

Machine learning for molecules

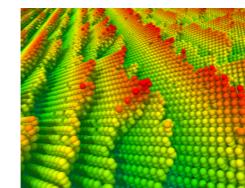
Gábor Csányi

Engineering Laboratory
 UNIVERSITY OF
CAMBRIDGE

Outline

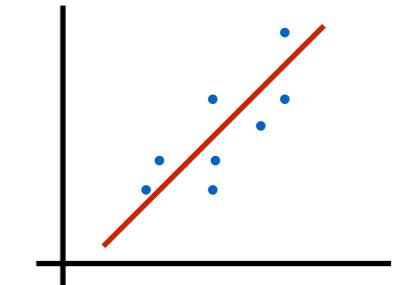
- Interatomic potentials

- Multi scale modeling, molecular dynamics



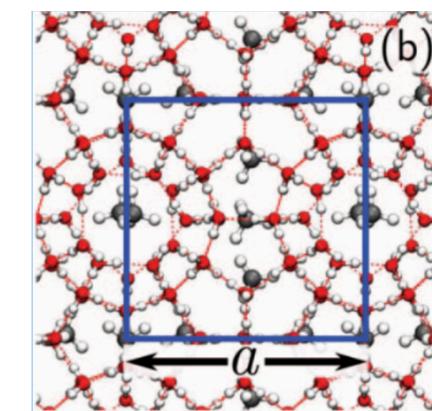
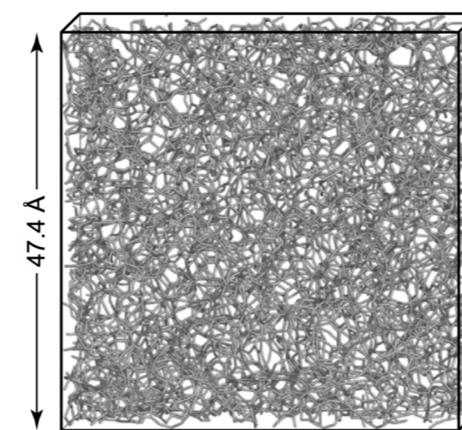
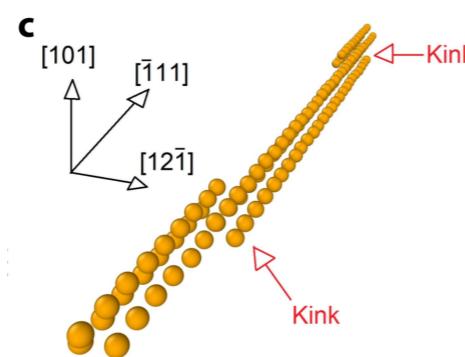
- Fitting functions in high dimensional spaces

- Descriptors, regression, regularisation

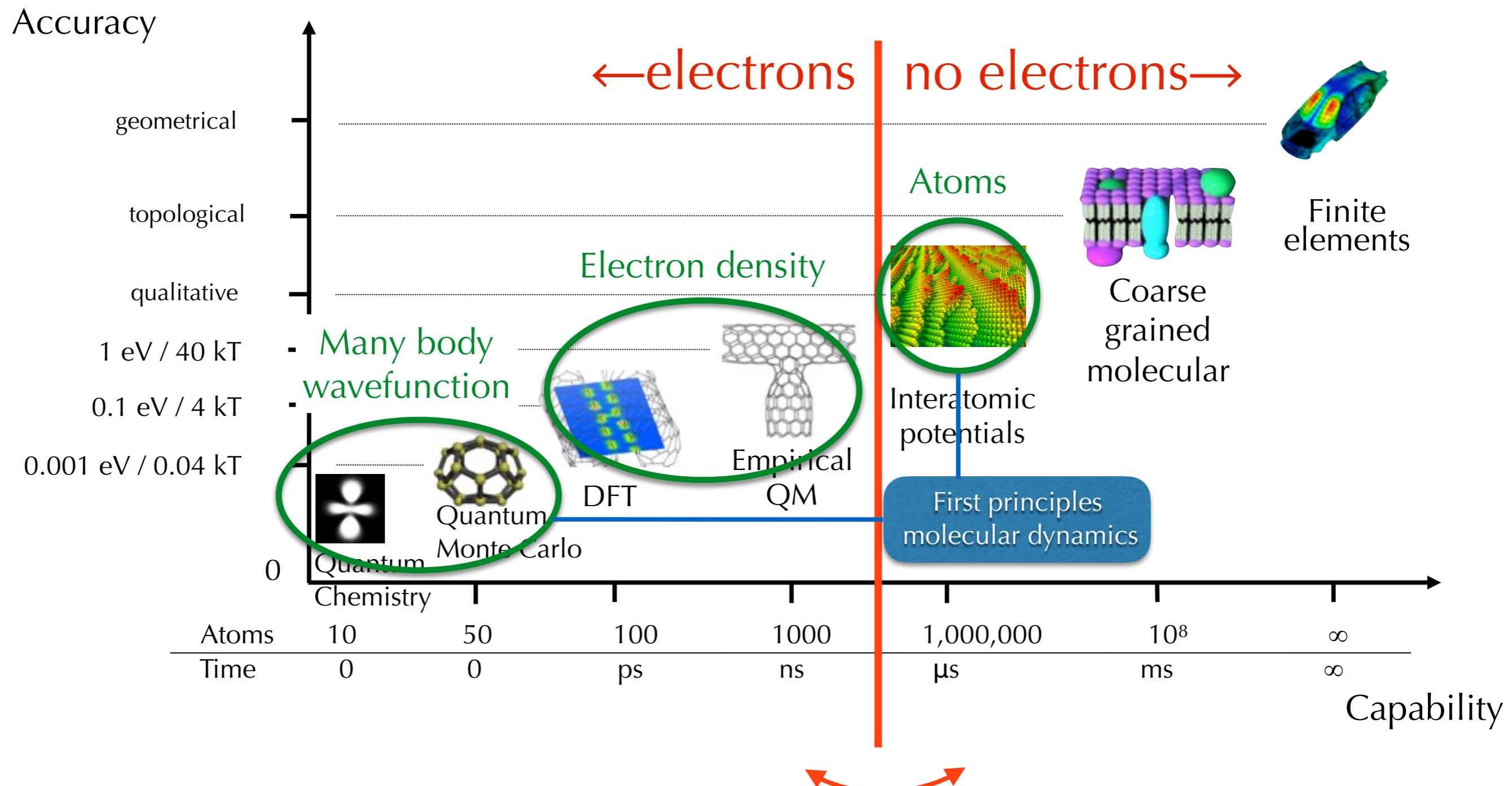


- Applications of machine learned models

- Long-standing materials problems, e.g. metal mechanics, structure and dynamics of amorphous and molecular materials

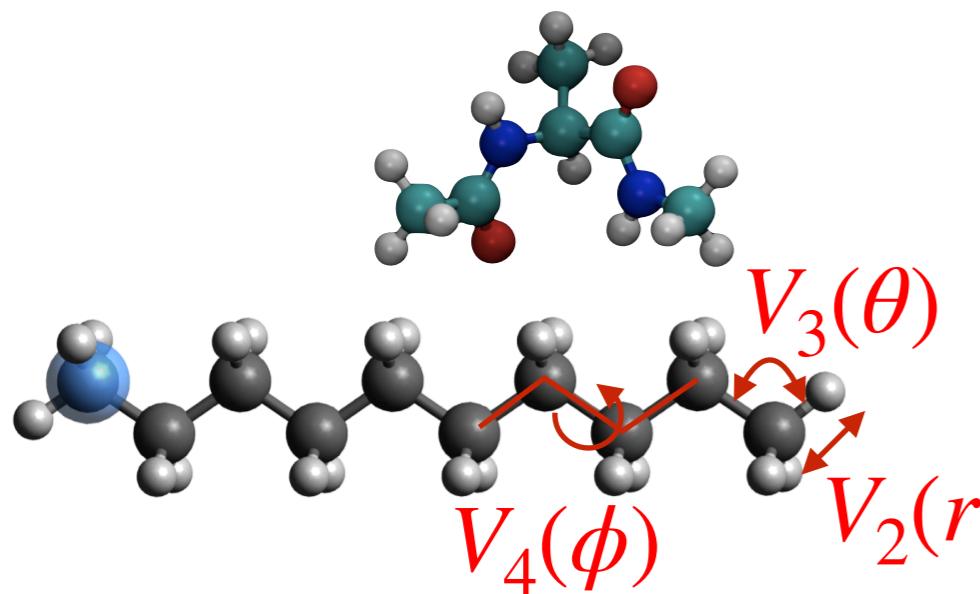


Multiple scales of materials modelling



GAP in accuracy and speed

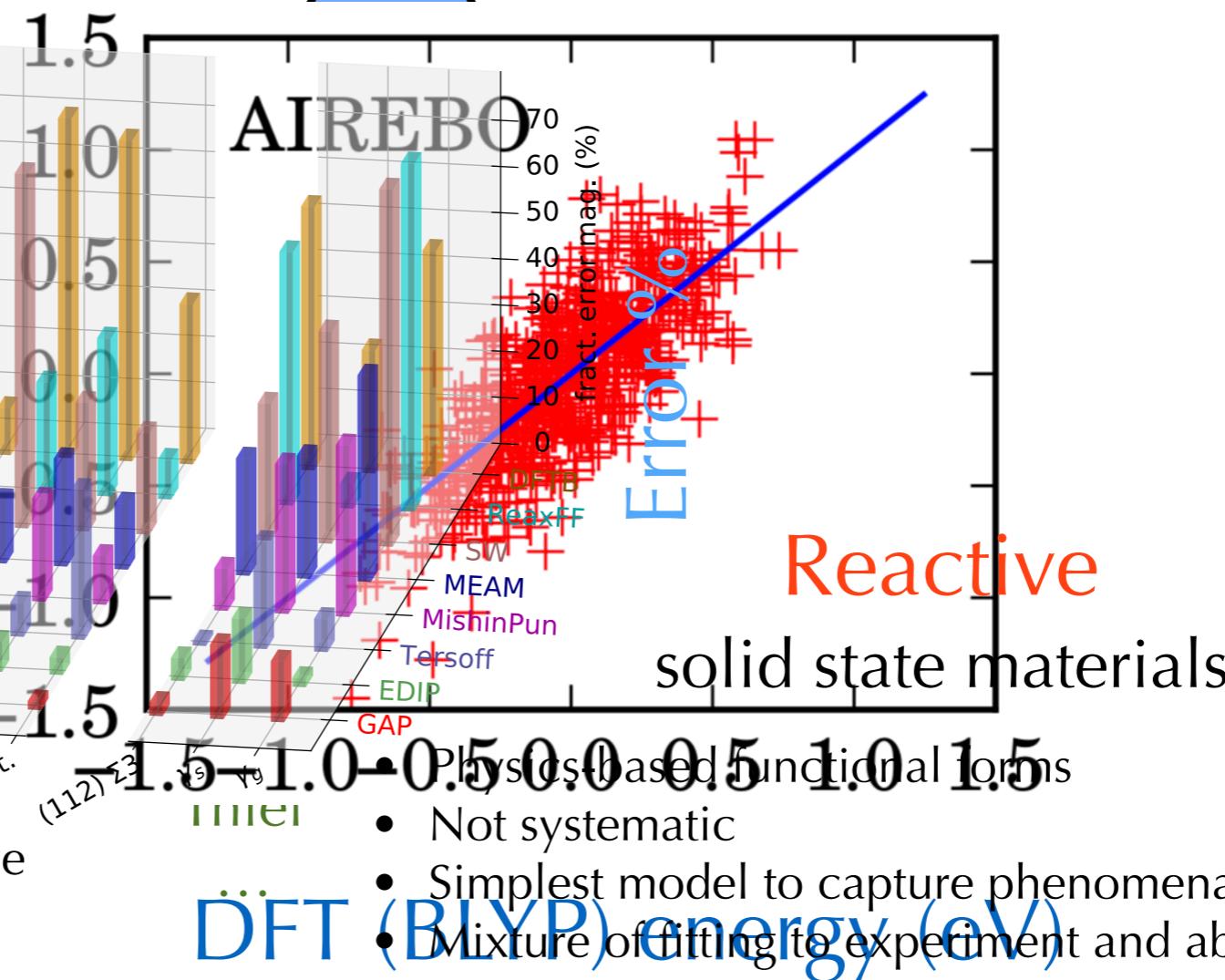
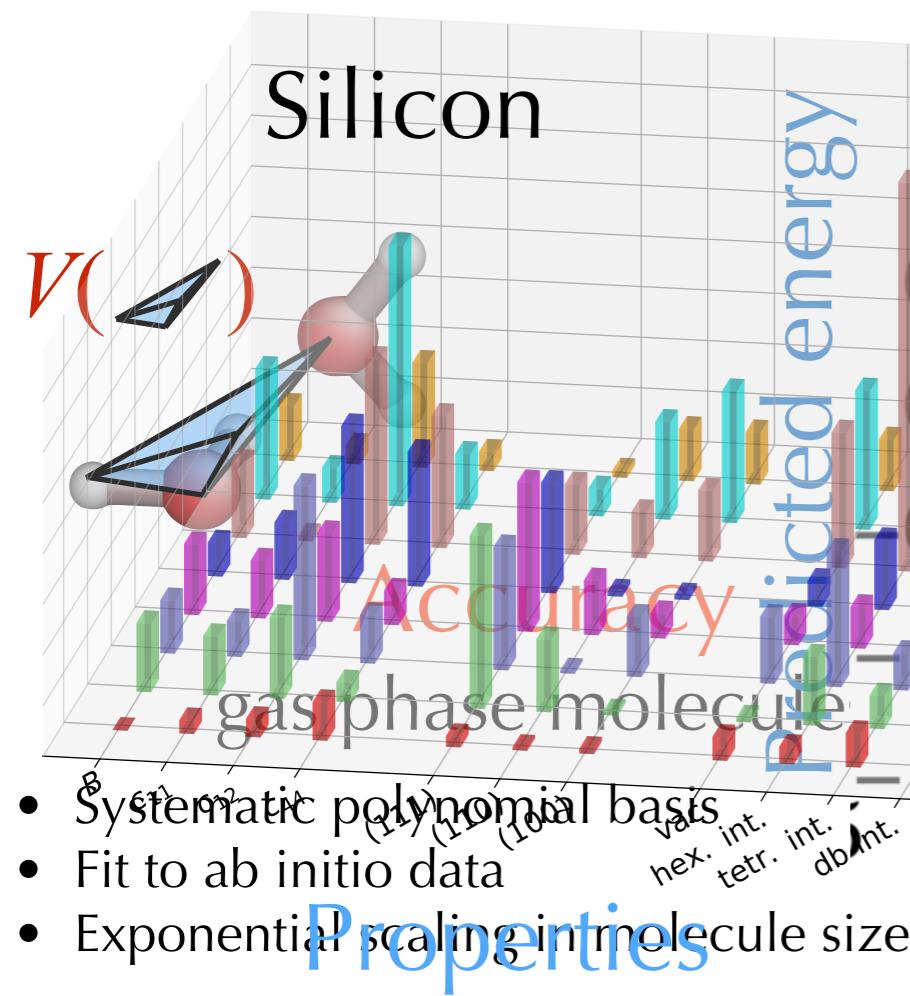
Interatomic potentials (force fields)



