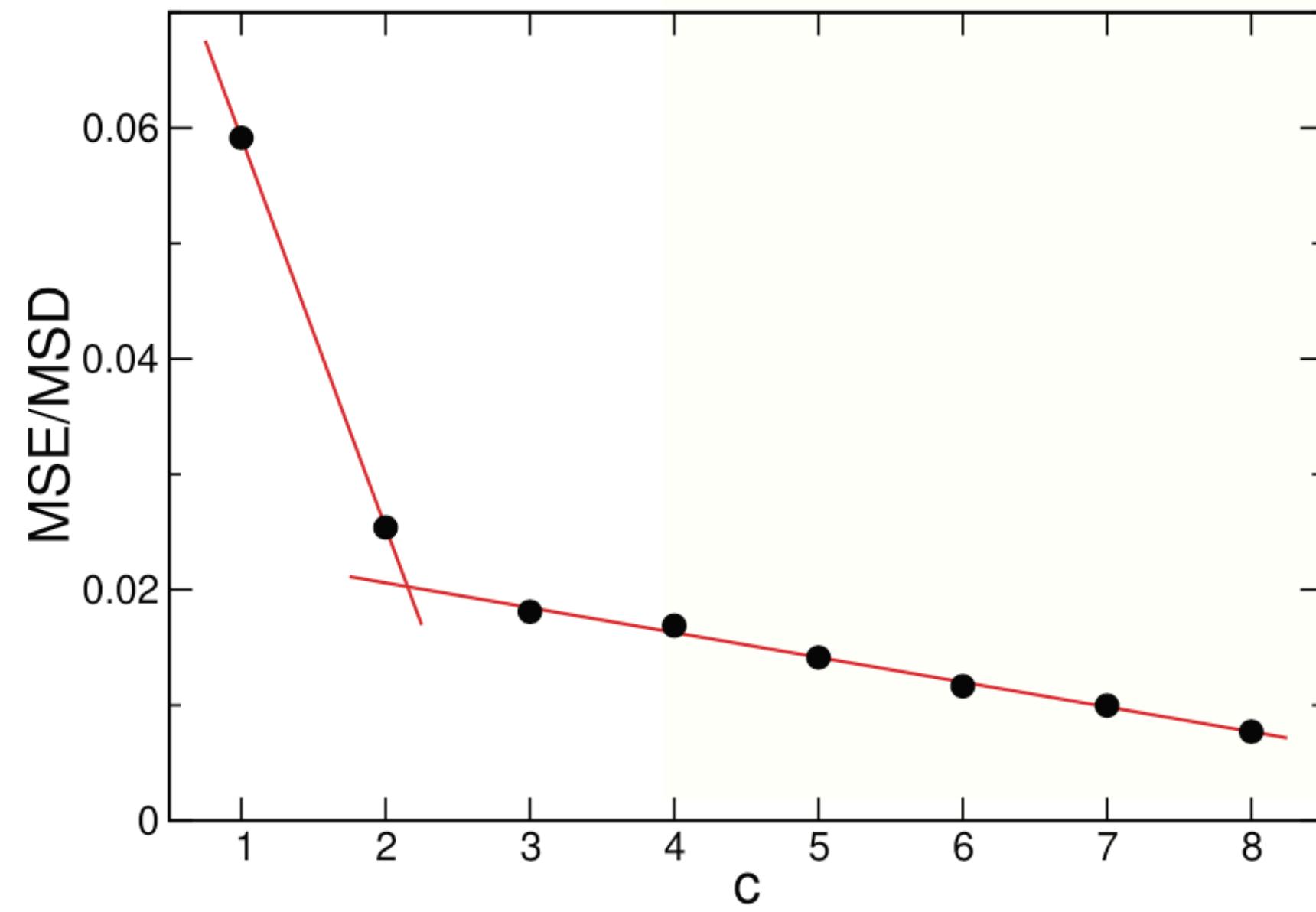
Clustering in latent space

- Gaussian mixture model: probabilities p_{ij} (posterior probability of i being from j th component)
- ‘elbow’ in entropy $S_K = - \sum_{i=1}^N \sum_{j=1}^K p_{ij} \ln(p_{ij})$ to find optimal number of clusters

