Science & Technology Facilities Council

Learning Interactions from Microscopic Observables

Albert Bartók-Pártay
Science & Technology Facilities Council
Rutherford Appleton Laboratory

in collaboration with the Csányi Group (Cambridge)

Modelling materials

Accuracy

←electrons | no electrons→

geometrical

topological

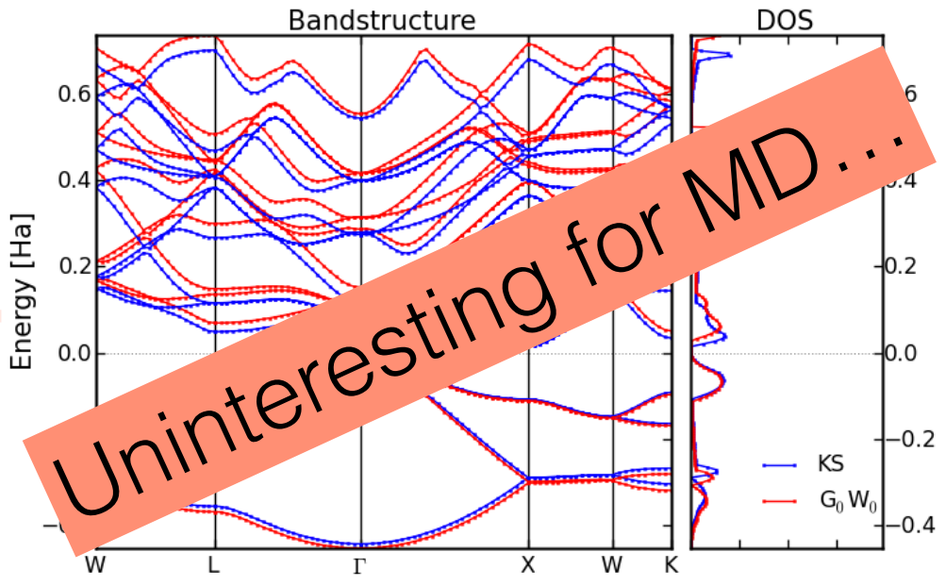
qualitative

1 eV / 40 kT

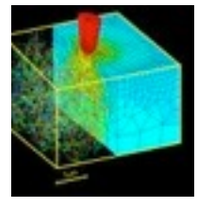
0.1 eV / 4 kT

0.0001 eV / 0.004 kT

0

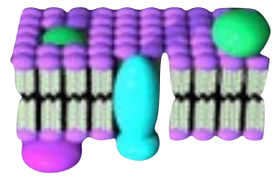


Uninteresting for MD…

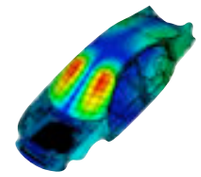


teratomic

potentials



Coarse grained molecular



Finite elements

Quantum Monte Carlo

DFT

QM/LCAO

Quantum Chemistry

Atoms	1	10	100	1000	1,000,000	10^8	∞
Time	0	0	ps	ns	µs	ms	∞

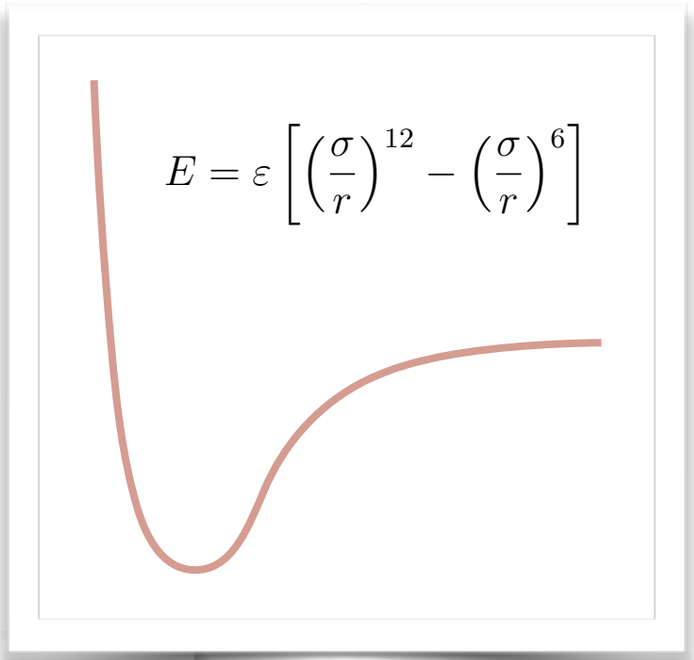
Capability

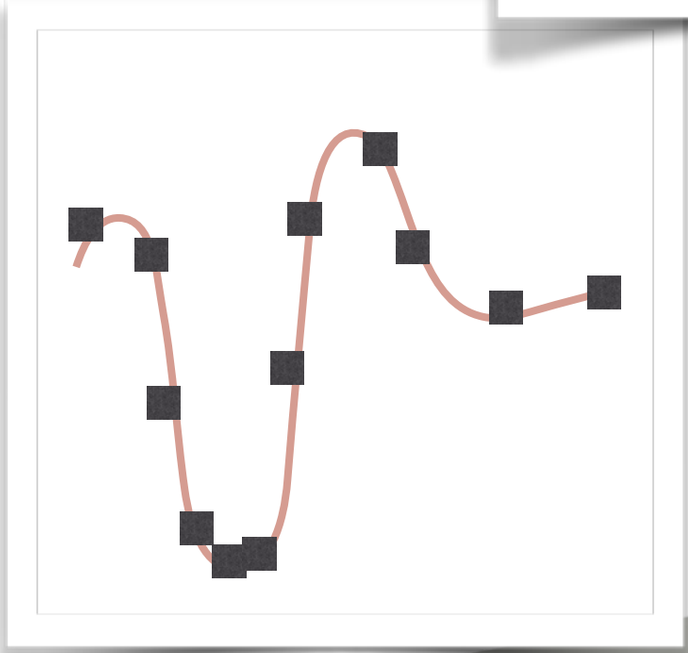
GAP in accuracy and speed

Science & Technology Facilities Council

Interatomic potentials

- Analytic potentials
  - fixed functional formula
  - based on physical understanding
  - few fixed parameters
  - fit to experimental and/or computational data
- Machine learning potentials
  - flexible functional form
  - no physical motivation
  - data driven - mostly computational





