Bridging the gap:

Gaussian Approximation Potential

Albert Bartók-Pártay

Gábor Csányi

Risi Kondor

Outline

- from QM to interatomic potentials
- potential based directly on detailed QM data
- high dimensional fit (Gaussian Processes)
- atomic neighbourhoods: bispectrum
- the first GAP for carbon
- other uses: defining the local energy

From QM...

- quantum mechanics is the 'ultimate truth'
- expensive to solve
- sequence of approximations:
  - Full CI
  - QMC
  - DFT-LDA
  - tight binding

  - interatomic potentials

...to interatomic potentials

- energy is sum of atomic energies
- atomic energy depends on neighbouring atoms
- electronic problem is not solved
  - cluster expansion of total energy

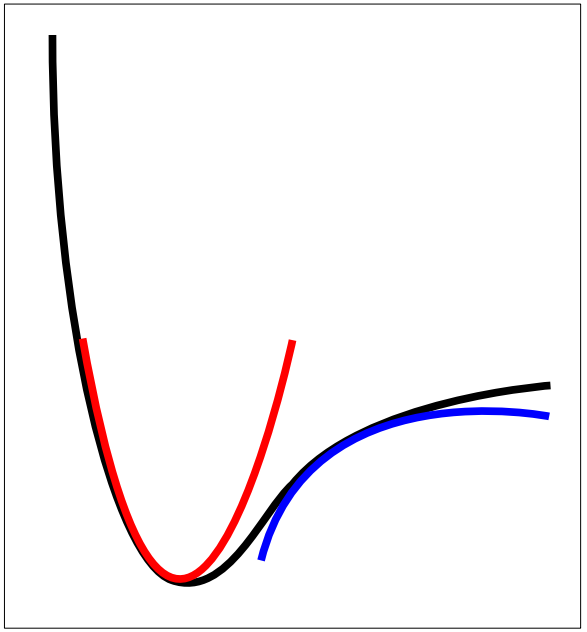
$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = V^{(0)} + \sum_i V^{(1)}(\mathbf{r}_i) + \sum_{i<j} V^{(2)}(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i<j<k} V^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \ldots$$

  - EAM expansion

$$V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \sum_i E(\rho_i) \qquad \rho_i = \sum_n^{\text{neighbours}} \rho(r_{in})$$

Generating potentials

- How is an interatomic potential generated?
  - empirical, analytic formula
  - choose target properties (even forces)
  - fit free parameters to reproduce properties
  - hope that the formula remains reasonably valid everywhere in the configurational space
- The GAP way:
  - no fixed formula
  - search in the space of smooth functions
  - identify target configurations
  - fit to arbitrary precision QM data
  - extend target set if needed



GAP

- energy is sum of atomic energies
- $\epsilon(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \ldots)$ is not practical
- matrix $\mathbf{r}_i \cdot \mathbf{r}_j$ is complete but not invariant to permutations
- symmetric polynomials are also complete but not invariant to rotation
  - CH_5^+ : all terms, chosen the rotationally invariant ones
- our solution
  - atomic energy is a functional of atomic density
  - express atomic density in rotationally invariant terms

GAP: function fitting

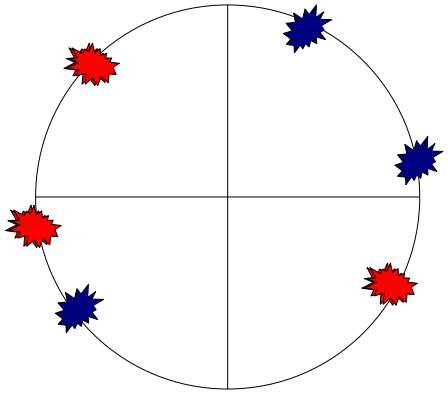
- how to choose the target configurations?
- an optimal way to interpolate many-dimensional functions

$$f(x) = \sum_i w_i \exp\left(-\frac{(x - x_i)^2}{2\sigma^2}\right)$$

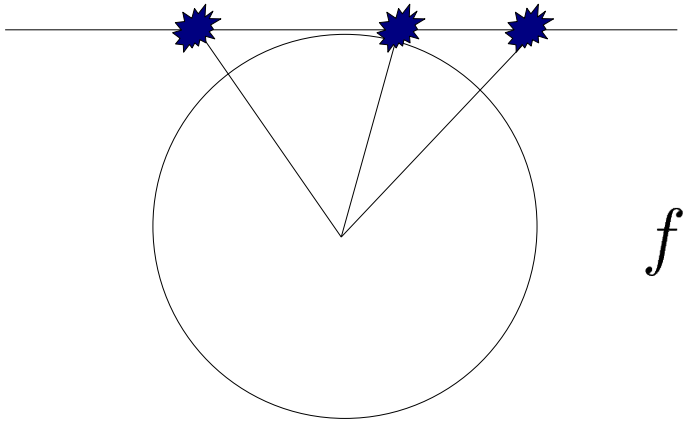
- magic trick: finding which fitting points are optimal for reproducing a very large data set
- calculate accurate forces and energies (DFT)
- perform fit using sum of derivatives of ϵ_i (forces) and sum of ϵ_i (energy) as the target function