Transferability
organic force fields

- Accurate for small displacements
- Typically no reactions (some exceptions)
- Ultimate fit is to macroscopic experimental properties

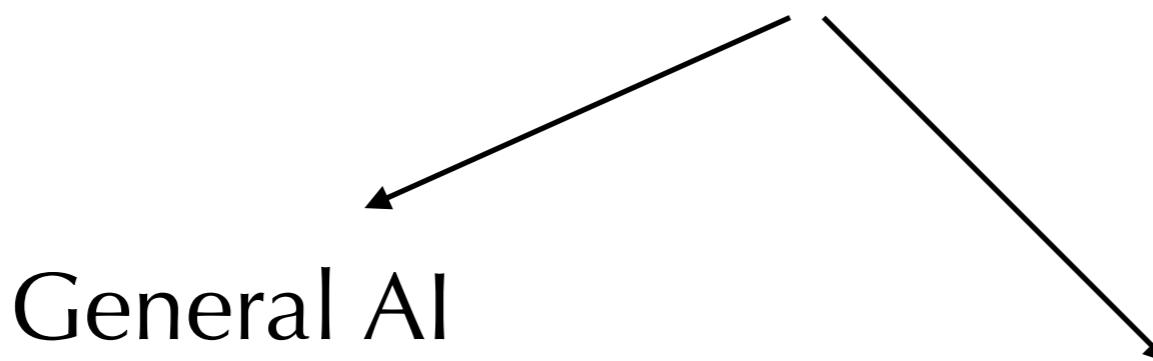
AMBER
CHARMM
COMPASS
OPLS
...



$V(\sum \rho(r))$

Tersoff
EAM
BOP
...

Artificial Intelligence (AI)



1950s: Computers! Intelligent robots?



1960s: Expert systems and databases



1980s: Artificial Neural Networks ! Intelligent robots?

1990s: Machine learning, kernel methods, high dimensional regression

2000s: Deep Neural Networks ! ... ?



“shallow” : 1 layer (same as kernels)

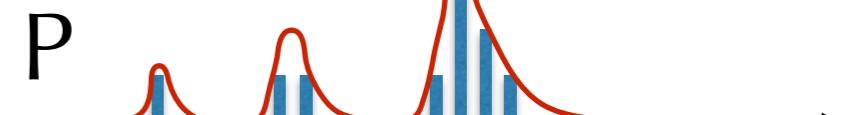
“deep” : many layers

$$f(x) = \sum g_i \left(w_i^{(1)}, \sum h_j \left(w_j^{(2)}, \sum \dots \right) \right)$$

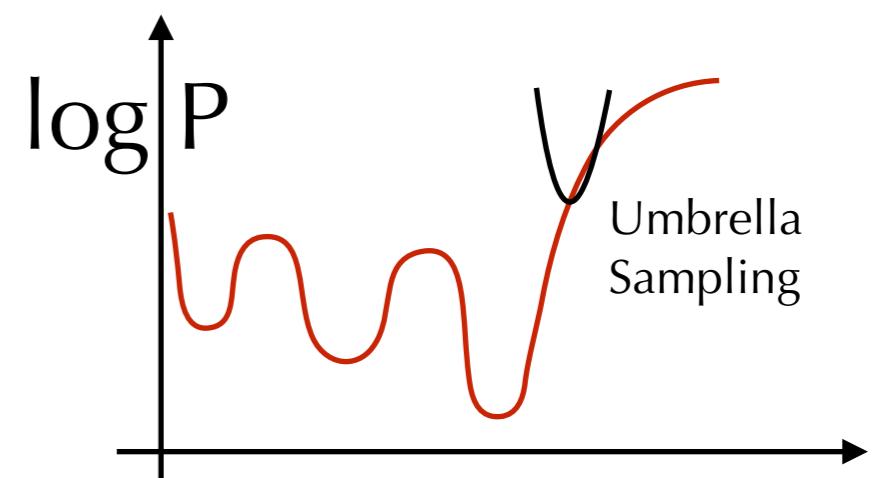
Machine Learning

sed
ion)

Unsupervised
(Probability density estimation)



Normalisation → global problem



“energy” vs “free energy”

Regression

More stringent

Properties of the input

Less stringent

- Linear regression

$$f(\mathbf{x}) = \sum_i^N c_i \mathbf{b}_i(\mathbf{x})$$

basis functions \mathbf{b}

- Kernel regression

$$f(\mathbf{x}) = \sum_i^N c_i K(\mathbf{x}, \mathbf{x}_i)$$

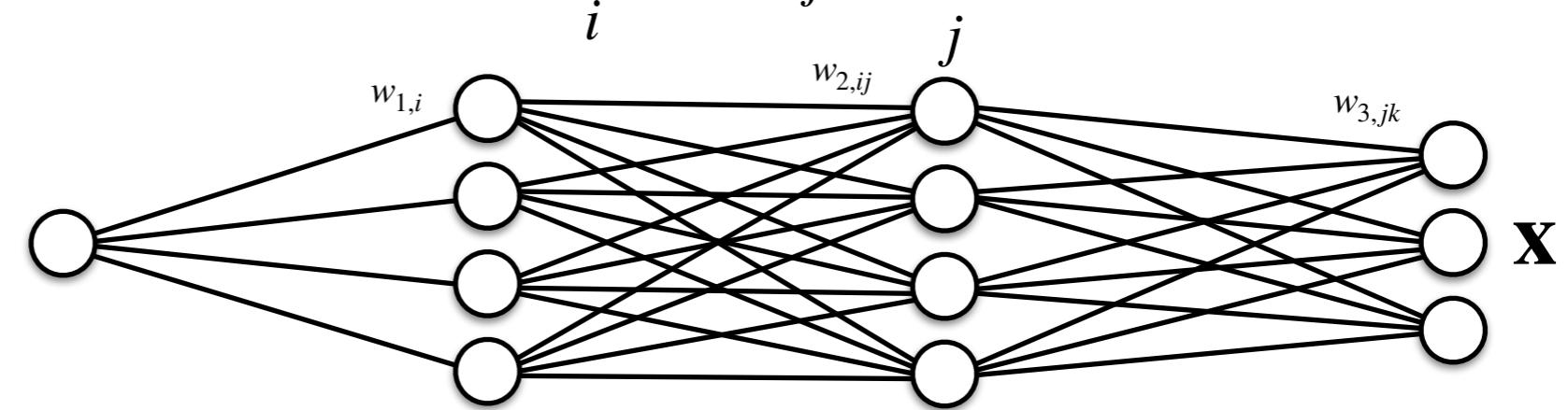
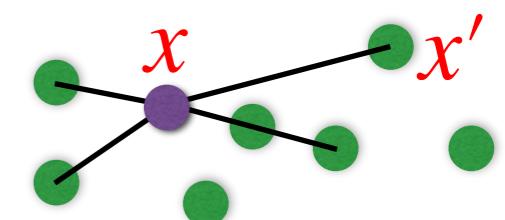
similarity kernel
 $K(\mathbf{x}, \mathbf{x}')$

- Also linear but in transformed space
- Equivalent to neural network with 1 hidden layer
- Also known as “Gaussian process regression”