Science & Technology Facilities Council

PES fitting

$$E_{\text{total}} = \sum_{i}^{\text{atoms}} \varepsilon(\text{neighbourhood}_i) \qquad E_{\text{total}} = \sum_{i}^{\text{atoms}} \varepsilon(\{\mathbf{r}_{ij}\}_{j}^{\text{neighbours}_i}) + \text{long range}$$

$$\varepsilon(\mathbf{x}) = \sum_{n} \alpha_n k(\mathbf{x}_n, \mathbf{x})$$

atomic environment

- Represent a local neighbourhood configuration: “descriptor”, “fingerprint”, “feature vector”, “symmetry function”
  - Rotational, reflectional, translational and permutational invariance
  - Faithfulness - no two different configuration give the same representation
  - Continuous, differentiable and smooth
- Fits are based on comparisons of configurations
  - Gaussian kernel
  - Dot-product kernel aka linear regression
  - Neural network kernel

$$k(\mathbf{x}, \mathbf{x}') = \exp(-|\mathbf{x} - \mathbf{x}'|^2/2\sigma^2)$$
$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$$
$$k(\mathbf{x}, \mathbf{x}') \approx V - |\mathbf{x} - \mathbf{x}'|^2$$

Science & Technology Facilities Council

Gaussian Process fitting - weight space view

- target function modelled as a linear combination of basis functions $$f_i = f(\mathbf{d}_i, \mathbf{w}) = \sum_h w_h \phi_h(\mathbf{d}_i)$$
- put prior on weights $$P(\mathbf{w}) = \mathrm{Normal}(\mathbf{w}; \mathbf{0}, \sigma_w \mathbf{I})$$
- compute the covariance of two observations $$\langle f_i \, f_j \rangle = \left\langle \sum_{hh'} w_h w_{h'} \phi_h(\mathbf{d}_i) \phi_{h'}(\mathbf{d}_j) \right\rangle = \sigma_w^2 \sum_h \phi_h(\mathbf{d}_i) \phi_h(\mathbf{d}_j)$$
- covariance is defined from the basis functions - but they are not actually needed! $$C(\mathbf{d}, \mathbf{d}') \equiv \sigma_w^2 \sum_h \phi_h(\mathbf{d}) \phi_h(\mathbf{d}')$$

Science & Technology Facilities Council

Gaussian Process fitting - weight space view

- distribution of observed function values is also a Gaussian

$$P(\mathbf{f}) = \mathrm{Normal}(\mathbf{f}; \mathbf{0}, \mathbf{C}) \propto \exp\left(-\frac{1}{2}\mathbf{f}^\top \mathbf{C}^{-1}\mathbf{f}\right)$$

- prediction is the most likely value given the prior and data (see Bayes theorem)

$$P(f|\mathbf{f}) = \frac{P(\mathbf{f}, f)}{P(\mathbf{f})}$$

- result is a closed form approximation

$$f = \mathbf{k}^\top \mathbf{C}^{-1}\mathbf{f}$$

$$k_i = \langle f\ f_j \rangle$$

Science & Technology Facilities Council

Gaussian Process to fit total energies

- model: total energy is a sum of atomic many-body energies

$$\varepsilon_i = \varepsilon(\mathbf{d}_i, \mathbf{w}) = \sum_h w_h \phi_h(\mathbf{d}_i)$$

- result: covariance of total energies is sum up atomic covariance functions

$$E = \sum_i \varepsilon_i + \text{long range}$$

- we have just defined a covariance function for total energies!

$$\langle E_N E_M \rangle = \left\langle \sum_{i\in N} \varepsilon(\mathbf{d}_i) \sum_{j\in M} \varepsilon(\mathbf{d}_j) \right\rangle = \left\langle \sum_{i\in N}\sum_{j\in M}\sum_{hh'} w_h w_{h'} \phi_h(\mathbf{d}_i)\phi_{h'}(\mathbf{d}_j) \right\rangle =$$

$$\sum_{i\in N}\sum_{j\in M}\sum_{hh'} \langle w_h w_{h'}\rangle \phi_h(\mathbf{d}_i)\phi_{h'}(\mathbf{d}_j) = \sigma_w^2 \sum_{i\in N}\sum_{j\in M}\sum_{h} \phi_h(\mathbf{d}_i)\phi_h(\mathbf{d}_j) = \sigma_w^2 \sum_{i\in N}\sum_{j\in M} C(\mathbf{d}_i, \mathbf{d}_j)$$

Science & Technology Facilities Council

Gaussian Process to fit to derivatives

- compute the covariance between energies and forces/ virials

$$\left\langle \frac{\partial E_N}{\partial \xi_k} E_M \right\rangle = \frac{\partial \langle E_N E_M \rangle}{\partial \xi_k} = \sigma_w^2 \sum_{i \in N} \sum_{j \in M} \nabla_{\mathbf{d}_i} C(\mathbf{d}_i, \mathbf{d}_j) \cdot \frac{\partial \mathbf{d}_i}{\partial \xi_k}$$

Cartesian coordinate or cell deformation

- or compute the covariances between forces/virials

- new covariance functions of the same total energy model!