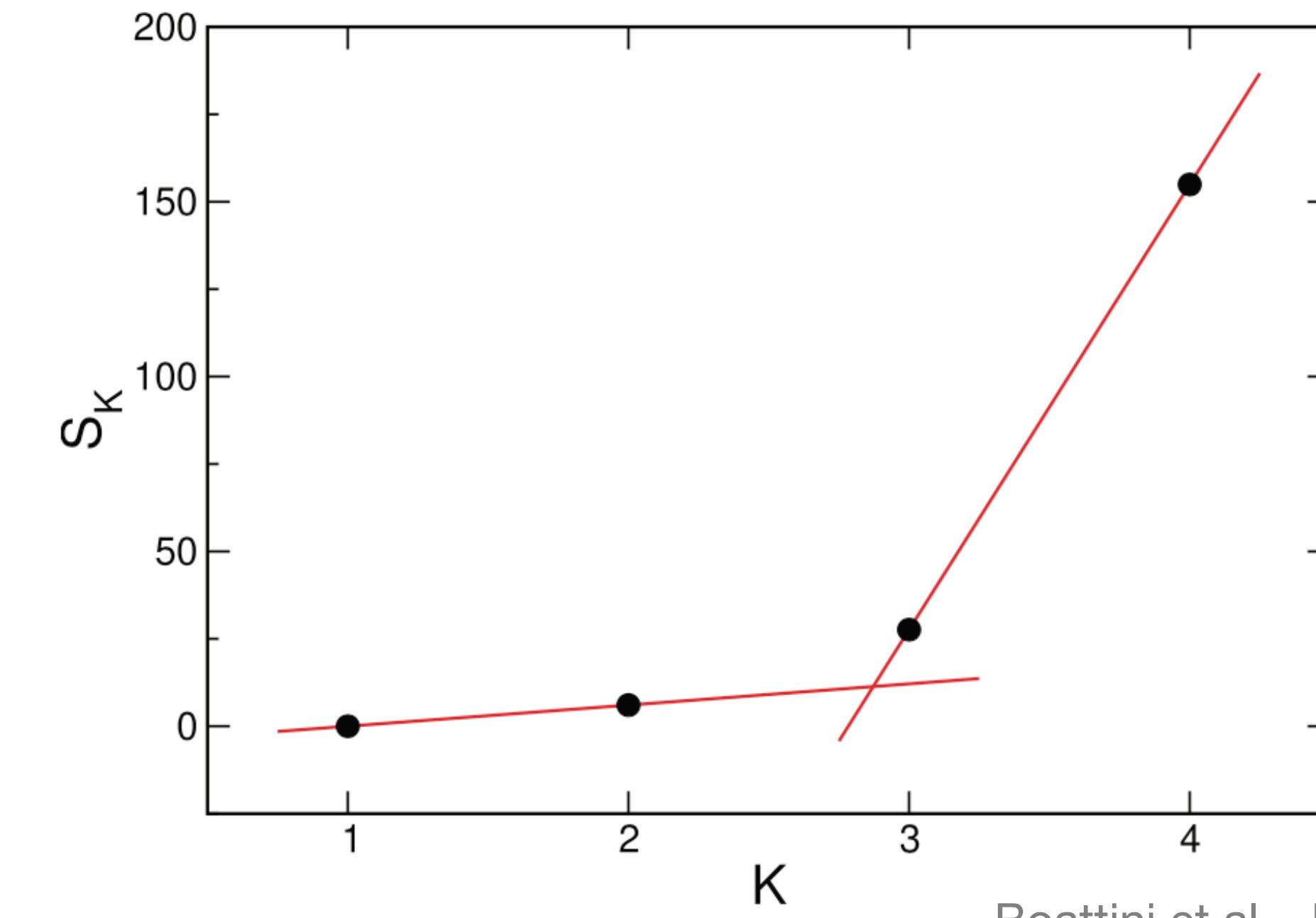
Classifying hard spheres



bottleneck dimension