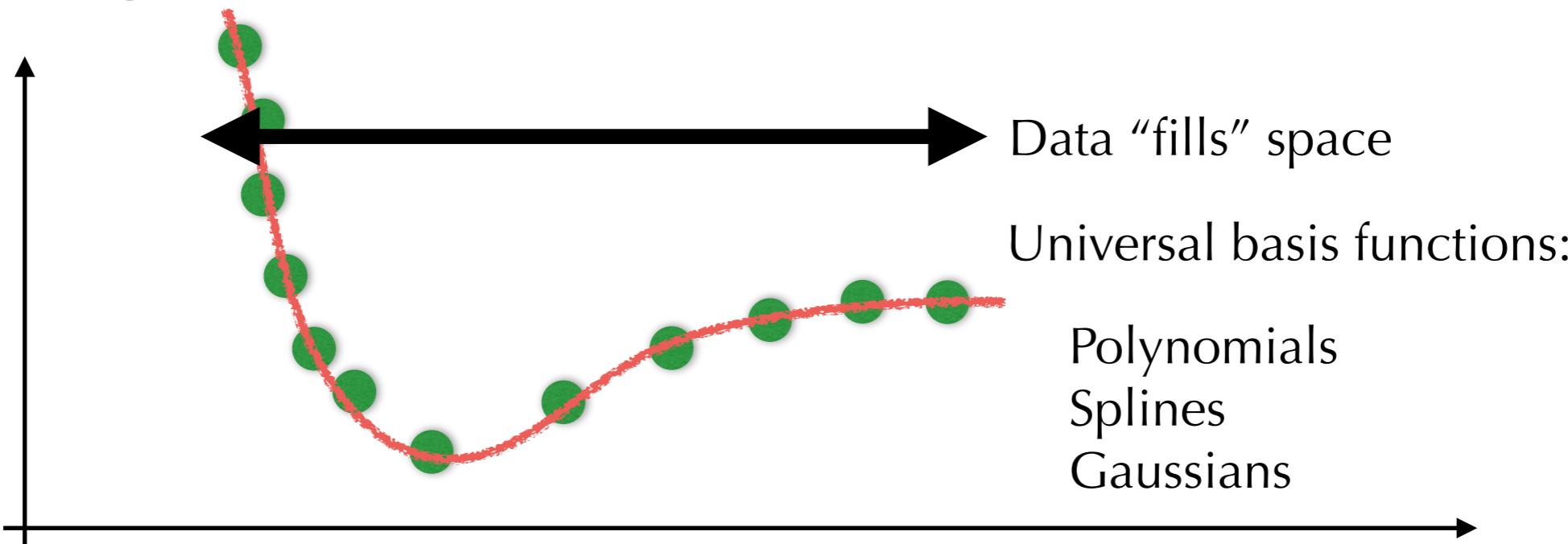
- Nonlinear regression

- >1 hidden layer neural networks

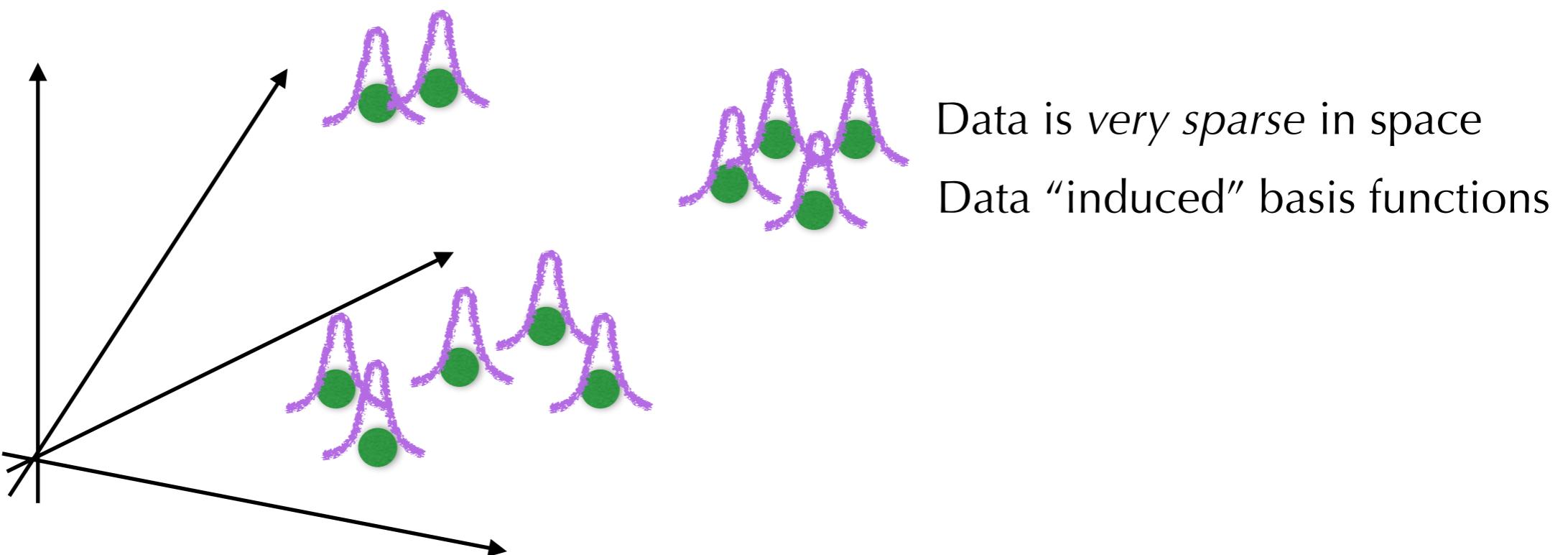
$$f(\mathbf{x}) = h(b_0 + \sum_i^{N_1} w_{1,i} h(b_{1,i} + \sum_j^{N_2} w_{2,ij} h(b_{2,j} + \sum_k^{\dim(\mathbf{x})} w_{3,jk} x_k)))$$



Fitting functions in few dimensions



Fitting functions in many dimensions



Linear regression with basis functions

Atomic configurations \mathcal{A}, \mathcal{B}

Kernel basis function $K(\mathcal{A}, \mathcal{B})$

N observations \mathbf{y}

M representative configs

$$f(\mathcal{A}) = \sum_{\mathcal{B} \in M} x_{\mathcal{B}} K(\mathcal{A}, \mathcal{B})$$

$$f(\mathcal{A}) = \mathbf{K}_{\mathcal{A}M} \mathbf{x}_M$$

$$\begin{bmatrix} K(\mathcal{A}_1, \mathcal{B}_1) & K(\mathcal{A}_1, \mathcal{B}_2) & \dots \\ K(\mathcal{A}_2, \mathcal{B}_1) & K(\mathcal{A}_2, \mathcal{B}_2) & \dots \\ \vdots & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} f(\mathcal{A}_1) \\ f(\mathcal{A}_2) \\ \vdots \end{bmatrix}$$

Design
matrix: \mathbf{K}

unknowns: \mathbf{x}

data: \mathbf{y}

Linear least squares problem $\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{K}\mathbf{x} - \mathbf{y}\|^2$ \mathbf{K} is $M \times N$

Solution

Suppose $M = N \quad \mathbf{K}^T = \mathbf{K}$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{K}^T - \mathbf{y}^T)(\mathbf{K}\mathbf{x} - \mathbf{y}) = 0$$

$$\mathbf{K}^T \mathbf{K}\mathbf{x} - \mathbf{K}^T \mathbf{y} = 0$$

$$\mathbf{K}\mathbf{x} - \mathbf{y} = 0$$

$$\mathbf{x} = \mathbf{K}^{-1}\mathbf{y}$$

But \mathbf{K} is likely not invertible (e.g. near degenerate basis)

$$\mathbf{x} = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}\mathbf{y}$$

But this is not what we do. Regularisation:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{K}\mathbf{x} - \mathbf{y}\|_{\Lambda^{-1}}^2 + \|\mathbf{x}\|_{\mathbf{K}}^2$$

Regularised solution

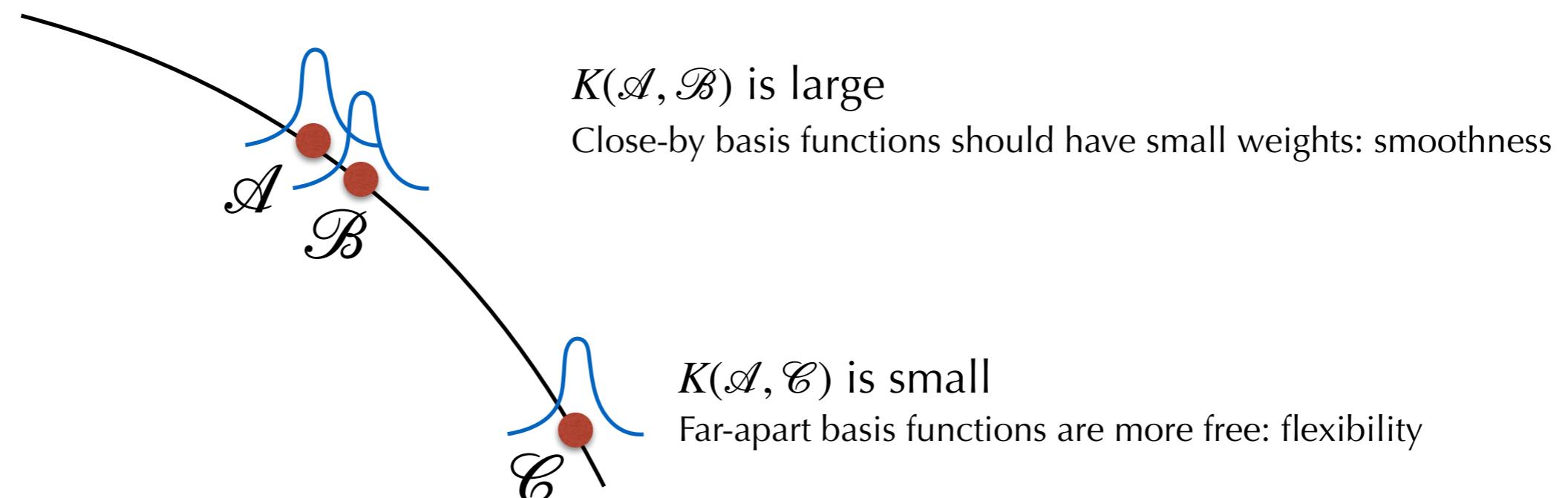
$$\min_{\mathbf{x}} \quad (\mathbf{K}\mathbf{x} - \mathbf{y})^T \Lambda^{-1} (\mathbf{K}\mathbf{x} - \mathbf{y}) + \mathbf{x}^T \mathbf{K}\mathbf{x}$$

$$\mathbf{K}^T \Lambda^{-1} \mathbf{K}\mathbf{x} - \mathbf{K}^T \Lambda^{-1} \mathbf{y} + \mathbf{K}\mathbf{x} = 0$$

$$\mathbf{K}(\Lambda^{-1} \mathbf{K} + \mathbf{I})\mathbf{x} - \mathbf{K}\Lambda^{-1} \mathbf{y} = 0$$

$$\mathbf{x} = (\mathbf{K} + \Lambda)^{-1} \mathbf{y}$$

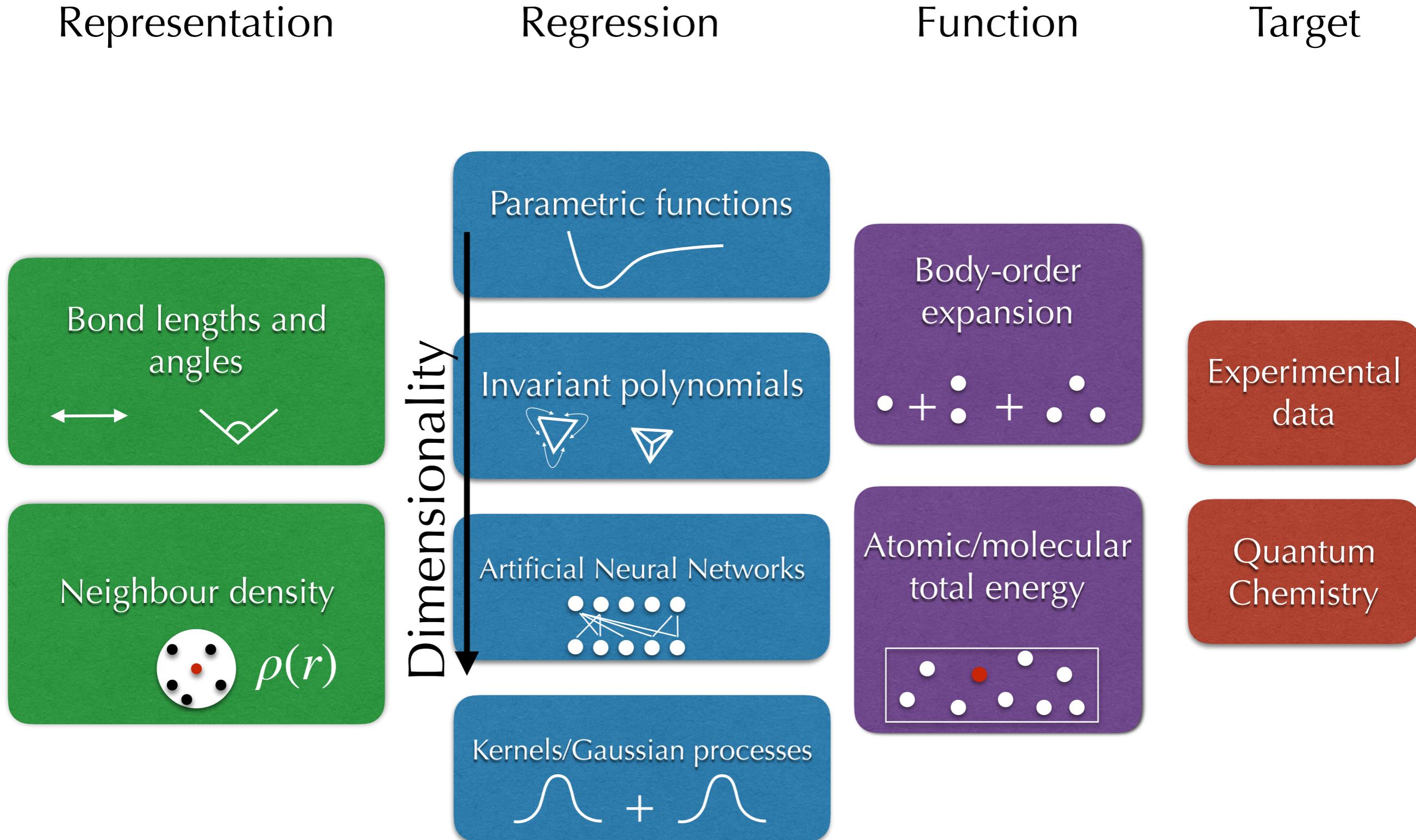
But why is adding $\mathbf{x}^T \mathbf{K}\mathbf{x}$ a good idea?