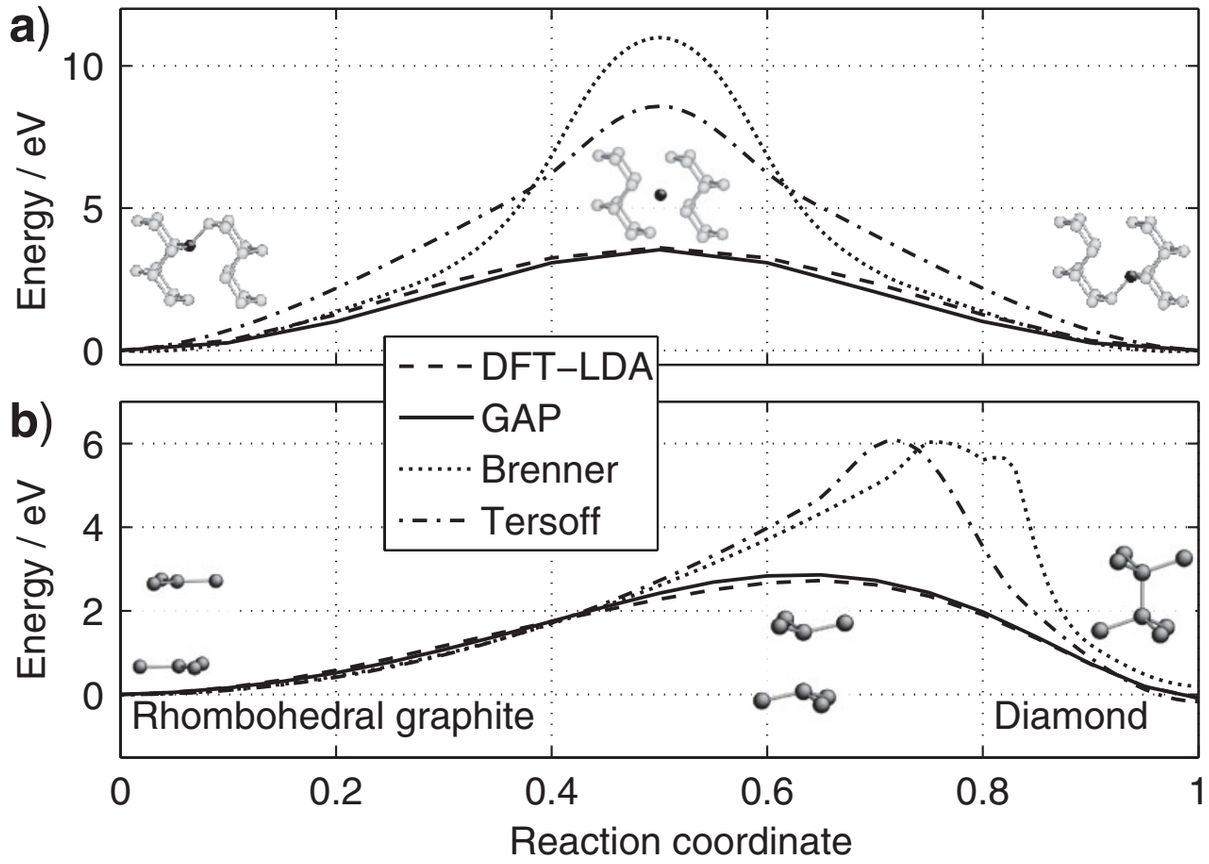
$$\left\langle \frac{\partial E_N}{\partial \xi_k} \frac{\partial E_M}{\partial \chi_l} \right\rangle = \frac{\partial^2 \langle E_N E_M \rangle}{\partial \xi_k \partial \chi_l} = \sigma_w^2 \sum_{i \in N} \sum_{j \in M} \frac{\partial \mathbf{d}_i^\top}{\partial \xi_k} (\nabla_{\mathbf{d}_i} C(\mathbf{d}_i, \mathbf{d}_j) \nabla_{\mathbf{d}_j}^\top) \frac{\partial \mathbf{d}_j}{\partial \chi_l}$$

E Solak *et al*, NIPS **15**, 529 (2003)

Science & Technology Facilities Council

GP to fit energies *and* forces

- Database: graphite and diamond forces
- Phonons great, but energetics *not*



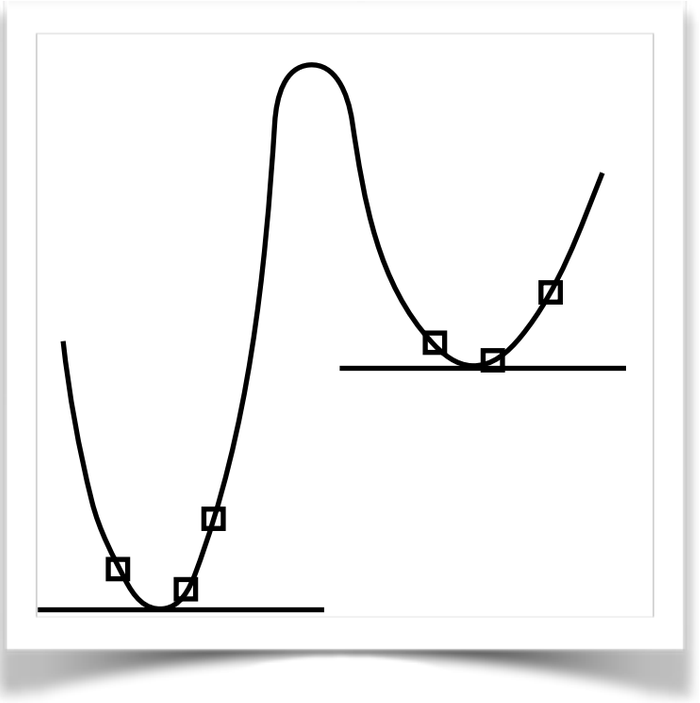
Science & Technology Facilities Council

GP to fit energies *and* forces

- forces provide rich information on PES

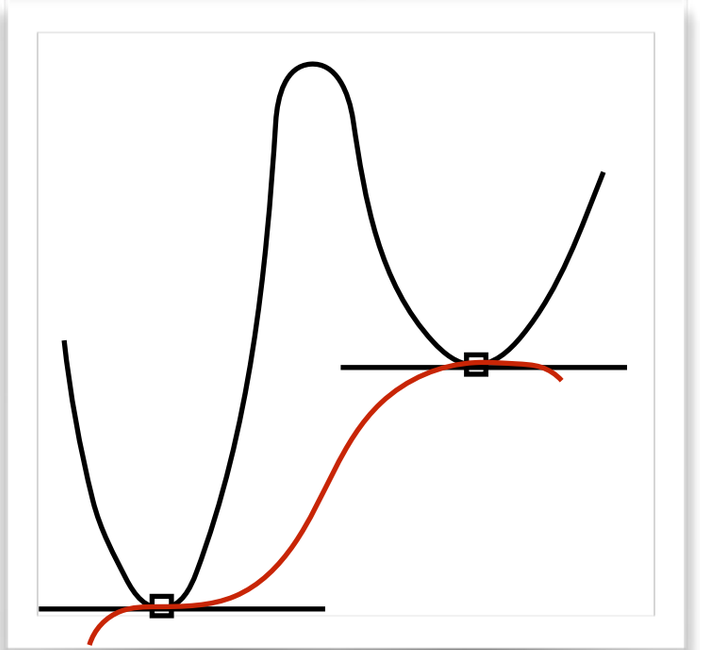
- energies set scale *and* connect minima correctly

- viral stresses capture the really soft response to deformation



Science & Technology Facilities Council

GP to fit to second derivatives?