Invariant representations

- Two examples:
  - 1D periodic function
  - 2D object



- 1D: $f(\phi) \rightarrow \hat{f}_n$
- 2D: project on the Riemann-sphere then express it in spherical harmonics basis:

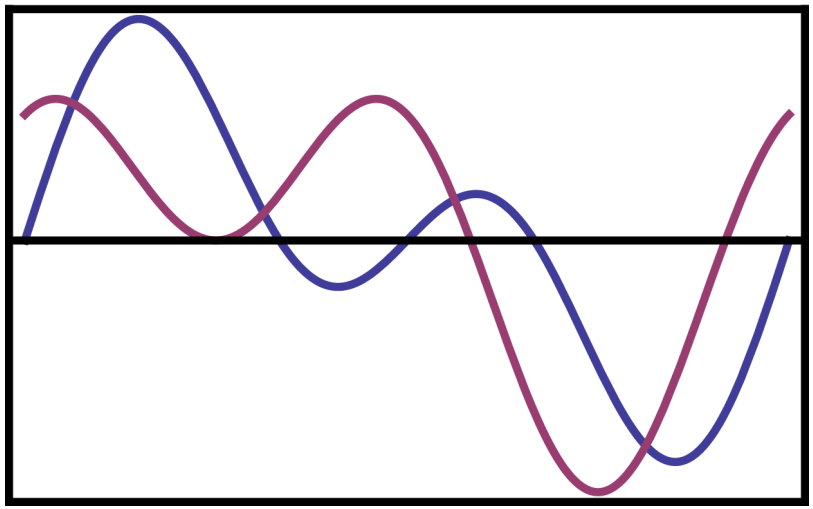


$$f(x,y) = f(\theta,\phi) \rightarrow f_{lm}$$

Invariant representations

- Power spectrum:
  - 1D: $g(\phi) = f(\phi + \phi_0)$ $\qquad \hat{g}_n = \exp(in\phi_0)\hat{f}_n$

    $p_n = \hat{f}_n^* \hat{f}_n \rightarrow \hat{f}_n^* \exp(-in\phi_0)\hat{f}_n \exp(in\phi_0) = \hat{f}_n^* \hat{f}_n$
  - 2D: $\mathbf{f}_l \rightarrow \mathbf{D}_l \mathbf{f}_l$ $\qquad \mathbf{f}_l^\dagger \mathbf{f}_l \rightarrow \mathbf{f}_l^\dagger \mathbf{D}_l^\dagger \mathbf{D}_l \mathbf{f}_l = \mathbf{f}_l^\dagger \mathbf{f}_l$
  - incomplete representation, phase information lost



Invariant representations

- Bispectrum (almost complete):
  - 1D: $b_n = \hat{f}^*_{n+m}\hat{f}_n\hat{f}_m \rightarrow$
    $\rightarrow \hat{f}^*_{n+m}\exp(-i(n+m)\phi_0)\hat{f}_n\exp(in\phi_0)\hat{f}_m\exp(im\phi_0) = \hat{f}^*_{n+m}\hat{f}_n\hat{f}_m$
  - 2D: $\mathbf{f}_{l_1}\otimes\mathbf{f}_{l_2} \rightarrow (\mathbf{D}_{l_1}\otimes\mathbf{D}_{l_2})\,\mathbf{f}_{l_1}\otimes\mathbf{f}_{l_2}$

$$\mathbf{D}_{l_1}\otimes\mathbf{D}_{l_2} = \mathbf{C}^{\dagger}_{l_1,l_2}\left[\bigoplus_{l=|l_1-l_2|}^{l_1+l_2}\mathbf{D}_l\right]\mathbf{C}_{l_1,l_2}$$