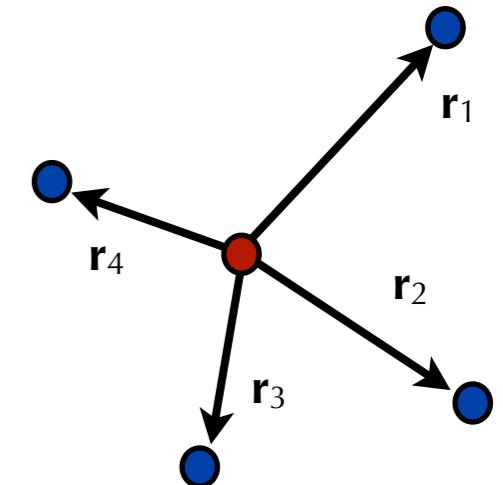
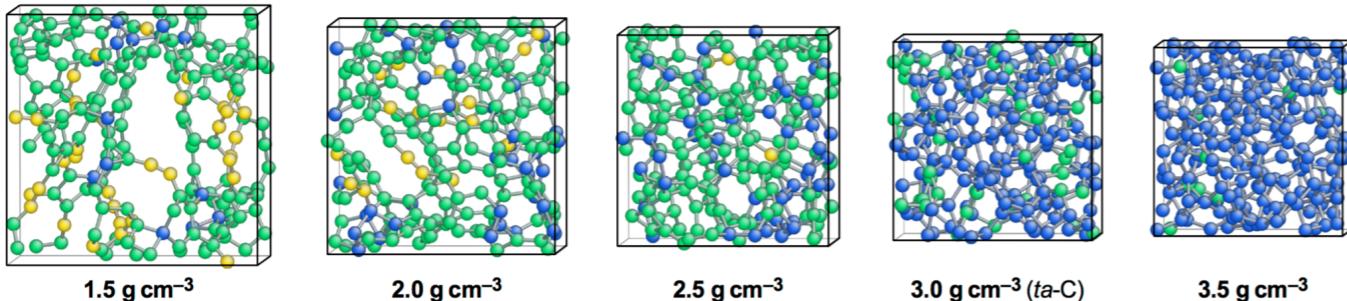
For $M < N$

$$\mathbf{x}_M^* = (\mathbf{K}_{MM} + \mathbf{K}_{MN} \Lambda^{-1} \mathbf{K}_{NM})^{-1} \mathbf{K}_{MN} \Lambda^{-1} \mathbf{y}_N$$

Anatomy of an interatomic potential



Building a potential for materials: a representation problem

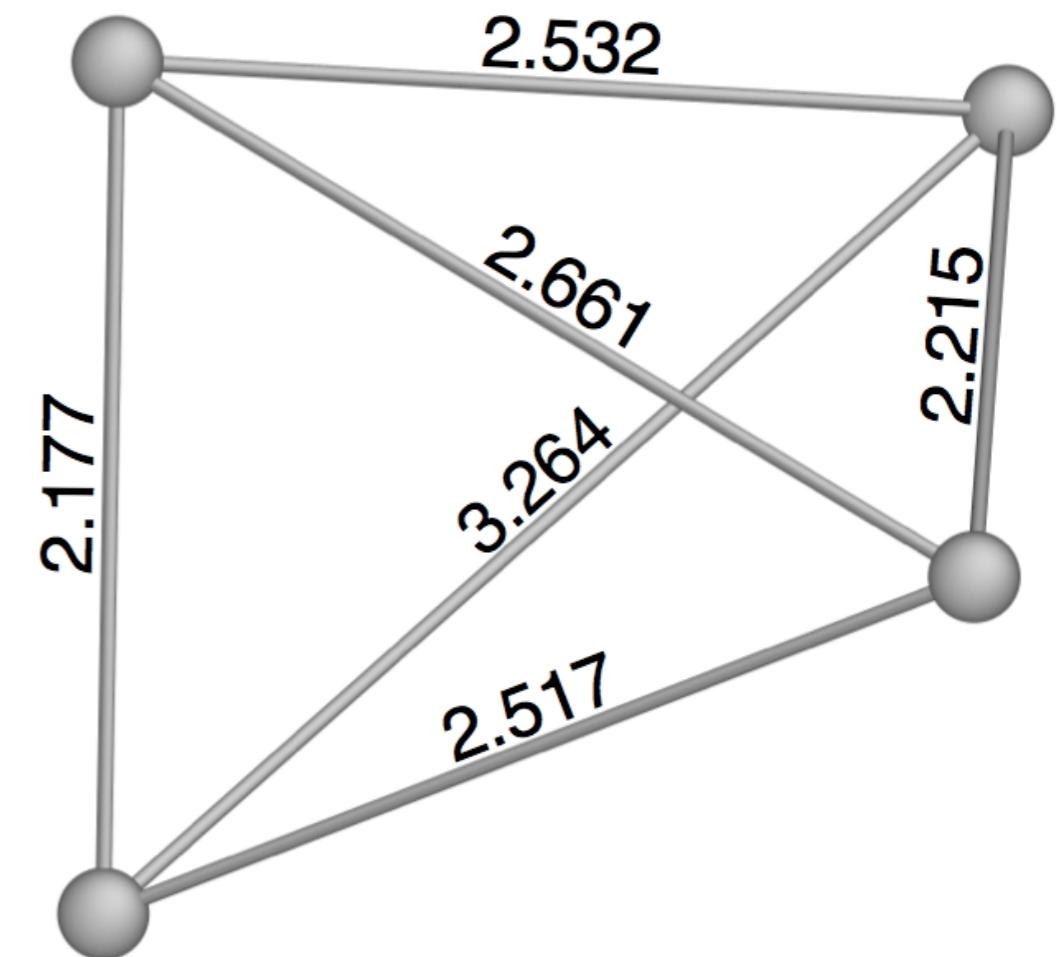
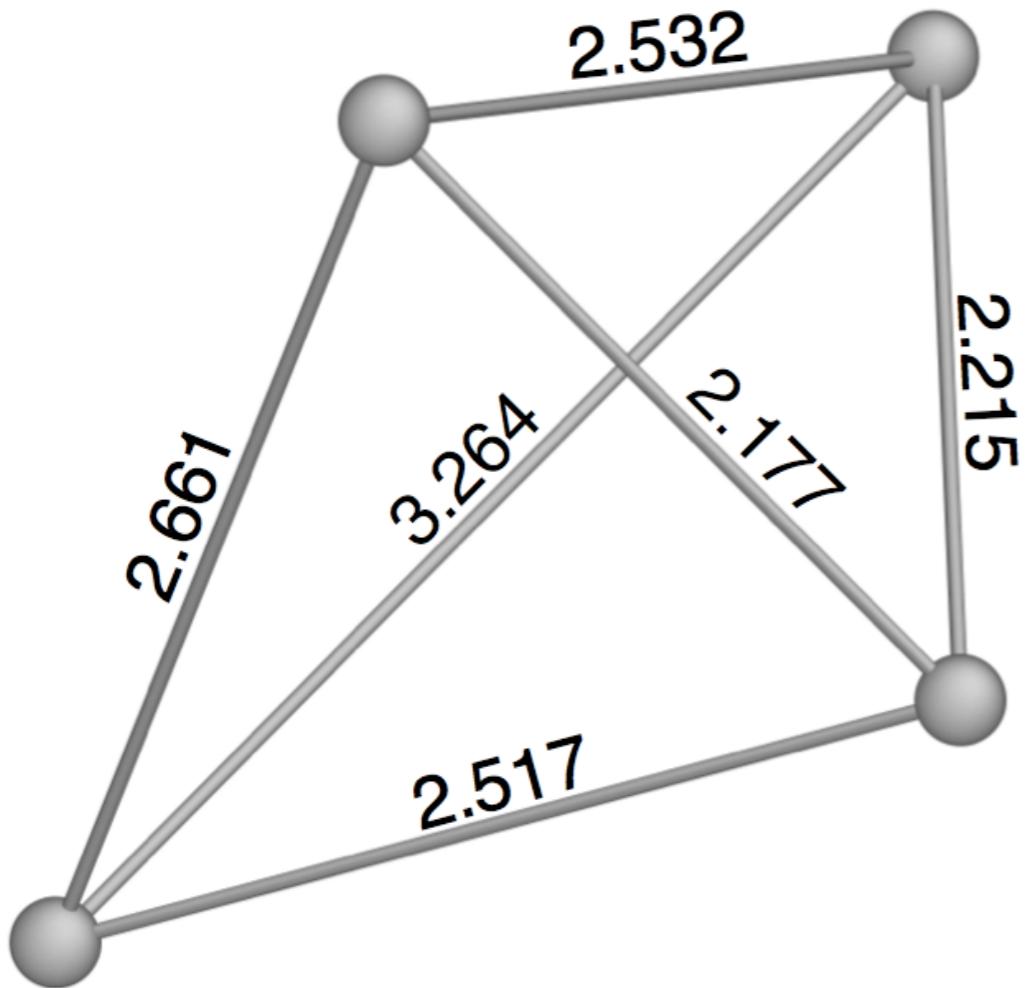


- Weyl matrix

$$\begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{r}_1 & \mathbf{r}_1 \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_1 \cdot \mathbf{r}_N \\ \mathbf{r}_2 \cdot \mathbf{r}_1 & \mathbf{r}_2 \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_2 \cdot \mathbf{r}_N \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{r}_N \cdot \mathbf{r}_1 & \mathbf{r}_N \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_N \cdot \mathbf{r}_N \end{bmatrix}$$

- Problems:
 - not permutation invariant
 - fixed number of neighbours (no reactions!)

Drop atom ordering ?



Not unique!

Atomic neighbour density function

$$\varepsilon(\mathbf{r}_1 - \mathbf{r}_i, \mathbf{r}_2 - \mathbf{r}_i, \dots) \equiv \varepsilon[\rho_i(\mathbf{r})]$$

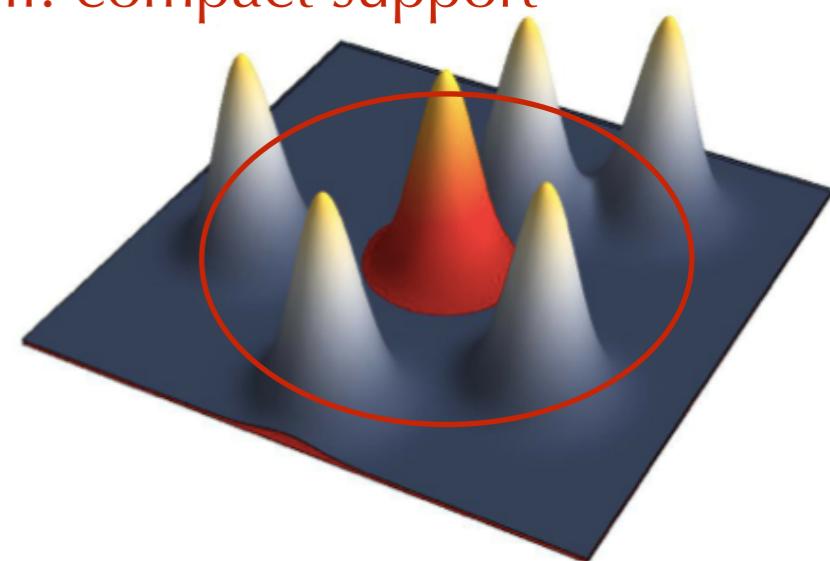
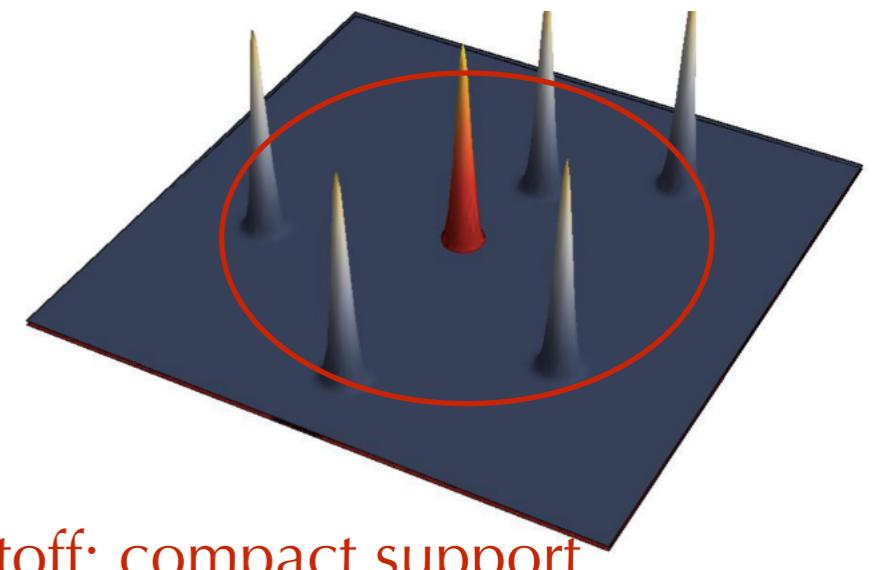
$$\rho_i(\mathbf{r}) = \sum_j \delta(\mathbf{r} - \mathbf{r}_{ij}) f_{\text{cut}}(|\mathbf{r}_{ij}|)$$