- if learning from minima - forces not very helpful
- second derivatives provide even richer information than first derivatives
- covariances must be a nightmare to derive!

Science & Technology Facilities Council

GP to fit the Hessian

$$E = E_0 + \sum_{i\alpha} \frac{\partial E}{\partial r_{i\alpha}} u_{i\alpha} + \frac{1}{2} \sum_{i\alpha j\beta} \frac{\partial^2 E}{\partial r_{i\alpha} \partial r_{j\beta}} u_{i\alpha} u_{j\beta}$$

- harmonic approximation to the total energy

$$E = E_0 - \mathbf{f}_0 \cdot \mathbf{u} + \frac{1}{2} \mathbf{u}^T \mathbf{H}_0 \mathbf{u}$$

- diagonalise the Hessian

$$\mathbf{H}_0 = \mathbf{E} \Lambda \mathbf{E}^T = \sum_k \lambda_k \mathbf{e}_k \mathbf{e}_k^T$$

- obtain forces in the harmonic approximation when displacements are along an eigenvector

$$\mathbf{f}_i^{\pm} = \mathbf{f}_0 \mp \Delta \mathbf{H}_0 \mathbf{e}_i = \mathbf{f}_0 \mp \Delta \lambda_i \mathbf{e}_i$$

Science & Technology Facilities Council

GP to fit the Hessian

- equilibrium forces cancel if done in both directions

$$\mathbf{f}_i^{\pm} = \mathbf{f}_0 \mp \Delta \mathbf{H}_0 \mathbf{e}_i = \mathbf{f}_0 \mp \Delta \lambda_i \mathbf{e}_i$$

- formula for the Hessian eigenvalues

$$\lambda_i = \frac{\mathbf{e}_i^T (\mathbf{f}_i^- - \mathbf{f}_i^+)}{2\Delta}$$

- covariance of Hessian eigenvalue with local energy terms - simple, and no need for new formula!

$$\langle \lambda_i \varepsilon \rangle = \frac{\mathbf{e}_i^T}{2\Delta} (\langle \mathbf{f}_i^- \varepsilon \rangle - \langle \mathbf{f}_i^+ \varepsilon \rangle)$$

- essentially an optimal (in some way) numerical differentiation

Science & Technology Facilities Council

Putting it together

- how to use all these kernel functions?
- put it all in a large covariance matrix - everything!
- this provides a unified model fitting all the data
- a bit too large? Use sparsification!

Quiñonero-Candela and Rasmussen JMLR **6**, 1939 (2005)

Science & Technology Facilities Council

Sparsification

- Training involves matrix “inversion”, *O(N*^{3}*)*
- Prediction is *O(N)*
- but atomic configurations are *very* correlated

$$f = \mathbf{k}^\top \mathbf{C}^{-1} \mathbf{f}$$

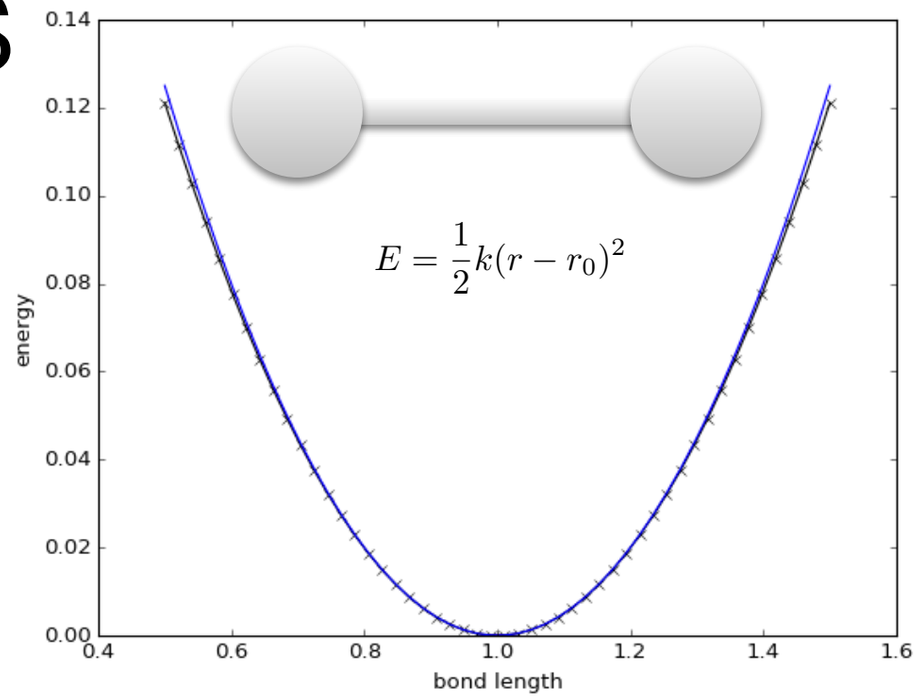
- possible to approximate **C** by a low-rank expression
- choose *M* representative data points
- Training complexity: *O(NM*^{2}*)*
- Prediction: *O(M)*

$$\mathbf{C}_{NN} = \mathbf{C}_{NM} \mathbf{C}_{MM}^{-1} \mathbf{C}_{MN}$$

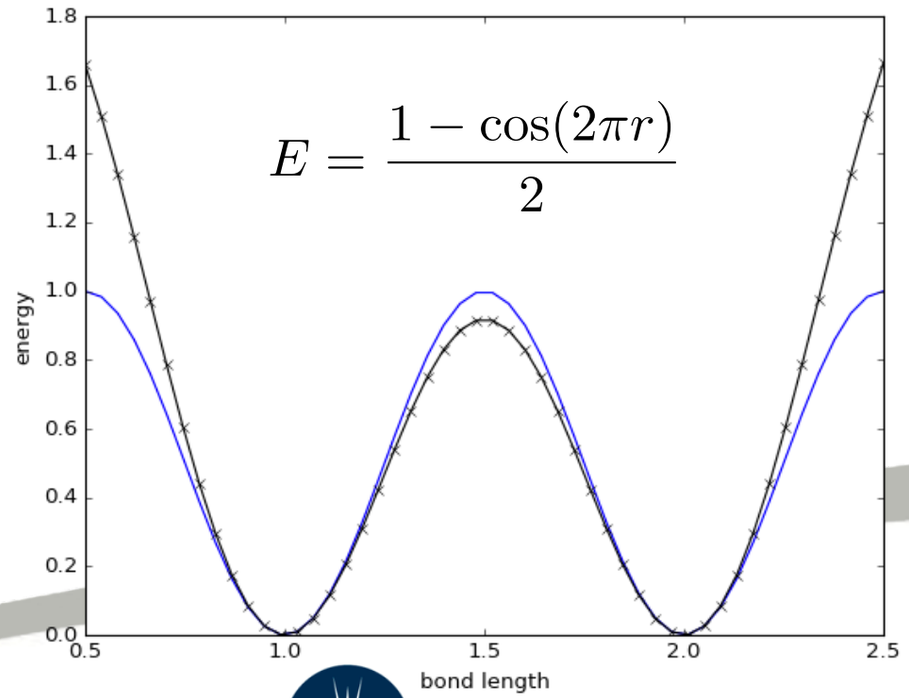
Quiñonero-Candela and Rasmussen JMLR **6**, 1939 (2005)
Snelson and Ghahramani NIPS 18