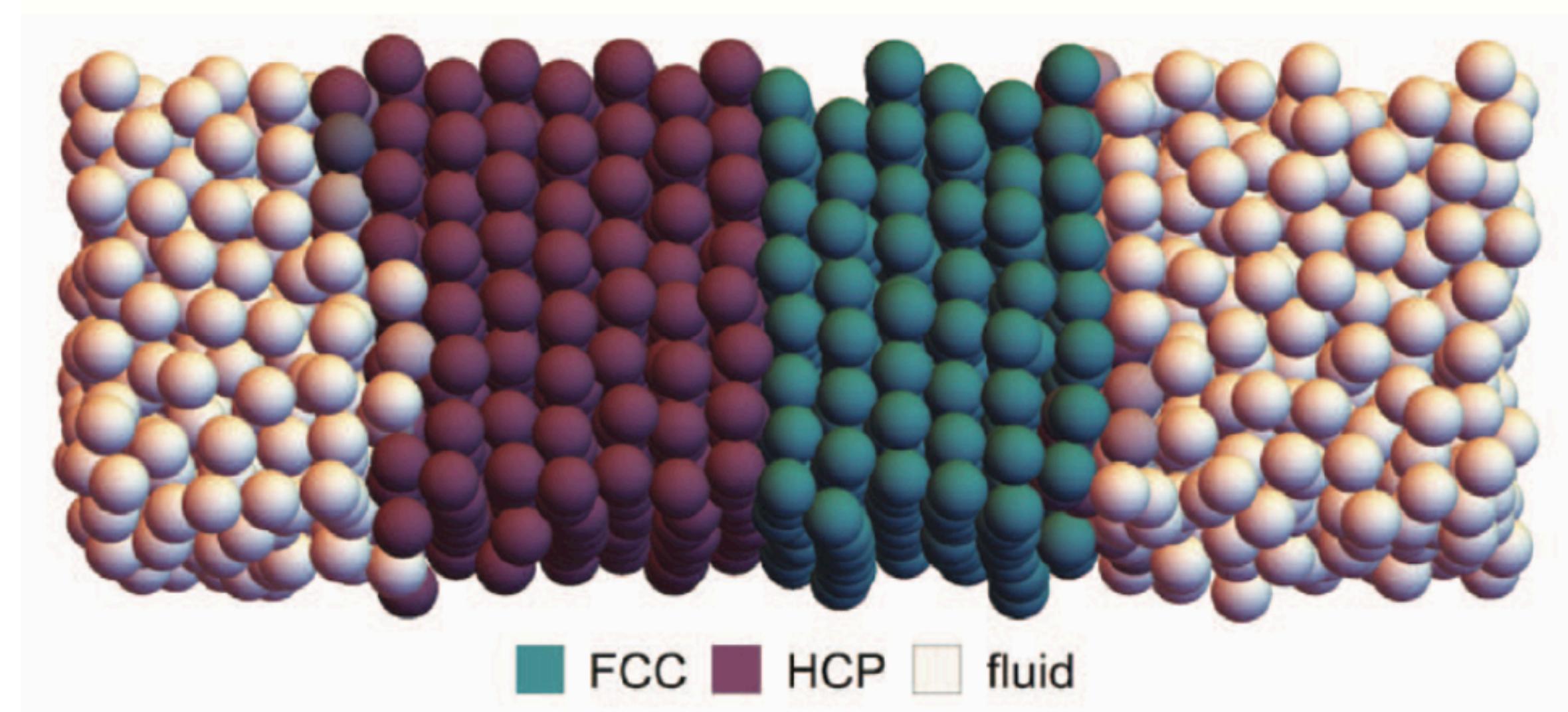
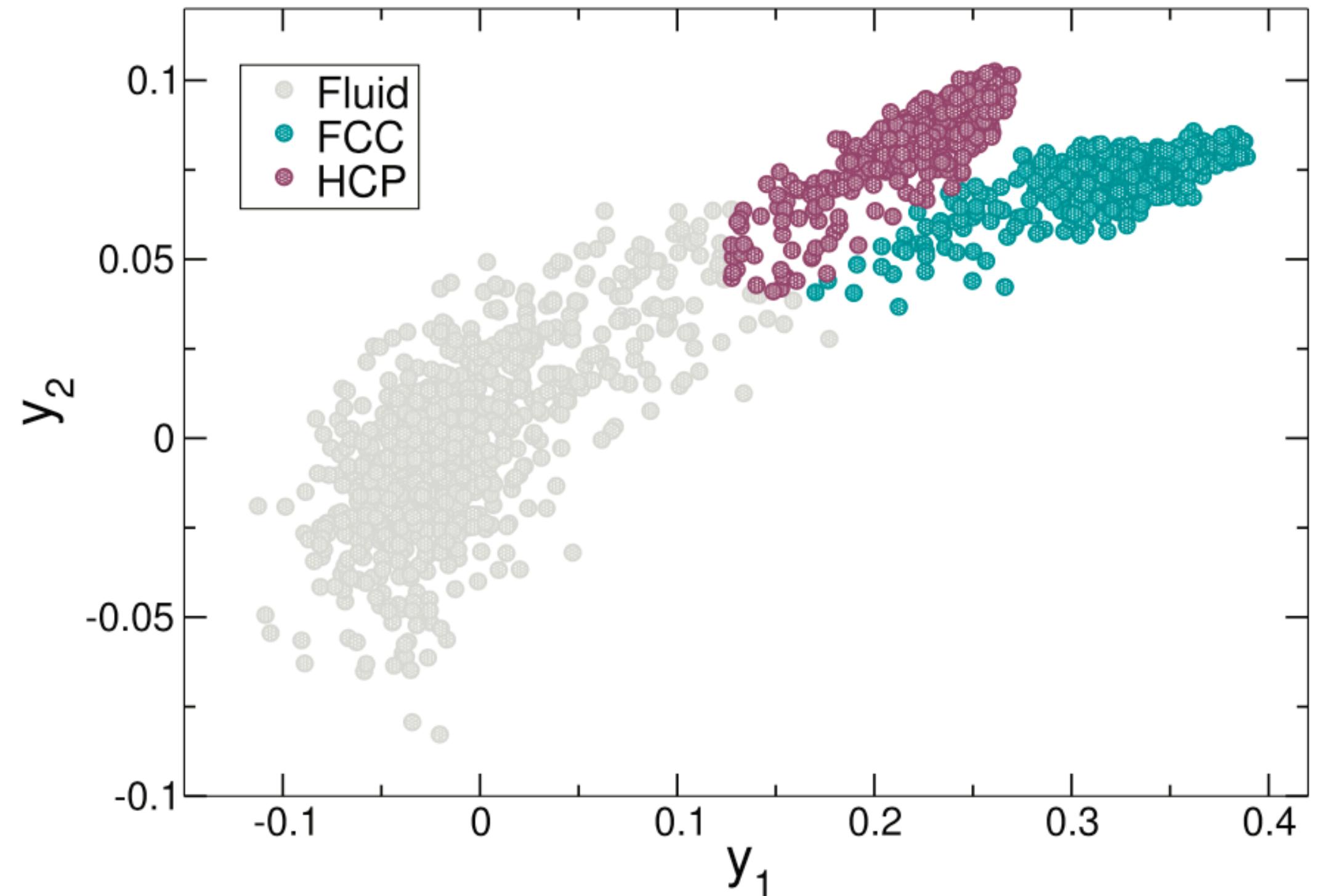