Or, smear the atoms a bit:

$$\rho_i(\mathbf{r}) = \sum_j \exp\left(-|\mathbf{r} - \mathbf{r}_{ij}|^2/2\sigma^2\right) f_{\text{cut}}(|\mathbf{r}_{ij}|)$$

- Translation ✓
- Permutation ✓
- Continuous ✓
- Now need rotationally invariants of $\rho(\mathbf{r})$

Project onto rotationally invariant basis sets:



- Behler-Parrinello symmetry functions
- Bispectrum (GAP 2010, SNAP)
- Wavelets (Mallat's scattering transforms)
- Histograms of lengths and angles (MBTR)

SOAP: Smooth Overlap of Atomic Positions

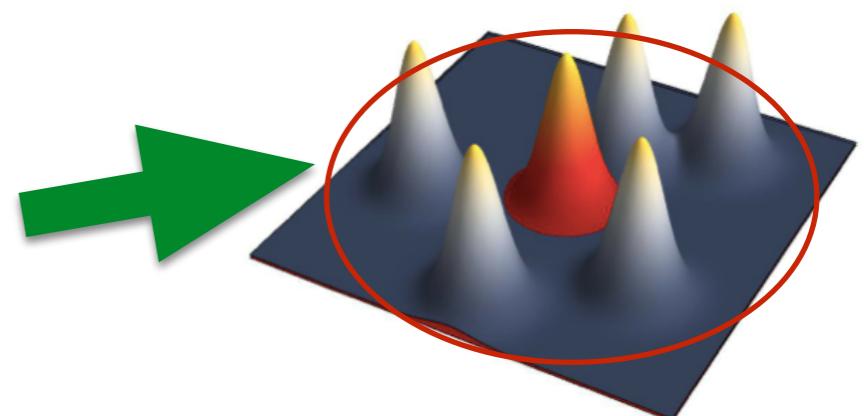
$$\rho_i(\mathbf{r}) = \sum_j \exp\left(-|\mathbf{r} - \mathbf{r}_{ij}|^2/2\sigma^2\right) f_{\text{cut}}(|\mathbf{r}_{ij}|) = \sum_{nlm} c_{nlm}^{(i)} g_n(r) Y_{lm}(\hat{\mathbf{r}})$$

- Overlap integral

$$S(\rho_i, \rho_{i'}) = \int \rho_i(\mathbf{r}) \rho_{i'}(\mathbf{r}) d\mathbf{r},$$



- Integrate over all 3D rotations:



cutoff: compact support

$$k(\rho_i, \rho_{i'}) = \int \left| S(\rho_i, \hat{R}\rho_{i'}) \right|^2 d\hat{R} = \int d\hat{R} \left| \int \rho_i(\mathbf{r}) \rho_{i'}(\hat{R}\mathbf{r}) d\mathbf{r} \right|^2$$

- After LOTS of algebra: SOAP kernel

$$k(\rho_i, \rho_{i'}) = \sum_{n,n',l} p_{nn'l}^{(i)} p_{nn'l}^{(i')} = \mathbf{p}^{(i)} \cdot \mathbf{p}^{(i')}$$

$$p_{nn'l} = \mathbf{c}_{nl}^\dagger \cdot \mathbf{c}_{n'l}$$

set of invariants:

“soap vector” or “power spectrum”
 \propto Steinhardt Q_l (with $g_n = 1$)

$$\mathbf{K}_{ij} \propto |k(\rho_i, \rho_j)|^\xi$$



Extensions

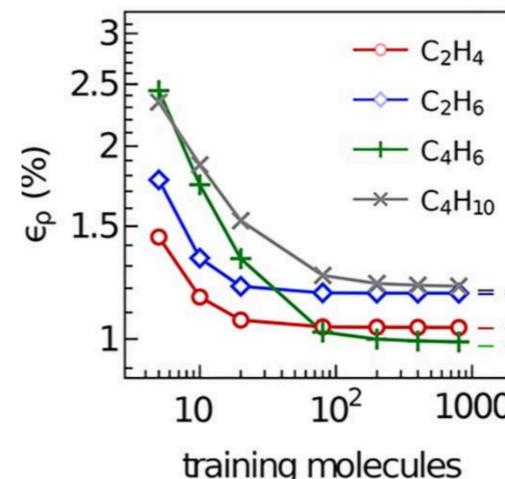
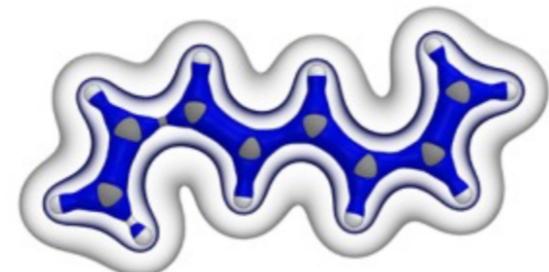
- Michele Ceriotti (2018-19)

- Many molecular properties, including NMR chemical shifts
- Kernel for learning tensor properties

$$k^\lambda(\rho, \rho') = \int d\hat{R} \mathbf{D}^\lambda(\hat{R}) \left| \int d\mathbf{r} \rho(r) \rho'(\hat{R}\mathbf{r}) \right|^2$$

\mathbf{D} : Wigner matrix

- Machine learning the electron density



- Ralf Drautz (2019)

- Explicit connection between all neighbour density based descriptors
- Further generalisation of invariants, body ordering interactions

Pair potential

$$B^{(1)} = c_{n00}$$

Soap vector (power spectrum)

$$B^{(2)} = \sum_m c_{n_1 lm}^\dagger c_{n_2 lm}$$

Bispectrum

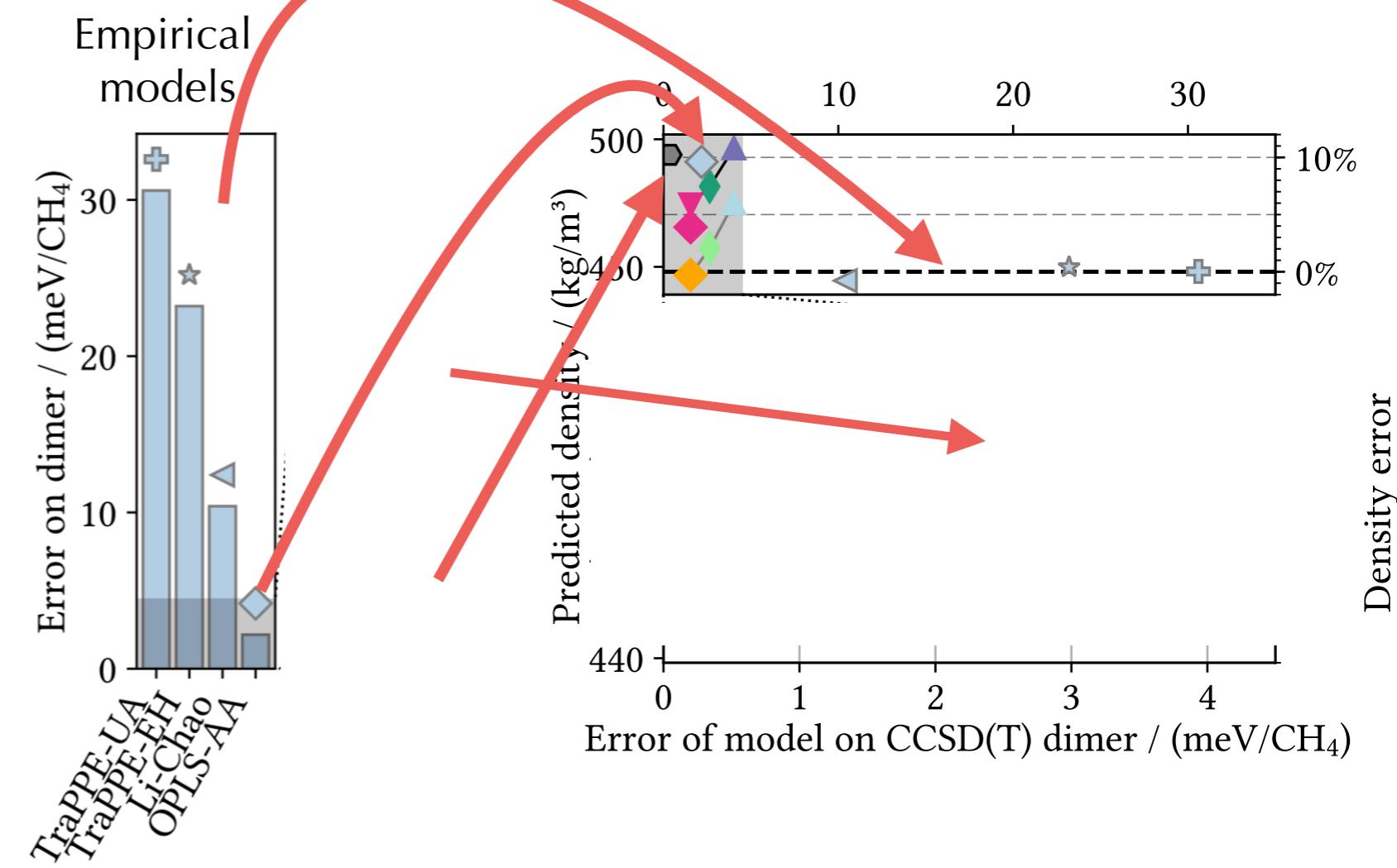
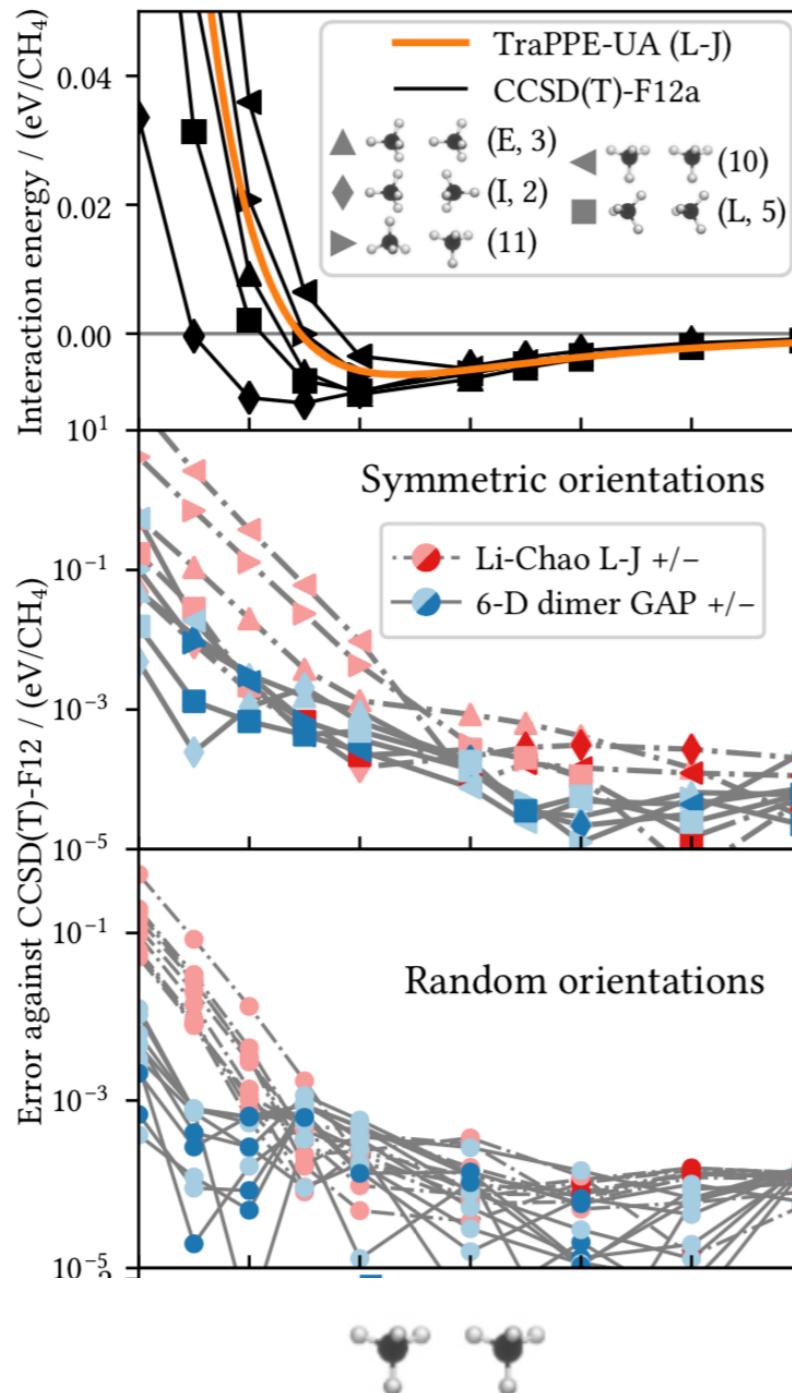
$$B^{(3)} = \sum_{m_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} c_{n_1 l_1 m_1} c_{n_2 l_2 m_2} c_{n_3 l_3 m_3}$$