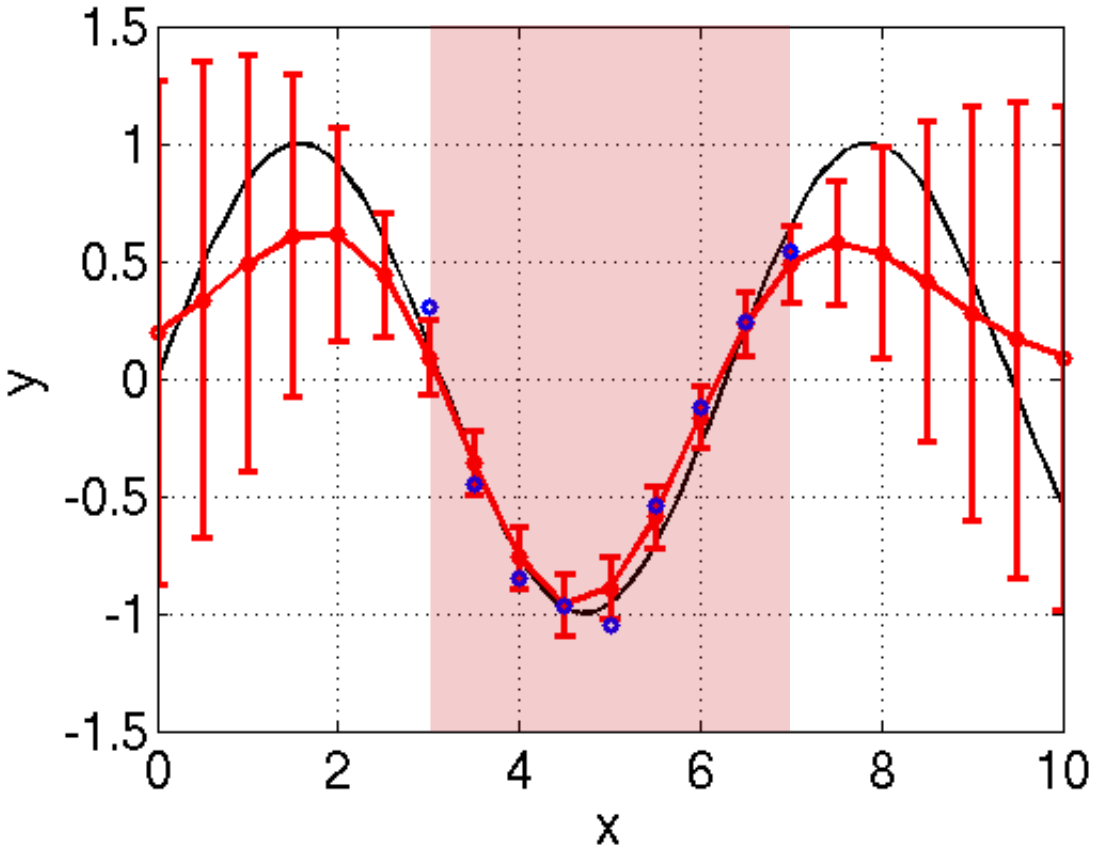
$$\mathbf{C}_{l_1,l_2}\mathbf{f}_{l_1}\otimes\mathbf{f}_{l_2} \rightarrow \left[\bigoplus_{l=|l_1-l_2|}^{l_1+l_2}\mathbf{D}_l\right]\mathbf{C}_{l_1,l_2}\mathbf{f}_{l_1}\otimes\mathbf{f}_{l_2} = \bigoplus_{l=|l_1-l_2|}^{l_1+l_2}\mathbf{g}_{l_1,l_2,l}$$

$$p_{l_1,l_2,l} = \mathbf{f}_l^{\dagger}\mathbf{g}_{l_1,l_2,l}$$

Invariant representations: 3D objects

- project on a 4D sphere
- $f(x, y, z) \rightarrow f(\alpha, \theta, \phi)$
- θ and ϕ are the same as the 3D polar coordinates
- express in 4D spherical harmonics basis: the Wigner D-matrices
- bispectrum is analogous to 3D case
- 4D CG-coefficients are direct products of 3D CG-coefficients

GP: a simple interpolation



target function: sin *x*
fit points between 3 and 7
random noise on fit points
expectation value and variance predicted