GP to fit the Hessian: Results

- toy model:
  - harmonic spring between two atoms
  - one teaching point at the minimum



- toy model II:
  - sinusoidal spring with two minima
  - two teaching points

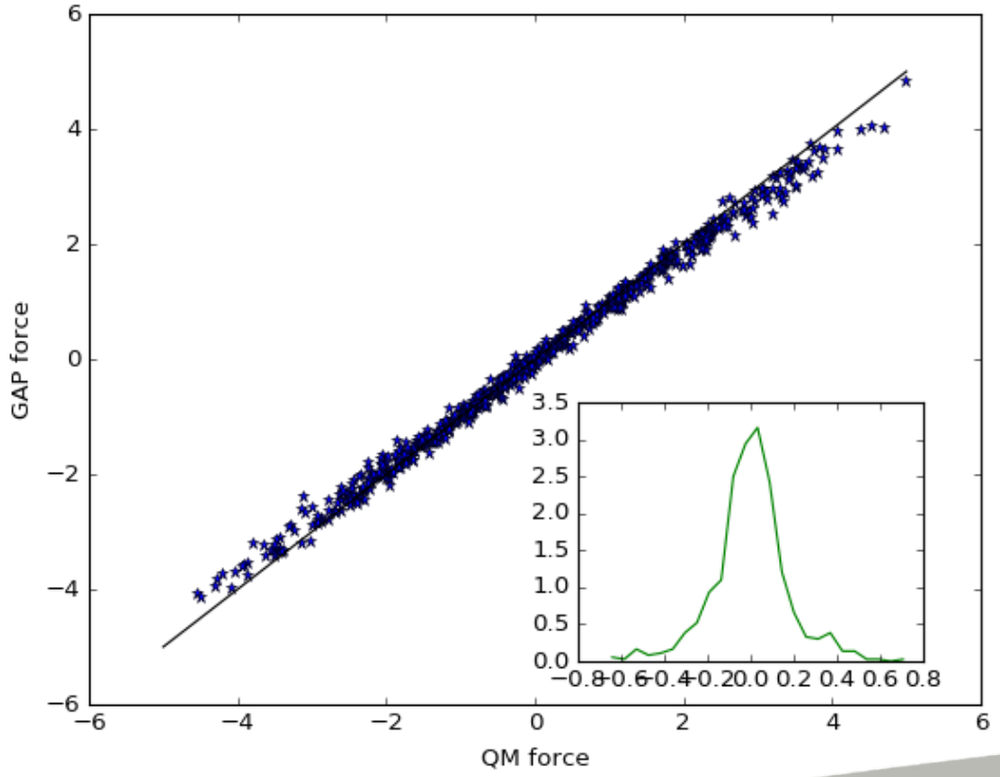


Science & Technology Facilities Council

GP to fit the Hessian: Results

- 8-atom cubic silicon cell with 21 Hessian eigenvalues

- test forces on 216-atom cell, randomised positions



Science & Technology Facilities Council

GP fit to charges

- fitting charges

$$q(\mathbf{d}) = \sum_h w_h \phi_h(\mathbf{d})$$

- training data: electrostatic potential around the molecule

$$\Phi(\mathbf{r}) = \sum_i \frac{q(\mathbf{d}_i)}{|\mathbf{r} - \mathbf{r}_i|}$$

- environment-dependent ESP charge fitting

$$\langle \Phi(\mathbf{r}) \Phi(\mathbf{r}') \rangle = \sum_{ij} \frac{C(\mathbf{d}_i, \mathbf{d}_j)}{|\mathbf{r} - \mathbf{r}_i| \cdot |\mathbf{r}' - \mathbf{r}_j|}$$

- essentially a regularisation!

Tannor *et al*, JACS **116**, 11875 (1994)

Science & Technology Facilities Council

GP fit to charges

	methanol	ethanol	acetone
O	-0.69	-0.70	-0.58
H_O	0.43	0.40 (0.41)	
C_O	0.29	0.49 (0.46)	0.80
C_β		-0.34 (-0.29)	-0.51

Science & Technology Facilities Council

What is the kernel?

- pair-based kernel: distance
- three-body kernel: symmetrised distances
- many-body kernel: SOAP

Fitting multiple models

- Example: total energy is sum of pair and triplets

$$E = \sum_{p \in \text{ pairs}} \varepsilon_p^{(2)} + \sum_{t \in \text{ triplets}} \varepsilon_t^{(3)}$$