number of clusters



Boattini et al., JCP 151, 154901 (2019)

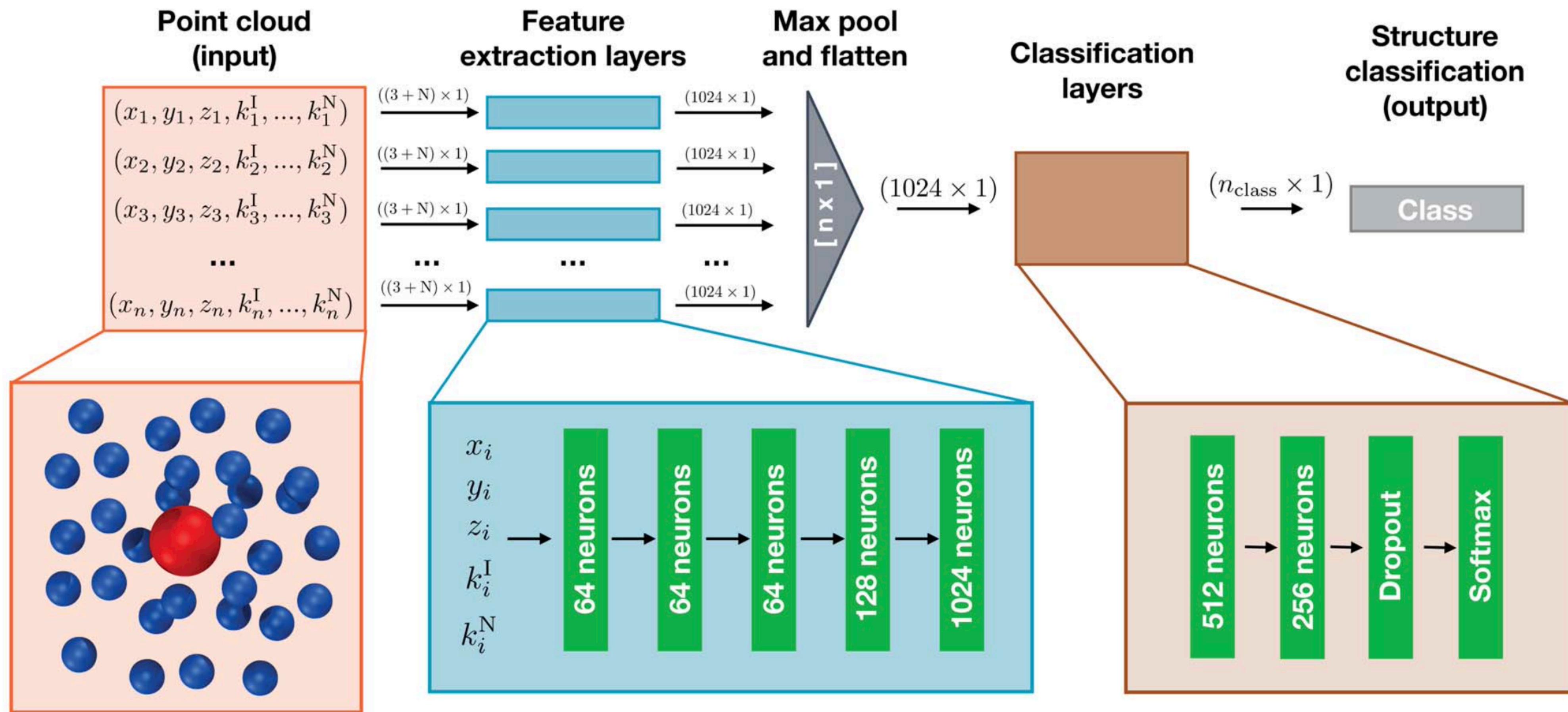
Classifying hard spheres



Boattini et al., JCP 151, 154901 (2019)

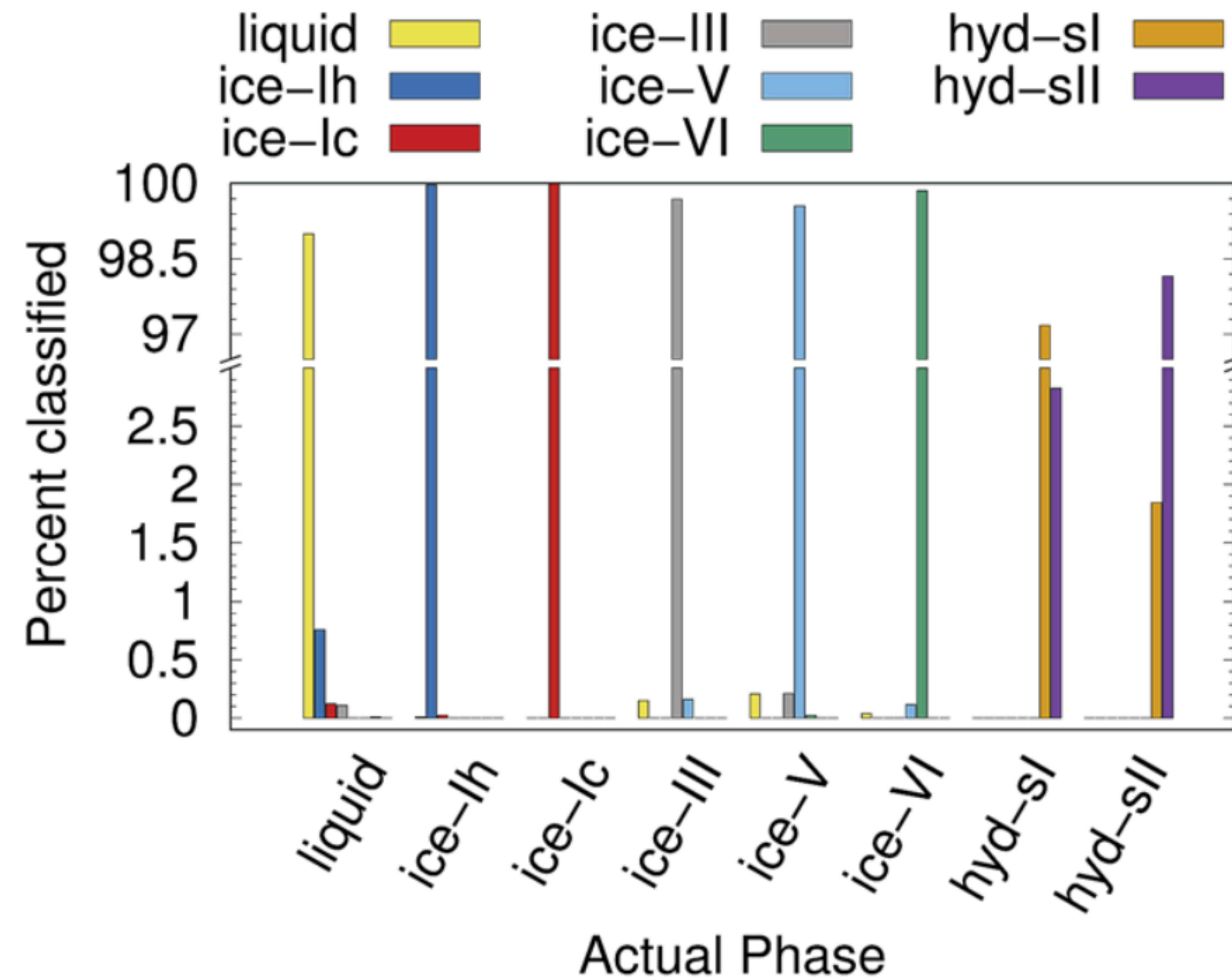
PointNet

- automatic feature extraction from point cloud



DeFever et al., Chem. Sci. 10, 7503 (2019)

Classifying ice phases



DeFever et al., Chem. Sci. 10, 7503 (2019)