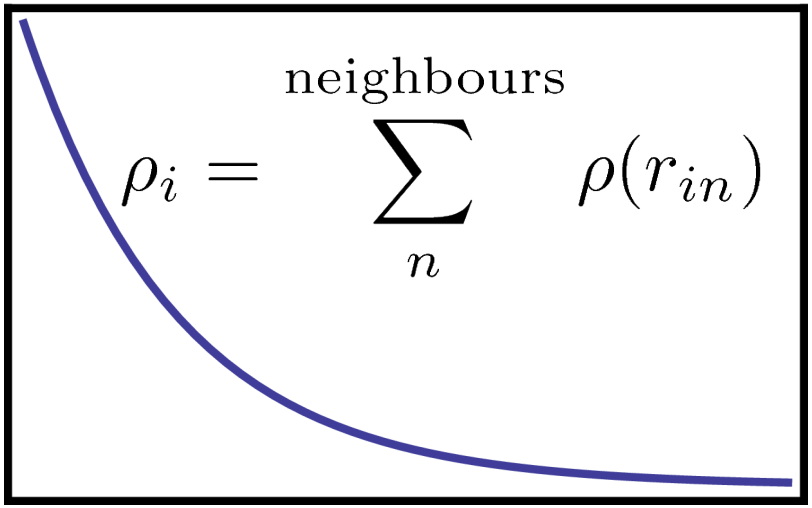
Target configurations

- Hot MD of interesting systems
  - surfaces
  - interstitial
  - vacancy
  - quenched liquid
- energies and forces of configuration samples with DFT

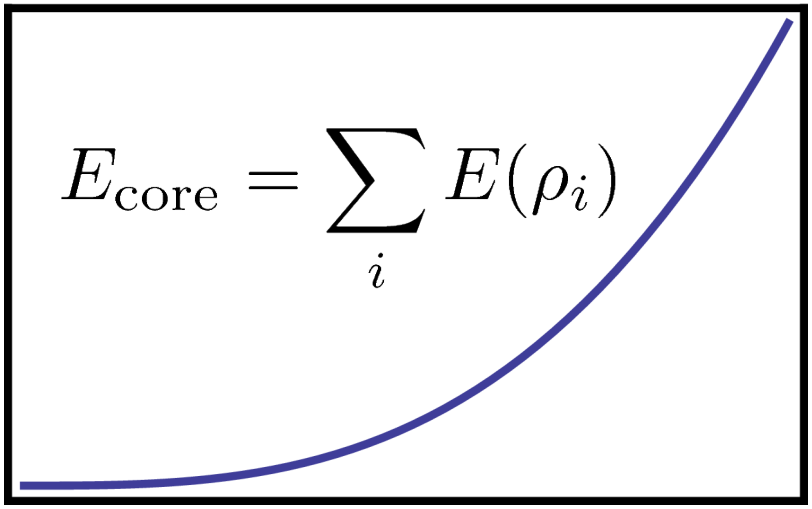
'Baseline' potential

- GP gives 0 as an answer when unsure
- add core repulsion to correct for very close atoms
  - EAM formula
  - fitted to high pressure DFT results
- add dispersion term for long-range interactions