$$B^{(4)} = \sum_{m_1 m_2 m_3 m_4} \begin{bmatrix} l_1 & l_2 & l_3 & l_4 \\ m_1 & m_2 & m_3 & m_4 \end{bmatrix} c_{n_1 l_1 m_1} c_{n_2 l_2 m_2} c_{n_3 l_3 m_3} c_{n_4 l_4 m_4}$$

$B^{(5)}, B^{(6)}$ etc

Fluid methane

- Dominated by weak dispersion interactions
- 2-body enough or need many-body effects?
- What level of quantum mechanics is required?



RETURN TO ISSUE

< PREV

ARTICLE

NEXT >

Equation of State of Fluid Methane from First Principles with Machine Learning Potentials

Max Veit*, Sandeep Kumar Jain, Satyanarayana Bonakala, Indranil Rudra, Detlef Hohl and Gábor Csányi

Cite This: *J. Chem. Theory Comput.* 2019, 15, 4,

2574-2586

Publication Date: February 22, 2019

<https://doi.org/10.1021/acs.jctc.8b01242>

Copyright © 2019 American Chemical Society

RIGHTS & PERMISSIONS

Article Views

Altmetric

Citations

Share Add to

393

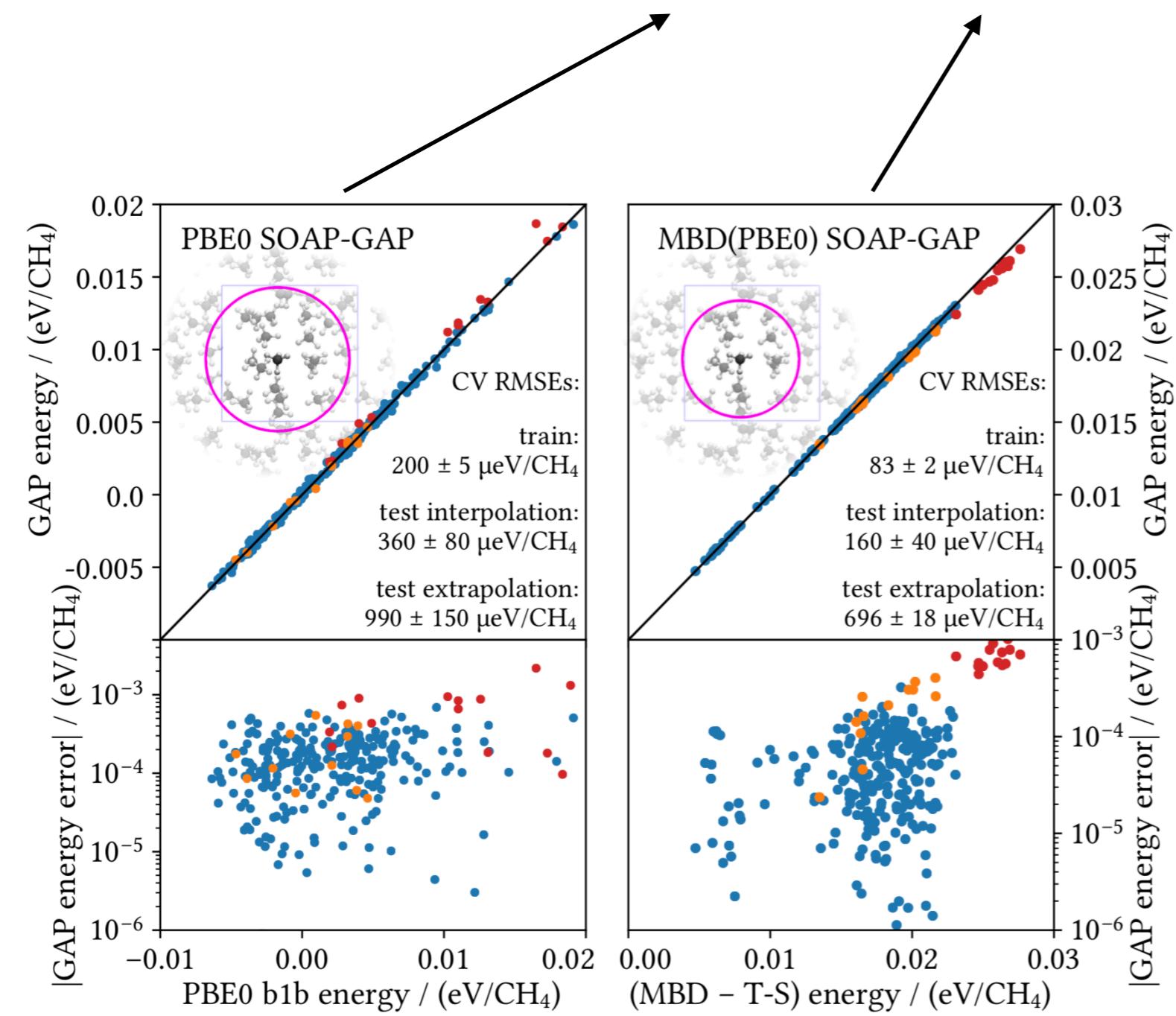
7

1

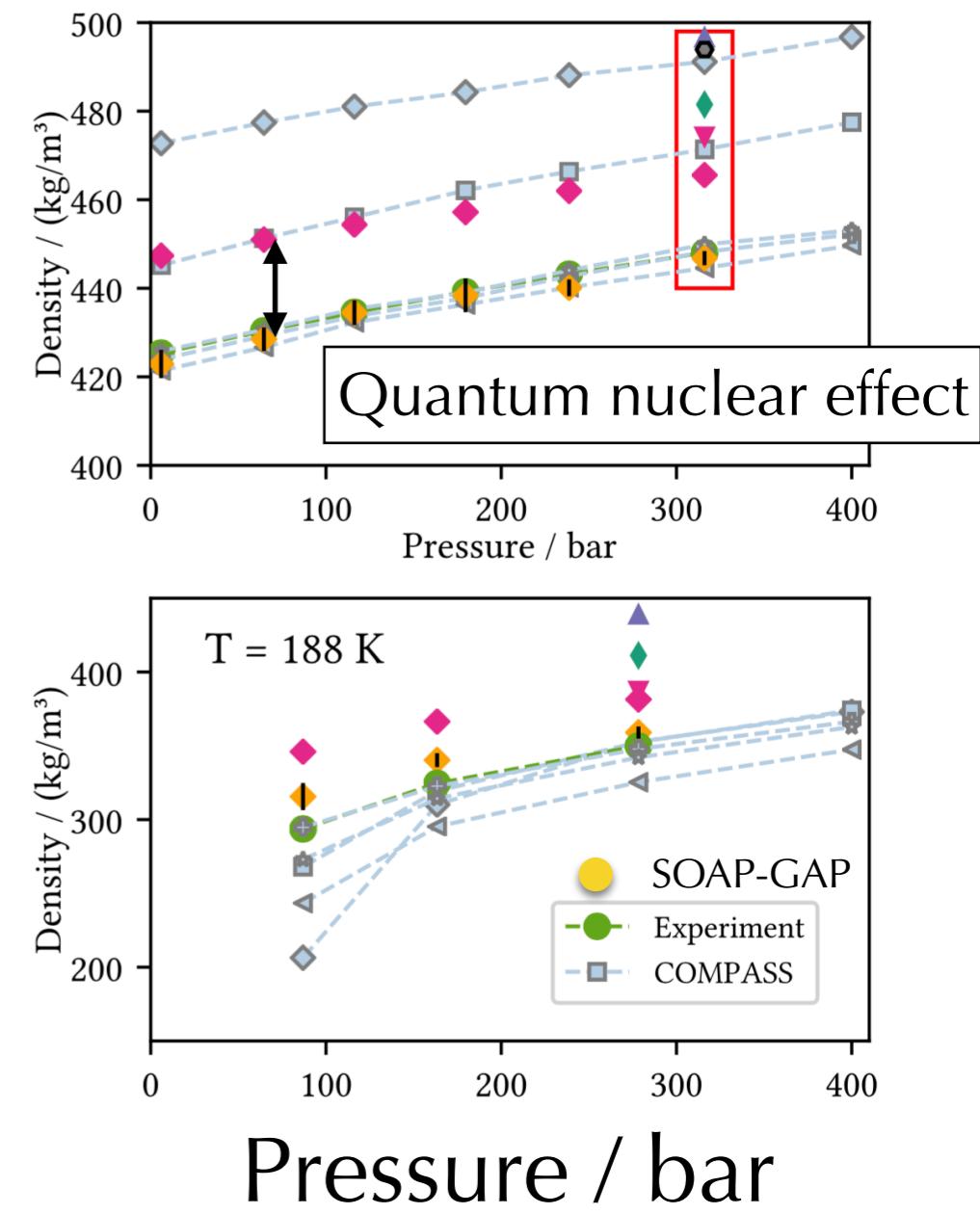
LEARN ABOUT THESE METRICS

Many-body SOAP-GAP model

$$E_{\text{potential}} = N E_{\text{CH}_4} + E_{\text{short-range}} + E_{\text{dispersion}}$$