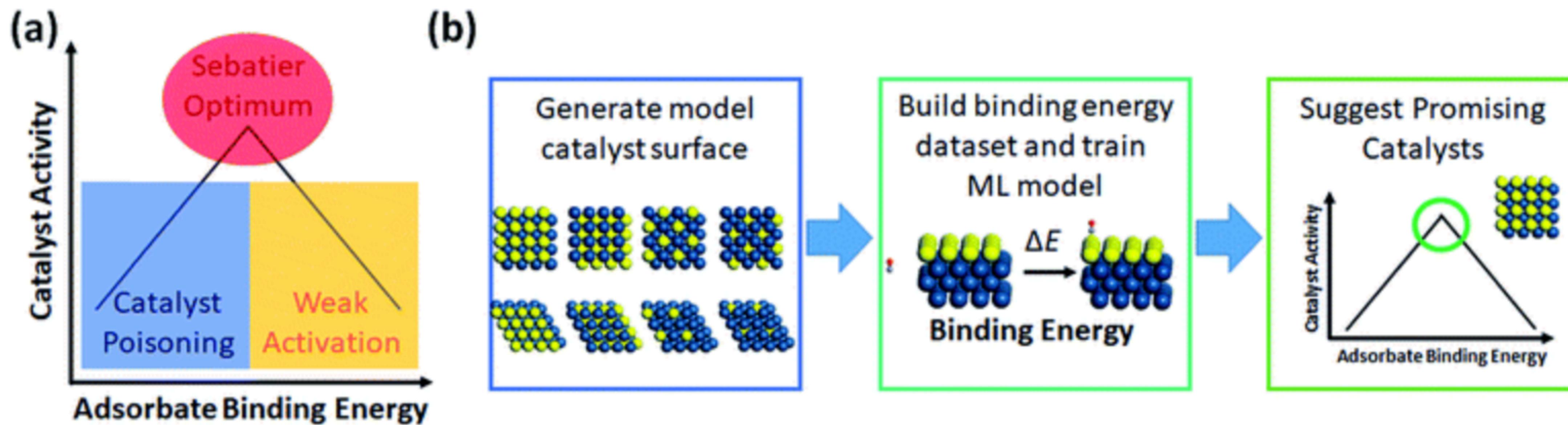
Property prediction

Screening and discovery

1. catalytic activity of surfaces
2. band gaps for solar cells
3. crystal graph convolutional neural network

Optimising catalyst activity

- predicting binding energies of metallic surfaces
- use electronic properties as input descriptors
- neural network for binding energy predictions



Ho Gu et al., J. Mater. Chem. A 7, 17096 (2019)

Input descriptors

- originally 13 electronic properties, ‘electronic fingerprints’
- 250 bimetallic systems

ΔE^*_{CO} : Binding energy of adsorbed CO on a metal surface

f : Filling of a d -band

ε_d : Center of a d -band

W_d : Width of a d -band

γ_1 : Skewness of a d -band

γ_2 : Kurtosis of a d -band

W : Work function

r_0 : Atomic radius

r_d : Spatial extent of d -orbitals

IE : Ionization potential

EA : Electron affinity

χ_0 : Pauling electronegativity

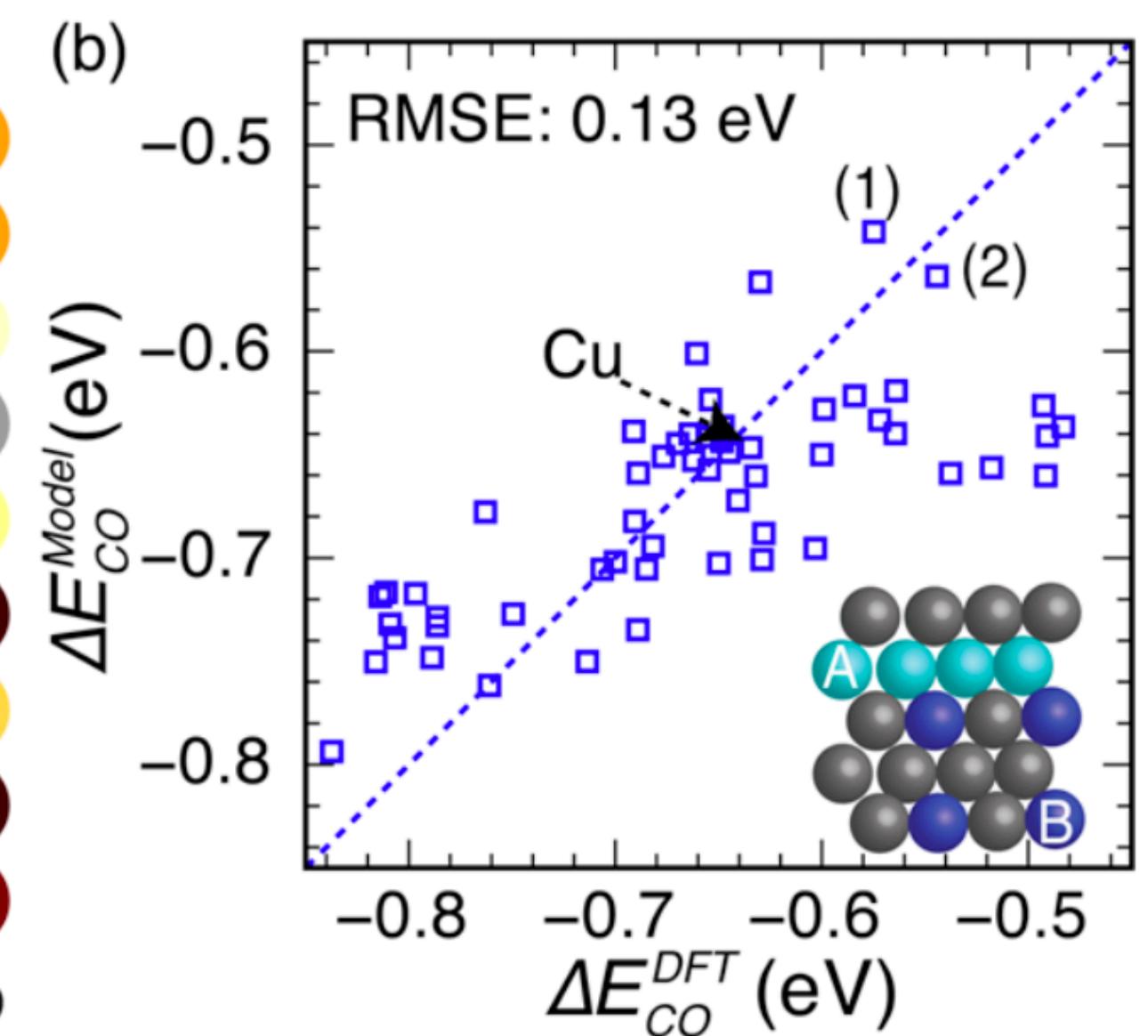
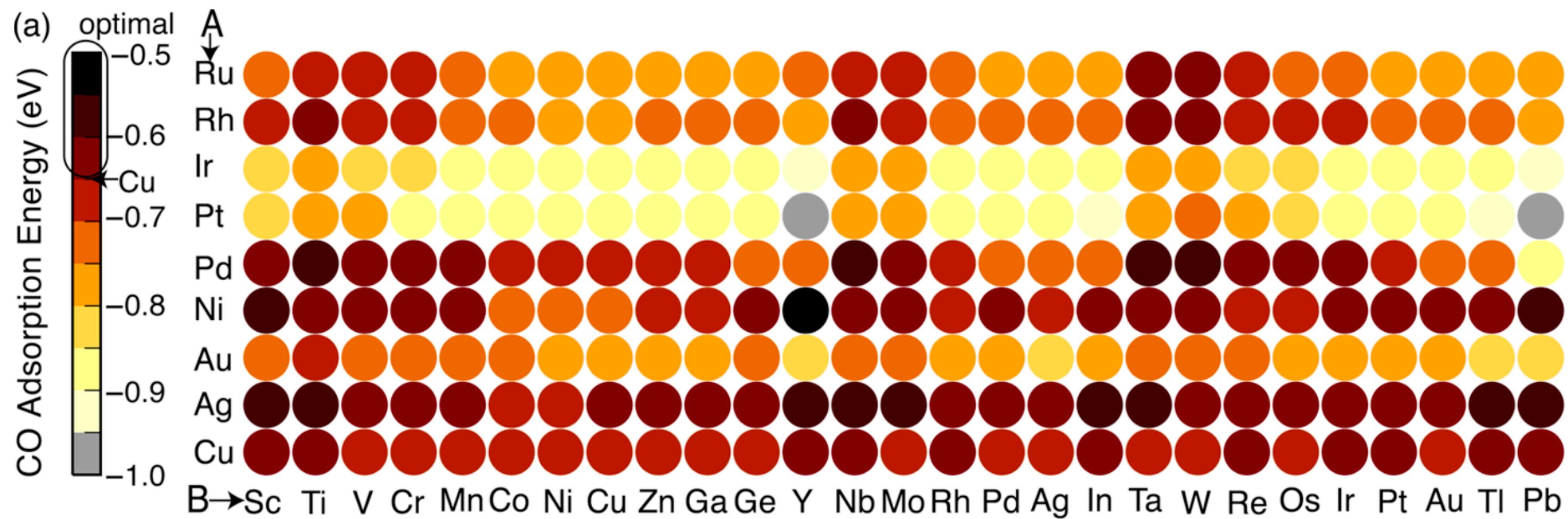
χ : Local Pauling electronegativity