- each term is a GP itself

$$\varepsilon^{(2)}(\cdot\cdot) = \sum_h w_h^{(2)} \phi_h^{(2)}(\cdot\cdot)$$

$$\varepsilon^{(3)}(\cdot\cdot) = \sum_h w_h^{(3)} \phi_h^{(3)}(\cdot\cdot)$$

- covariance is a sum pair and triplet covariances

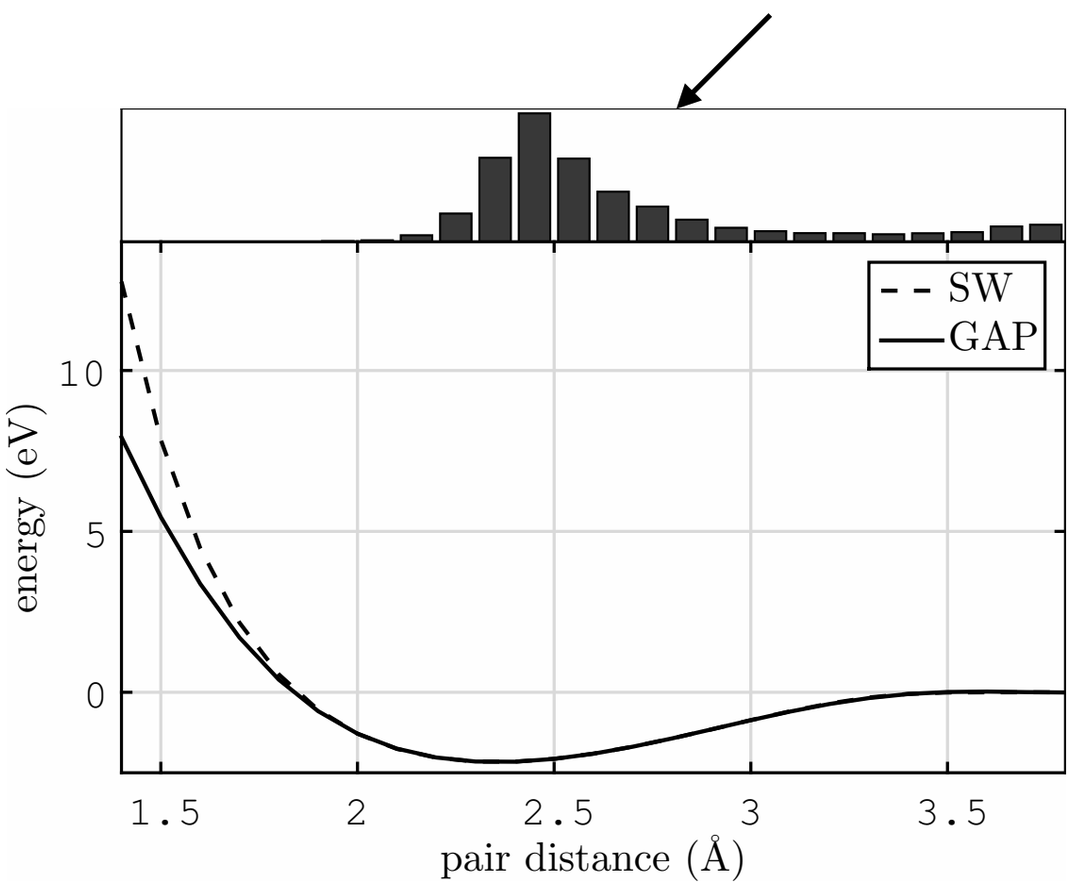
$$\langle E_N E_M \rangle = \sigma^2_{w^{(2)}} \sum_{p \in \text{pairs}_N} \sum_{q \in \text{pairs}_M} C^{(2)}(p,q) + \sigma^2_{w^{(3)}} \sum_{t \in \text{triplets}_N} \sum_{u \in \text{triplets}_M} C^{(3)}(t,u)$$

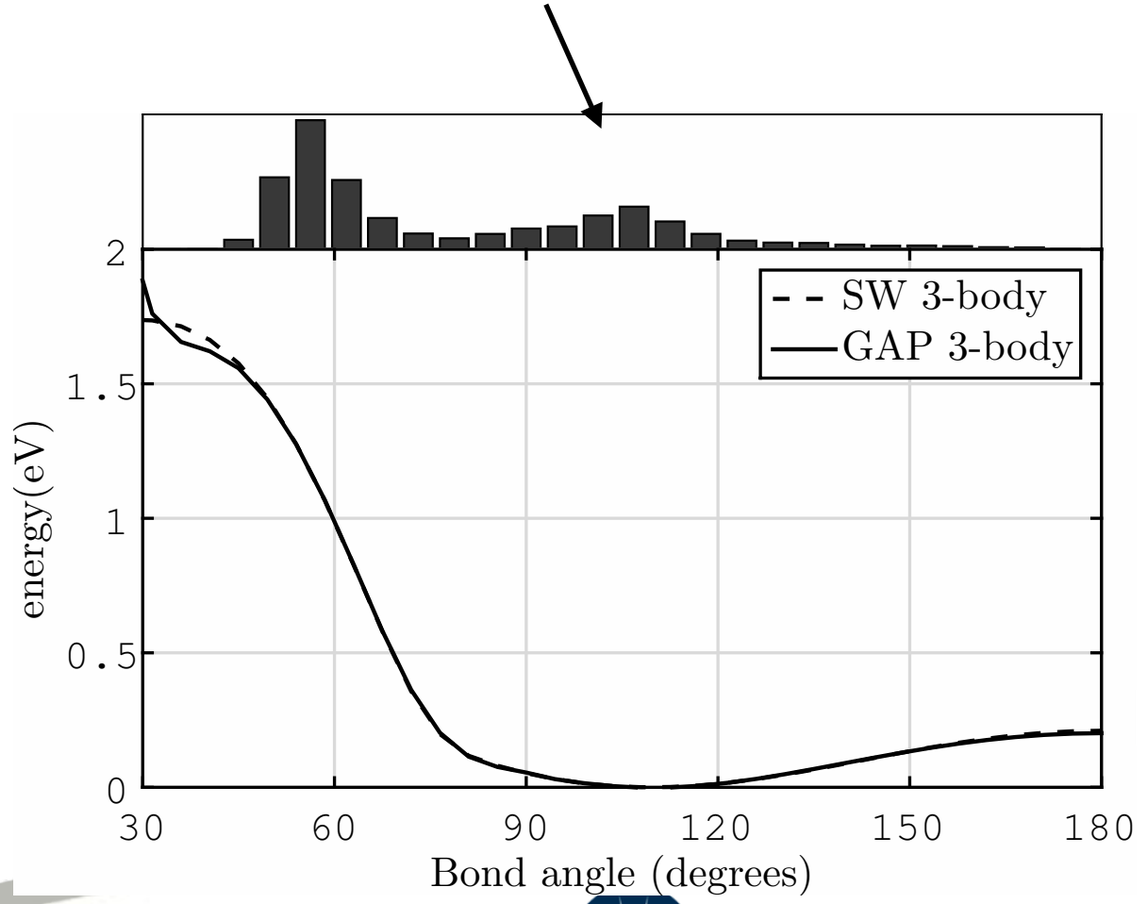
Science & Technology Facilities Council

GAP fitting of multiple models

Fit of a two and three-body model on Si SW-data

$$\langle E_N E_M \rangle = \sigma^2_{w^{(2)}} \sum_{p \in \text{pairs}_N} \sum_{q \in \text{pairs}_M} C^{(2)}(p,q) + \sigma^2_{w^{(3)}} \sum_{t \in \text{triplets}_N} \sum_{u \in \text{triplets}_M} C^{(3)}(t,u)$$