Density prediction



General kernel for energy of molecules

www.nature.com/scientificdata

SCIENTIFIC DATA

110110
011101
11011110
011101101

OPEN

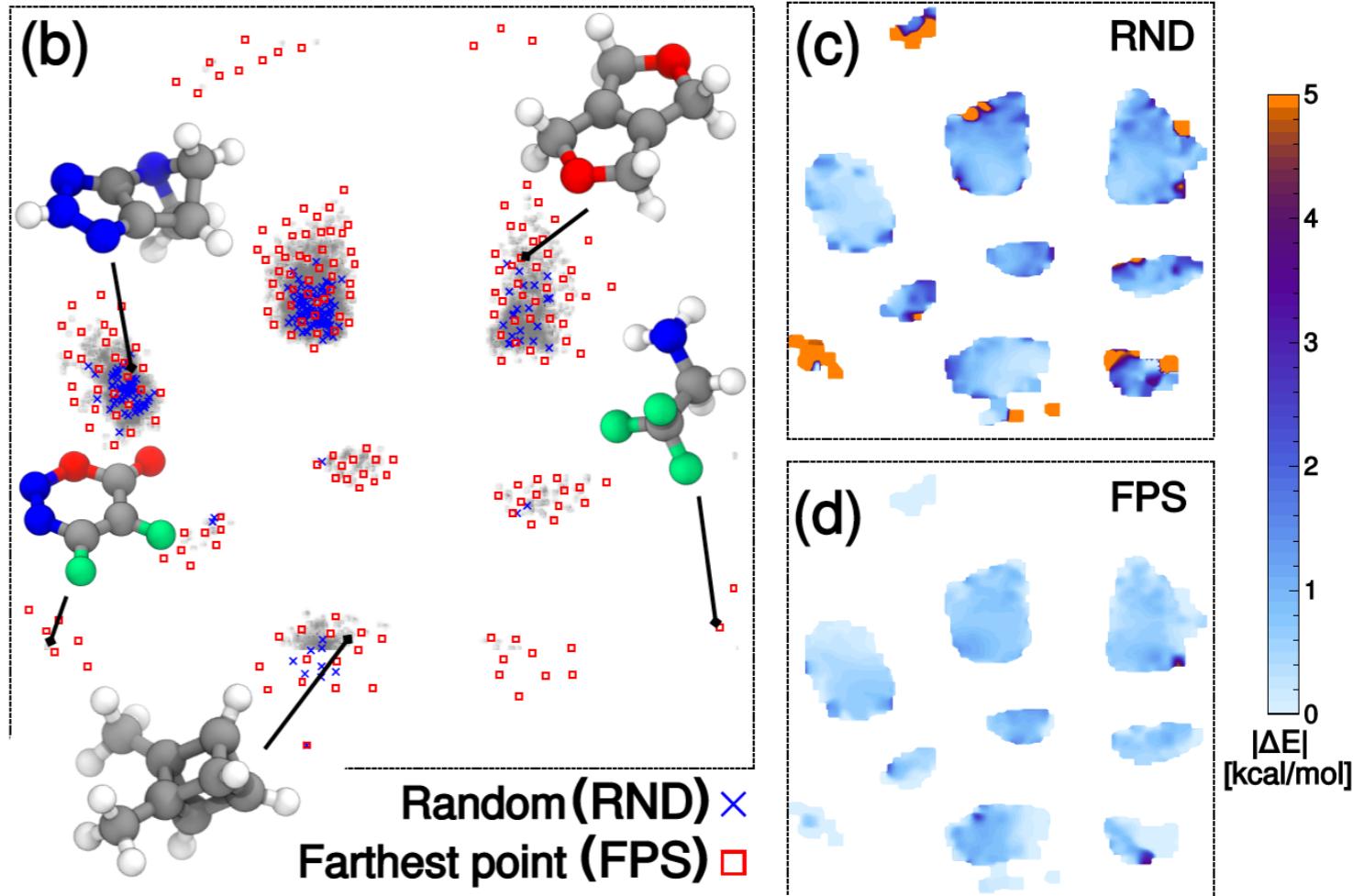
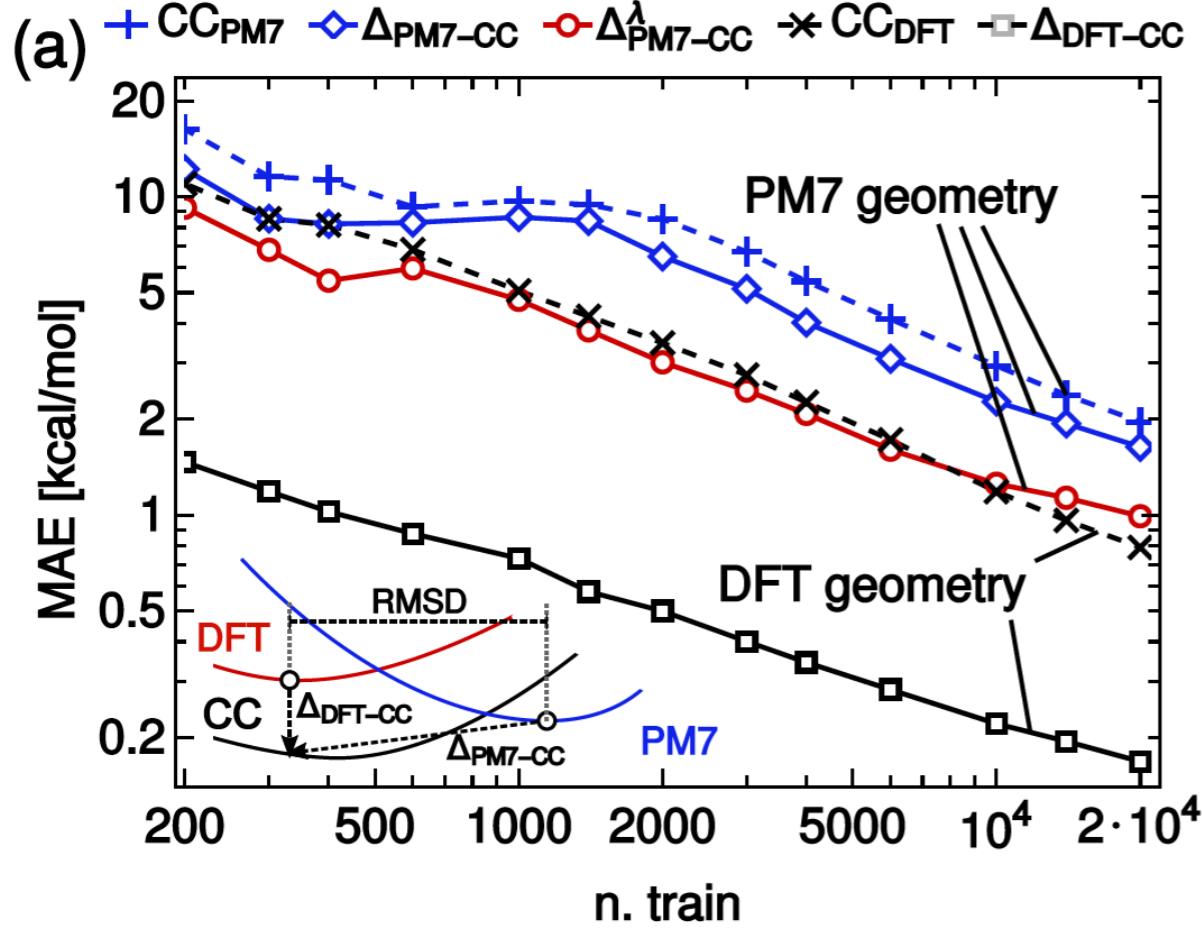
SUBJECT CATEGORIES

- » Quantum chemistry
- » Density functional theory
- » Computational chemistry

Quantum chemistry structures and properties of 134 kilo molecules

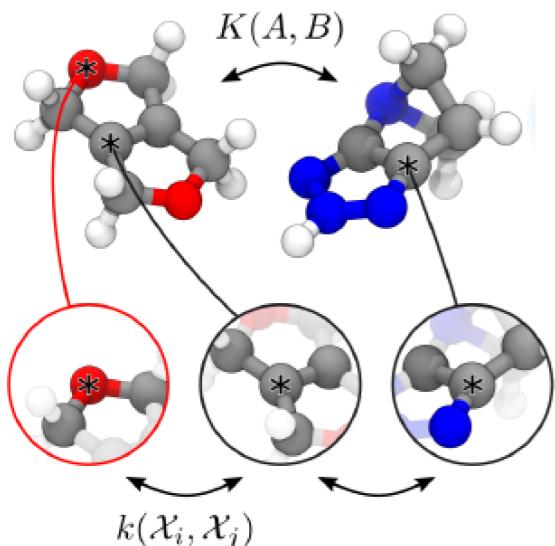
Raghunathan Ramakrishnan¹, Pavlo O. Dral^{2,3}, Matthias Rupp¹ & O. Anatole von Lilienfeld^{1,4}

Computational *de novo* design of new drugs and materials requires rigorous and unbiased exploration of chemical compound space. However, large uncharted territories persist due to its size scaling



From local environment to global structure

(work with Michele Ceriotti and Sandip De)



Construct a kernel between
structures from a "base" kernel
between each pair of atoms

$$C_{ij}(A, B) = k(\mathcal{X}_i^A, \mathcal{X}_j^B)$$

- Average Kernel
- Best-match Kernel
- Regularised match Kernel

$$\bar{K}(A, B) = \frac{1}{N^2} \sum_{ij} C_{ij}(A, B)$$

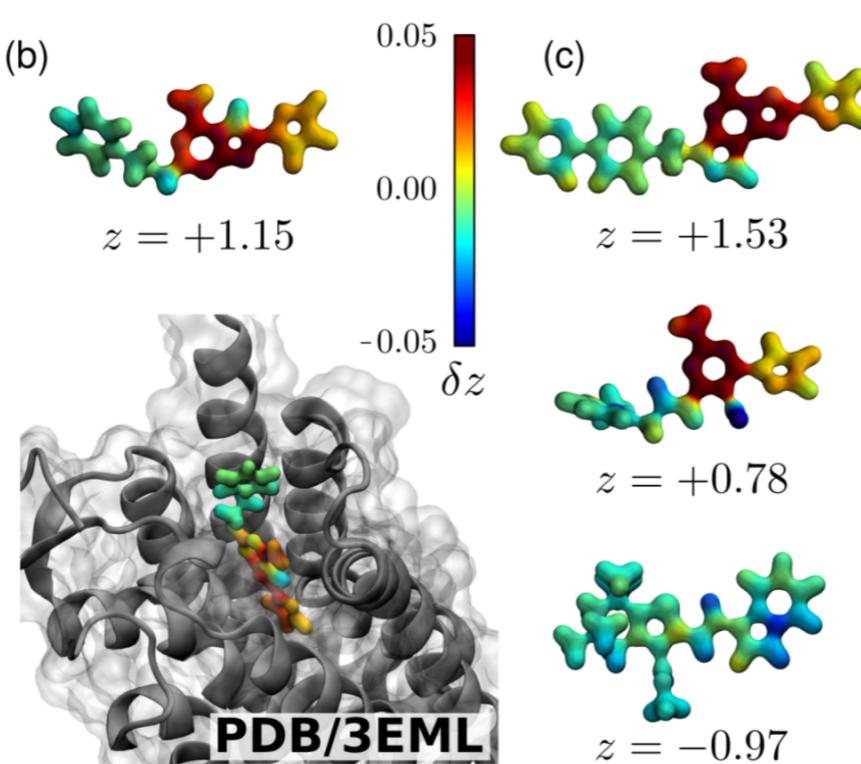
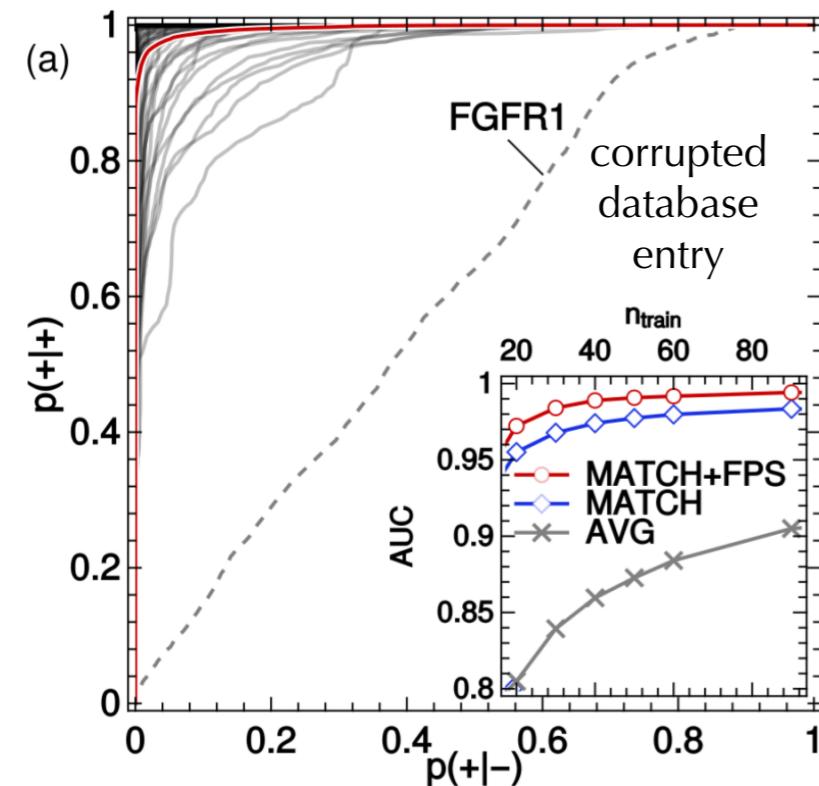
$$\hat{K}(A, B) = \max_{\mathbf{P} \in \mathcal{U}(N, N)} \sum_{ij} C_{ij}(A, B) P_{ij}$$

$$\hat{K}^\gamma(A, B) = \text{Tr} \mathbf{P}^\gamma \mathbf{C}(A, B),$$

$$\mathbf{P}^\gamma = \arg \min_{\mathbf{P} \in \mathcal{U}(N, N)} \sum_{ij} P_{ij} (1 - C_{ij} + \gamma \ln P_{ij})$$