V_{ad}^2 : Adsorbate-metal interatomic d coupling matrix element squared¹⁵

Ma et al., Phys. Chem. Lett. 6, 3528 (2015)

CO adsorption at Cu-based alloys

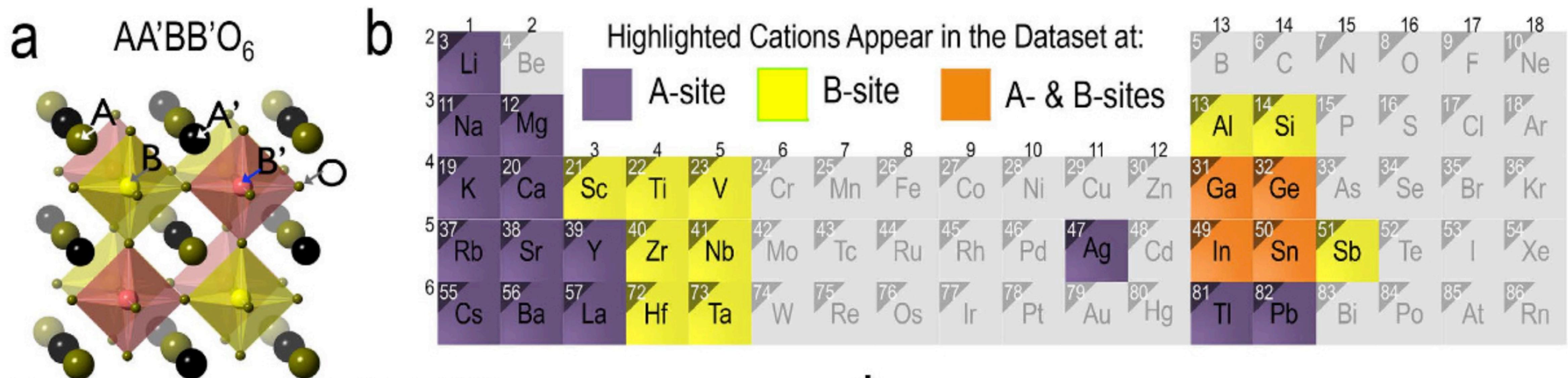
- 2nd generation core-shell alloys, $\text{Cu}_3\text{B}-\text{A}@\text{CuML}$



Ma et al., Phys. Chem. Lett. 6, 3528 (2015)