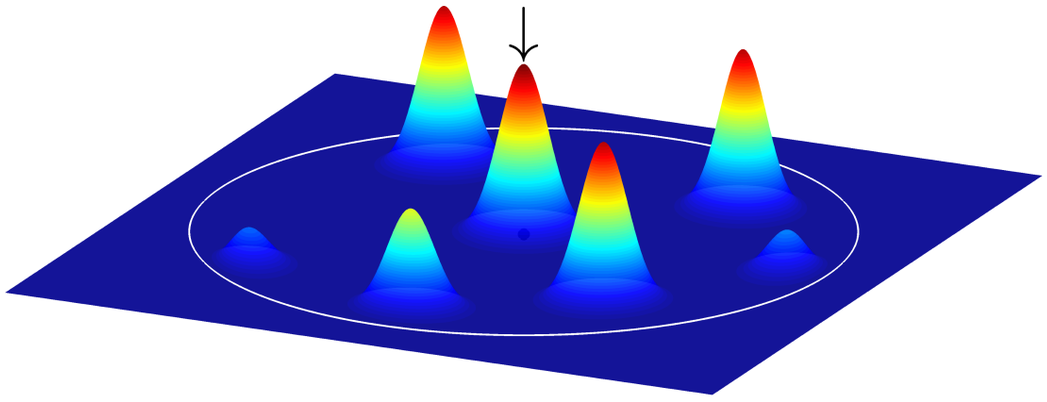




Science & Technology Facilities Council

Many-body kernel

- represent atomic neighbourhood environment by atomic density



$$\rho_i(\mathbf{r}) \equiv \sum_j^{\text{neigh.}} \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_{ij}|^2}{2\sigma_{\text{atom}}^2}\right)$$

- atomic energy is a linear functional of density

$$\varepsilon_i = \varepsilon[\rho_i] = \int w(\mathbf{r})\rho_i(\mathbf{r})\mathrm{d}\mathbf{r}$$

- distribution of weights is Gaussian

$$\langle w(\mathbf{r})w(\mathbf{r}')\rangle = \delta(\mathbf{r}-\mathbf{r}')\sigma_w^2$$

- kernel becomes the overlap

$$C(\rho_i,\rho_j) = \langle\varepsilon_i\varepsilon_j\rangle = \left\langle \int w(\mathbf{r})\rho_i(\mathbf{r})w(\mathbf{r}')\rho_j(\mathbf{r}')\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \right\rangle = \sigma_w^2 \int \rho_i(\mathbf{r})\rho_j(\mathbf{r})\,\mathrm{d}\mathbf{r}$$

Science & Technology Facilities Council

Many-body kernel

- kernel needs to be rotationally invariant

$$C'(\rho_i, \rho_j) = \int |C(\rho_i, \hat{R}\rho_j)|^p \mathrm{d}\hat{R}$$

- integrate over all possible rotations

$$C''(\rho_i, \rho_j) = \frac{C'(\rho_i, \rho_j)}{\sqrt{C'(\rho_i, \rho_i)}\sqrt{C'(\rho_j, \rho_j)}}$$

- normalise

$$\rho_i(\mathbf{r}) = \sum_{nlm} c^{(i)}_{nlm} g_n(r) Y_{lm}(\hat{\mathbf{r}})$$

- it can be done analytically

$$p^{(i)}_{nn'l} \equiv \frac{1}{\sqrt{2l+1}} \sum_m c^{(i)}_{nlm} (c^{(i)}_{n'lm})^*$$

$$C'(\rho_i, \rho_j) = \sum_{n,n',l} p^{(i)}_{nn'l} p^{(j)}_{nn'l}$$

Science & Technology Facilities Council

Is SOAP a good similarity kernel?

- can we find two genuinely different configurations which are yet deemed equivalent? $\exists i, j \text{ where } \rho_i \not\equiv \rho_j \text{ such that } C''(\rho_i, \rho_j) = 1?$