$$\rho_i = \sum_{n}^{\text{neighbours}} \rho(r_{in})$$



$$E_{\text{core}} = \sum_i E(\rho_i)$$

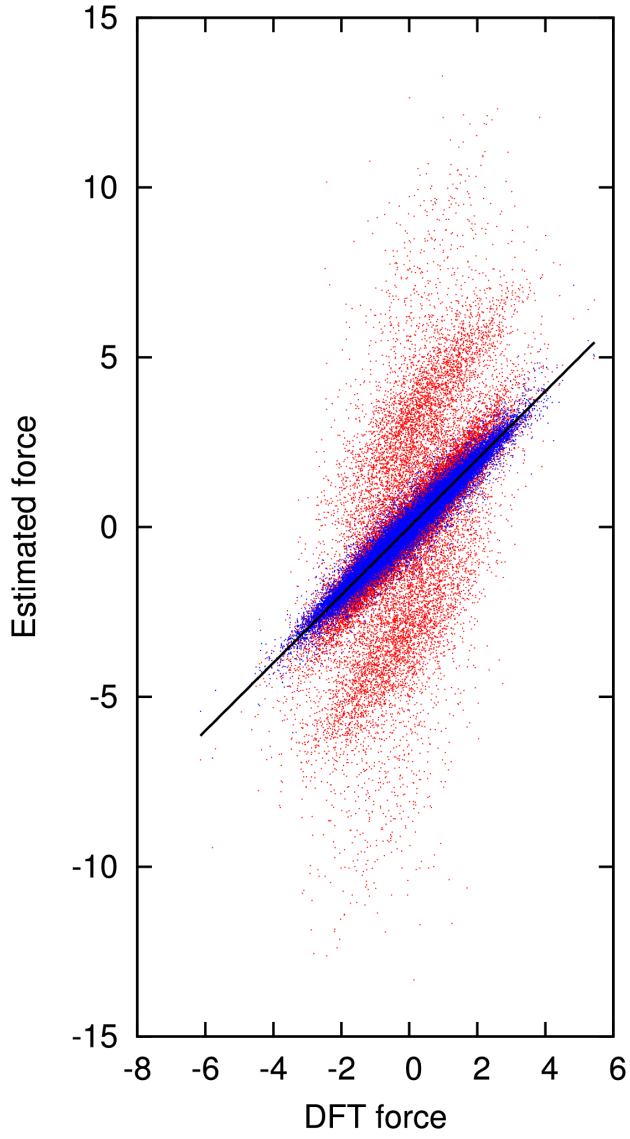


GAP for carbon

- target configurations
  - bulk phases
  - transition from diamond to graphite
  - 111 and 100 surfaces
  - vacancy
  - interstitial
  - amorphous carbon

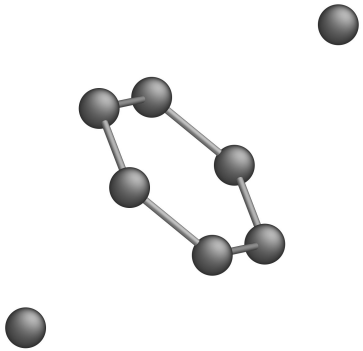
GAP test

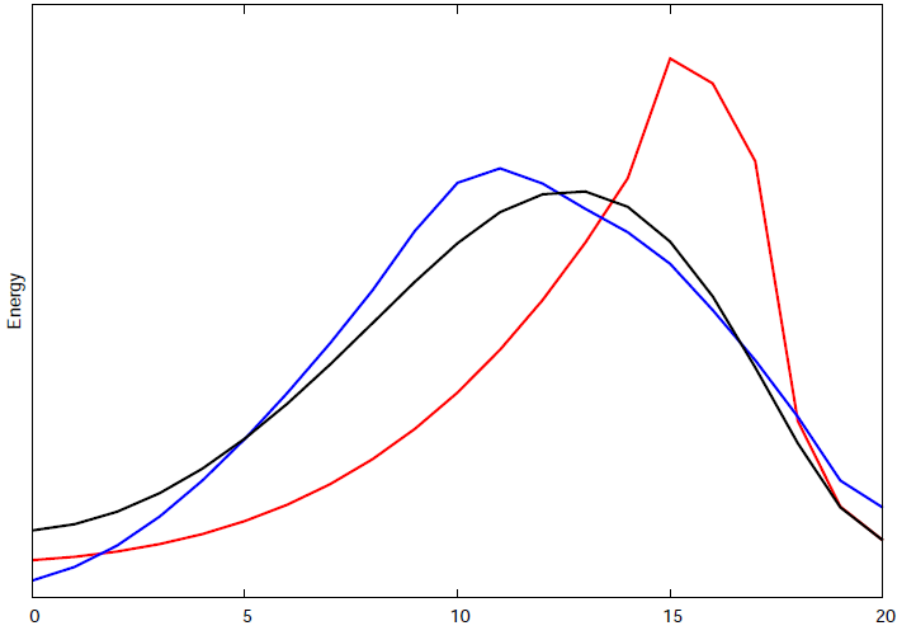
- force correlation
  DFT vs (REBO, GAP)
- σ = 0.26 eV/A
- 100 teaching points
- 2.5 A cutoff

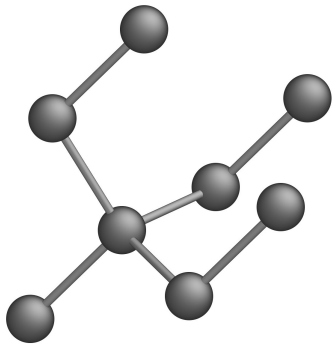


GAP test

- diamond to graphite transition along arbitrary reaction coordinate
- energy from DFT, REBO and GAP







Defining local energies

- how to obtain local energy from DFT
  - surface energy
  - visualisation
- best fit potential: use local energy
- equivalent to chemical potential of an atom
  - replacing atoms to infinity

Approximate local energies

- restricted part of configurational space
  - use force and energy information only along a minimisation from 'gas' phase or
  - blow up a configuration
- use it as post-processing tool
- visualise 'hot' atoms

Summary

- interpolation in high-dimensional space via GP
- target is quantum mechanical data
- extendable method
- carbon potential that captures sp^2-sp^3 transition
- new approach in defining local energies

- more than one atom type: $\rho(\mathbf{r}) = \sum_i c_i \delta(\mathbf{r} - \mathbf{r}_i)$
- electrostatics: subtract Coulomb energy