Bandgaps of double perovskites

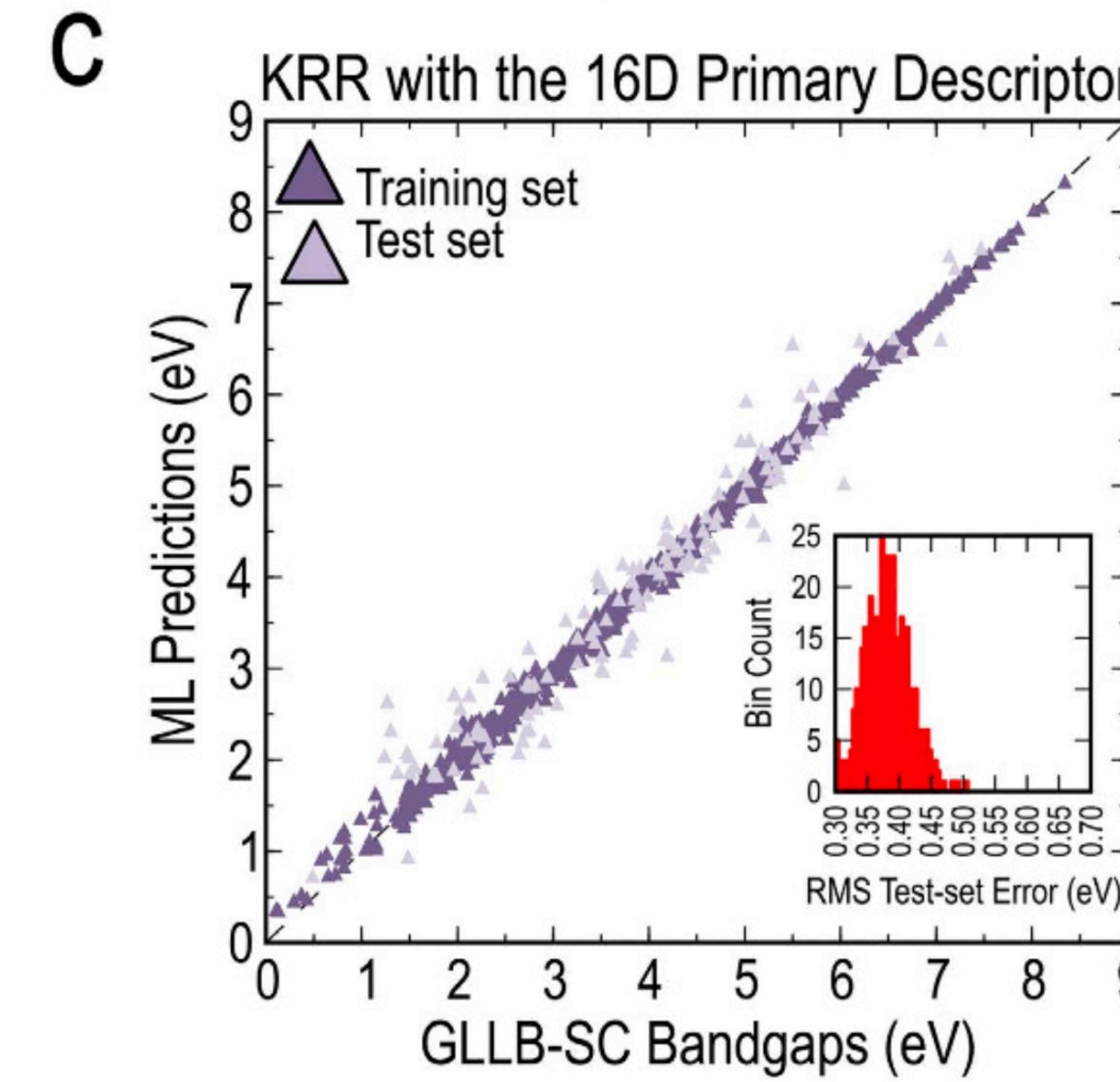
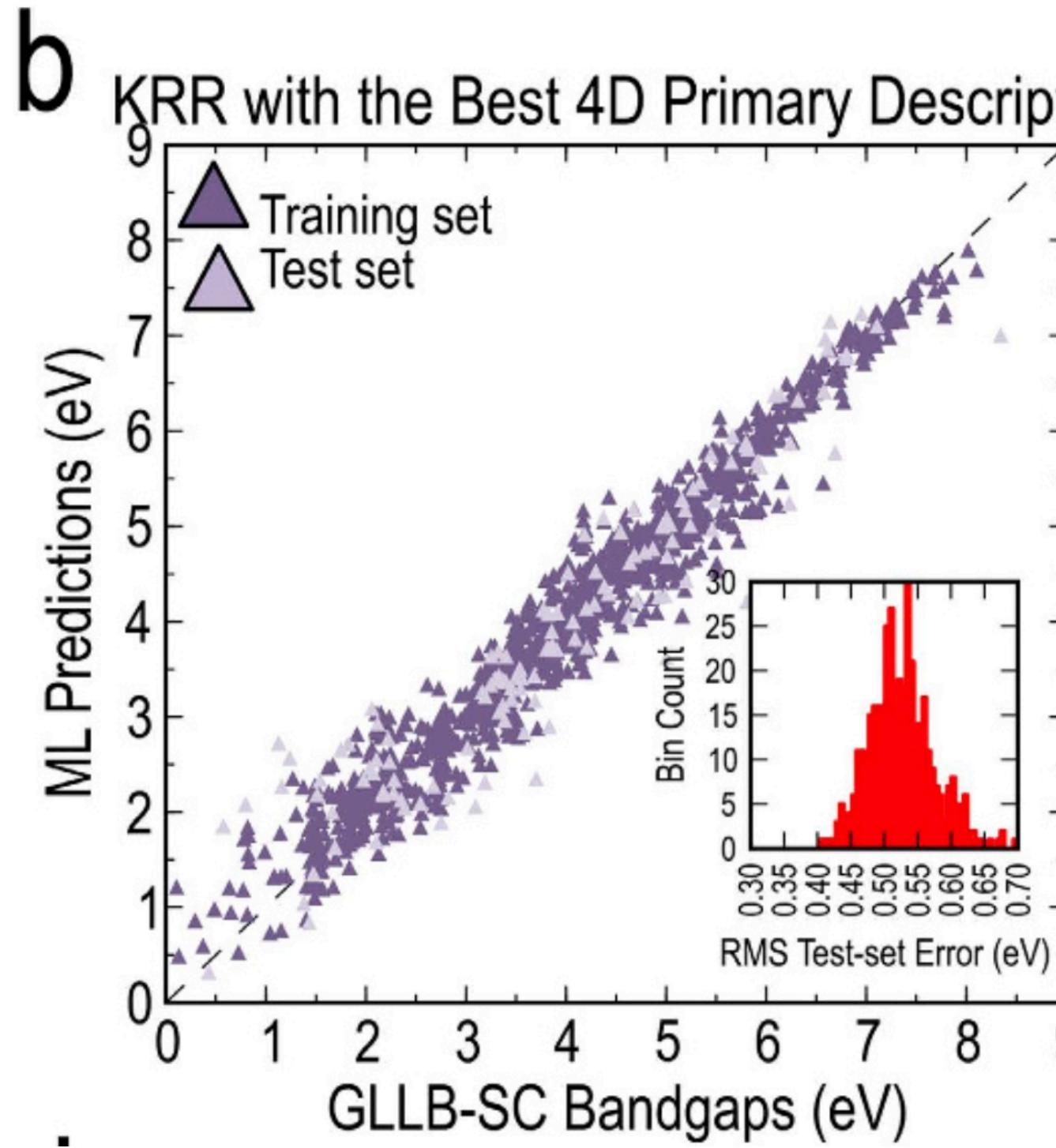
- dataset with ~1300 double perovskites
- DFT calculations with