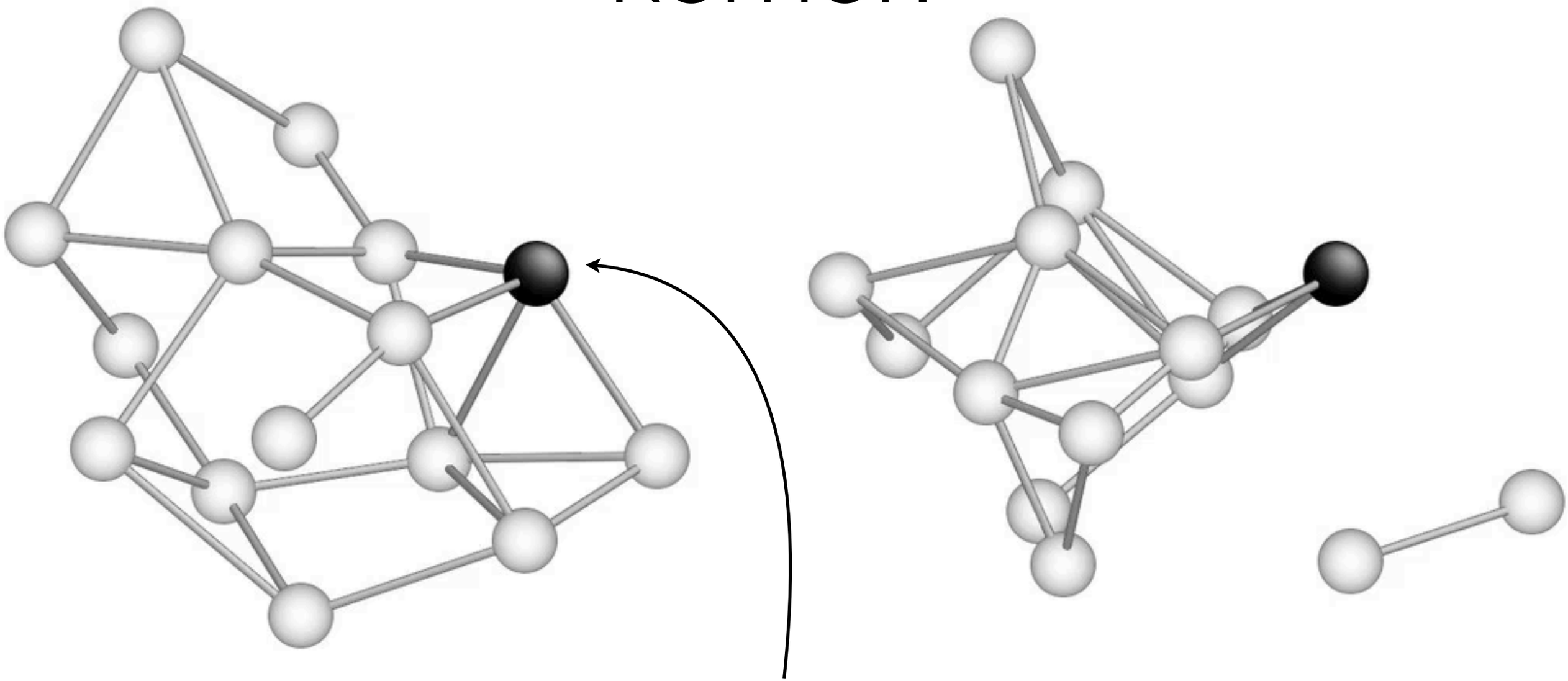
- no analytic proof

- numerical experiments: match an atomic density to a reference density

$$\underset{\rho}{\operatorname{argmax}}\, C''(\rho, \rho')$$

Science & Technology Facilities Council

Is SOAP a good similarity kernel?

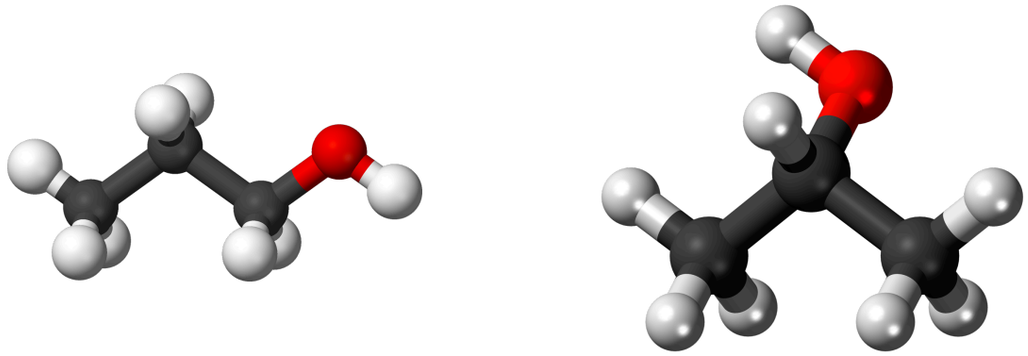


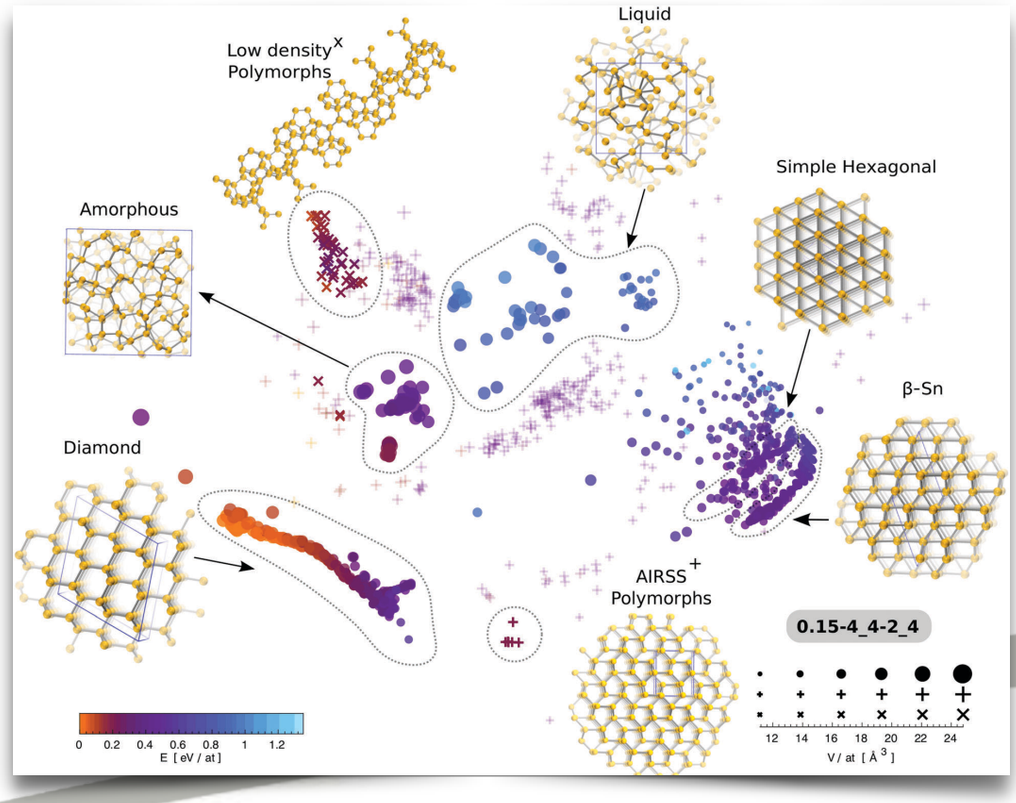
Science & Technology Facilities Council

SOAP for structure comparison

work with Sandip De and Michele Ceriotti

- SOAP provides a similarity measure between atoms
- use it to compare two molecules or structures
- build a matrix from all pairwise SOAP kernels
  - average the elements
  - find the best match





Science & Technology Facilities Council

Acknowledgments

- Dario Alfè
- Noam Bernstein
- Michele Ceriotti
- Gábor Csányi
- Sandip De
- Mike Gillan
- Ferenc Huszár
- James Kermode
- Fred Manby
- Alan Nichol
- Jonatan Öström
- Mike Payne
- Peter Pinski
- Thor Wikfeldt

- University of Cambridge
- Isaac Newton Trust
- Leverhulme Trust
- STFC
- CCP-NC

The Leverhulme Trust

UNIVERSITY OF CAMBRIDGE
Department of Engineering

Science & Technology Facilities Council