Classification: protein-ligand binding

SCIENCE ADVANCES | RESEARCH ARTICLE



PHYSICS

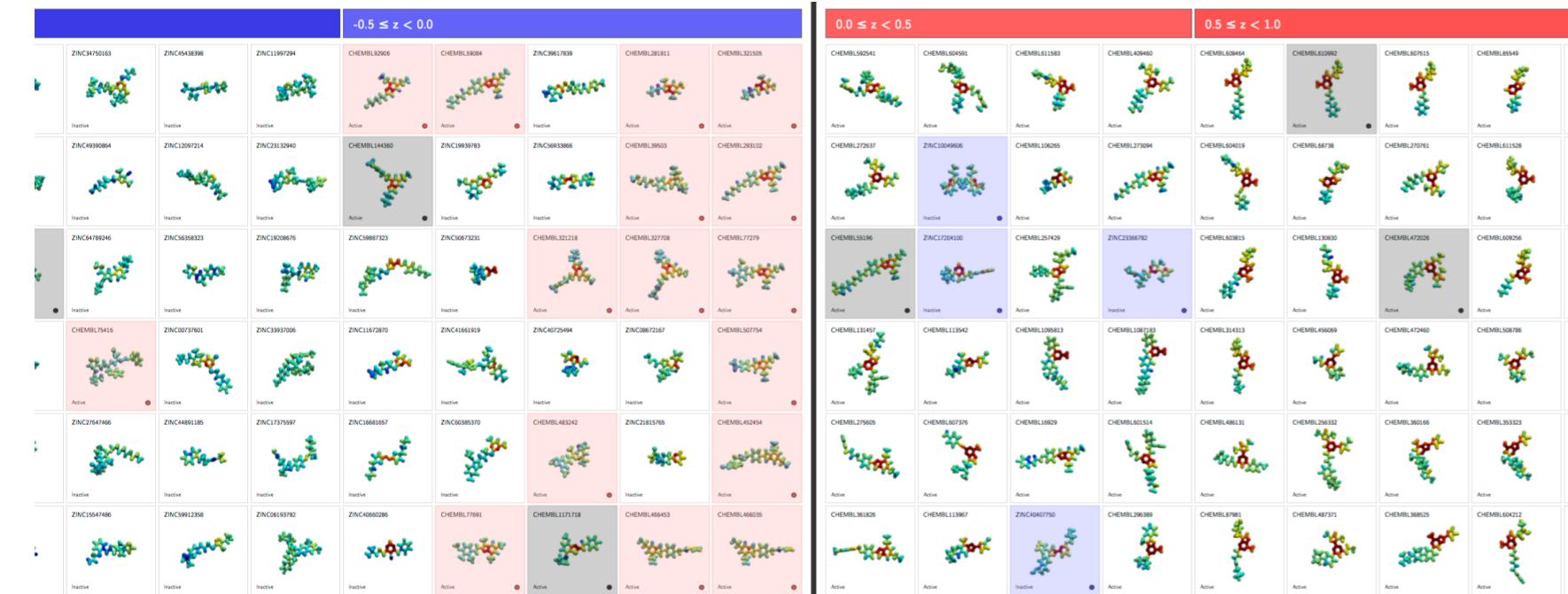
Machine learning unifies the modeling of materials and molecules

Albert P. Bartók,¹ Sandip De,^{2,3} Carl Poelking,⁴ Noam Bernstein,⁵ James R. Kermode,⁶ Gábor Csányi,⁷ Michele Ceriotti^{2,3*}



Apply the general kernel framework for molecular machine learning: classification

- SOAP representation
- entropy-regularised environment match kernel

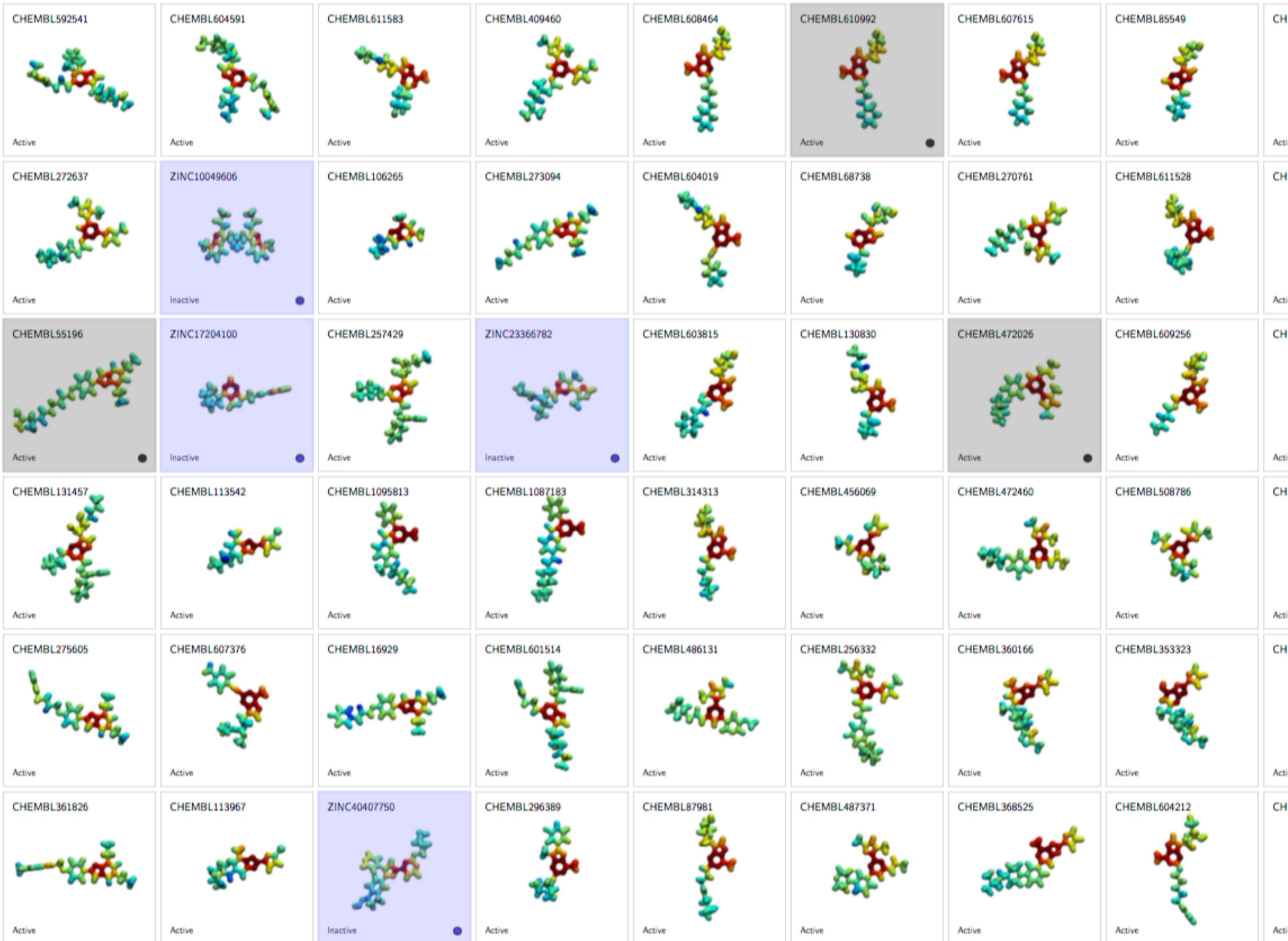


Rank ordering for all ligands and decoys

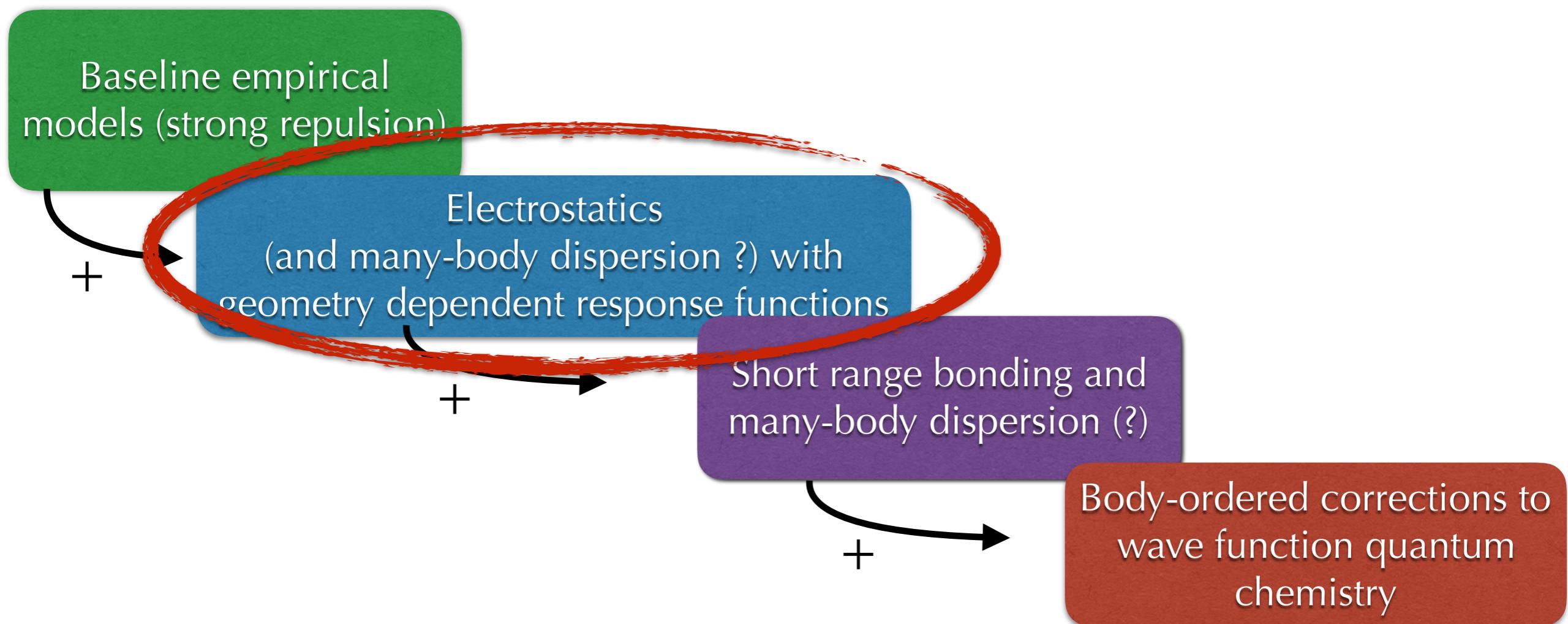
Atomistic “contribution score” identifies warheads

0.0 ≤ z < 0.5

0.5 ≤ z < 1.0



Perspective for meV accurate force fields



- Limitations of DFT: Need data sources using high accuracy wave function based quantum mechanics
- Combine with long range electrostatic interactions (Goedecker 2015)
- Combine exploration of configuration space and fitting (Active learning: Shapeev 2018, E 2018, Deringer 2019)
- Data efficiency: minimal necessary training databases → more widely applicable potentials
- Higher accuracy: SOAP is hard to beat for materials with defects and also good for molecules.
- Include magnetism and electronic entropy
- Eliminate fragility of high dimensional surrogate models (the “extrapolation problem”)