Pilania et al., Sci. Rep. 6, 19375 (2016)

Input descriptors and model

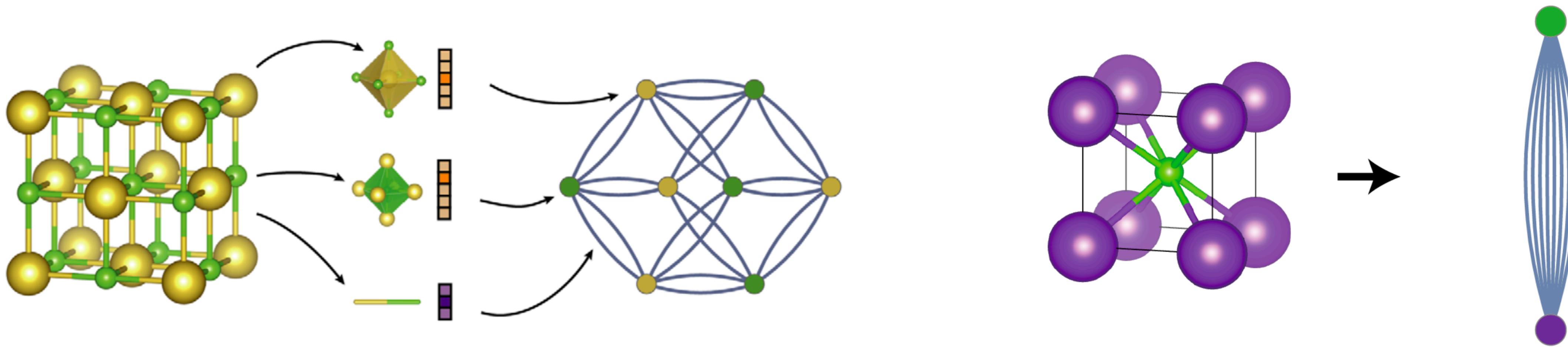
- 7 atomic features: electronegativity, ionization potential, highest occ./lowest unocc. KS level, s/p/d valence orbital radii; $4 \times 7 = 28$
- use Pearson correlation filter to reduce to max. 16 components
- use KRR with Gaussian kernel to predict band gap



Pilania et al., Sci. Rep. 6, 19375 (2016)

Crystal graph convolutional NN

- graph representation of a crystal: encode atomic information and bonding interactions
- undirected multigraph: nodes → atoms, edges → connections between atoms
- allow multiple edges between same pair of end nodes



Xie, Grossman, PRL 120, 145301 (2018)

Crystal graph convolutional NN

- feature vectors for each node and each edge
- convolutional NN: convolutional layers + pooling
- update feature vectors by ‘convolution’ with surrounding atoms
- pooling layer: overall feature vector of crystal
- combine with hidden layers for property prediction

Convolution function

- in general:

$$\boldsymbol{v}_i^{(t+1)} = \text{Conv}\left(\boldsymbol{v}_i^{(t)}, \boldsymbol{v}_j^{(t)}, \boldsymbol{u}_{(i,j)_k}\right), \quad (i, j)_k \in \mathcal{G}.$$

- simple convolution function:

$$\boldsymbol{v}_i^{(t+1)} = g\left[\left(\sum_{j,k} \boldsymbol{v}_j^{(t)} \oplus \boldsymbol{u}_{(i,j)_k}\right) \boldsymbol{W}_c^{(t)} + \boldsymbol{v}_i^{(t)} \boldsymbol{W}_s^{(t)} + \boldsymbol{b}^{(t)}\right]$$

Datasets and prediction

- materials project (<https://materialsproject.org>)
- 46744 materials
- atom features: group number, period number, electronegativity, covalent radius, valence electrons, first ionization energy, electron affinity block (s,p,d,f), atomic volume
- bond features: atom distance

Datasets and prediction

| Property | # of train data | Unit | MAE _{model} | MAE _{DFT} |
|------------------|-----------------|----------|----------------------|--------------------|
| Formation energy | 28 046 | eV/atom | 0.039 | 0.081–0.136 [28] |
| Absolute energy | 28 046 | eV/atom | 0.072 | ... |
| Band gap | 16 458 | eV | 0.388 | 0.6 [32] |
| Fermi energy | 28 046 | eV | 0.363 | ... |
| Bulk moduli | 2041 | log(GPa) | 0.054 | 0.050 [13] |
| Shear moduli | 2041 | log(GPa) | 0.087 | 0.069 [13] |
| Poisson ratio | 2041 | ... | 0.030 | ... |

