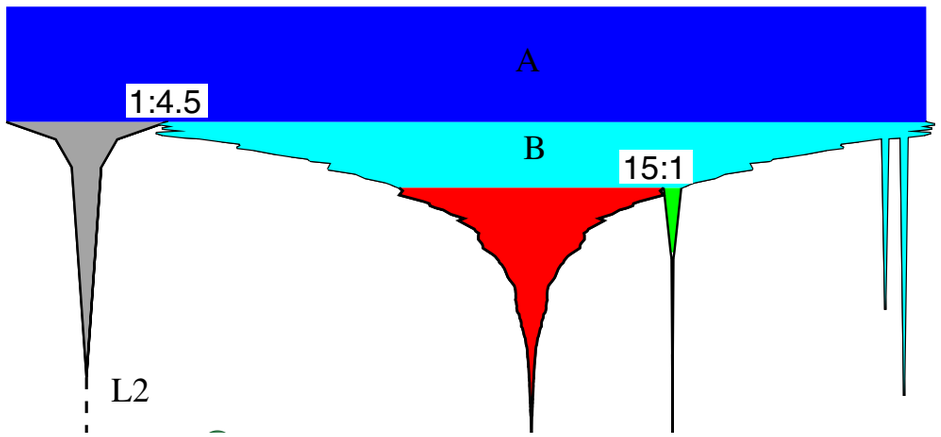
Applying machine learning to quantum mechanics

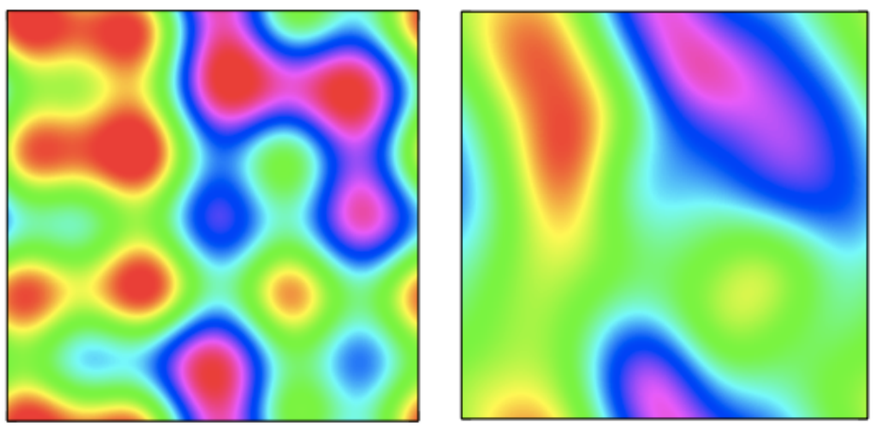
Gábor Csányi
Engineering Laboratory

UNIVERSITY OF
CAMBRIDGE

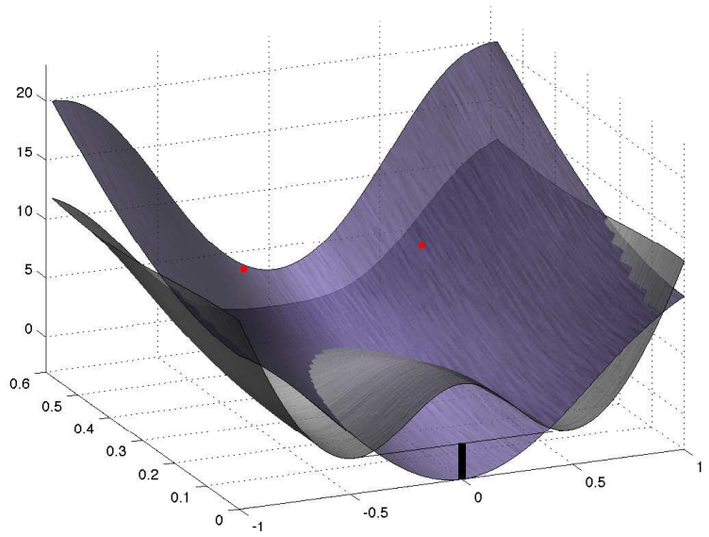
Configuration exploration:
Nested Sampling



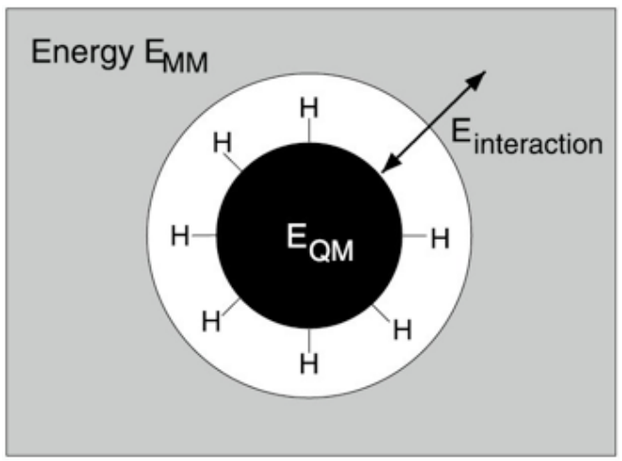
Free energy surface reconstruction
Bayesian Inference



Force-fields by
Machine Learning



Adaptive QM/MM
with particle exchange



Swift introduction to molecular modelling

The world works according to the Schrödinger equation:

$$-i\hbar\frac{\partial\Psi}{\partial t} = \mathcal{H}\Psi$$

$$\Psi \equiv \Psi(x_1, x_2, \ldots) \qquad x_i \in R^3$$

$$\mathcal{H} = \sum_i \frac{1}{2m_i}\nabla^2_{x_i} + \sum_{ij}\frac{Z_i Z_j}{|r_i - r_j|}$$

"matter waves"
tendency to spread

Long range Coulomb
attraction/repulsion
macroscopically observable!

For fermions, antisymmetry:

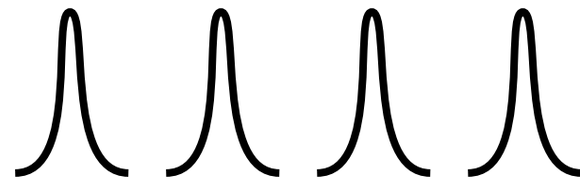
$$\Psi(\ldots, x_i, \ldots, x_j, \ldots) = -\Psi(\ldots, x_j, \ldots, x_i, \ldots)$$

- Range of validity: particles are electrons and atomic nuclei, stay the same
- Thus far not contradicted by experiment

Newtonian molecular dynamics

At ambient conditions nuclear Ψ do not overlap (other than for H…), separable solution:

$$\Psi = \Psi_1(R_1)\Psi_2(R_2)\ldots$$



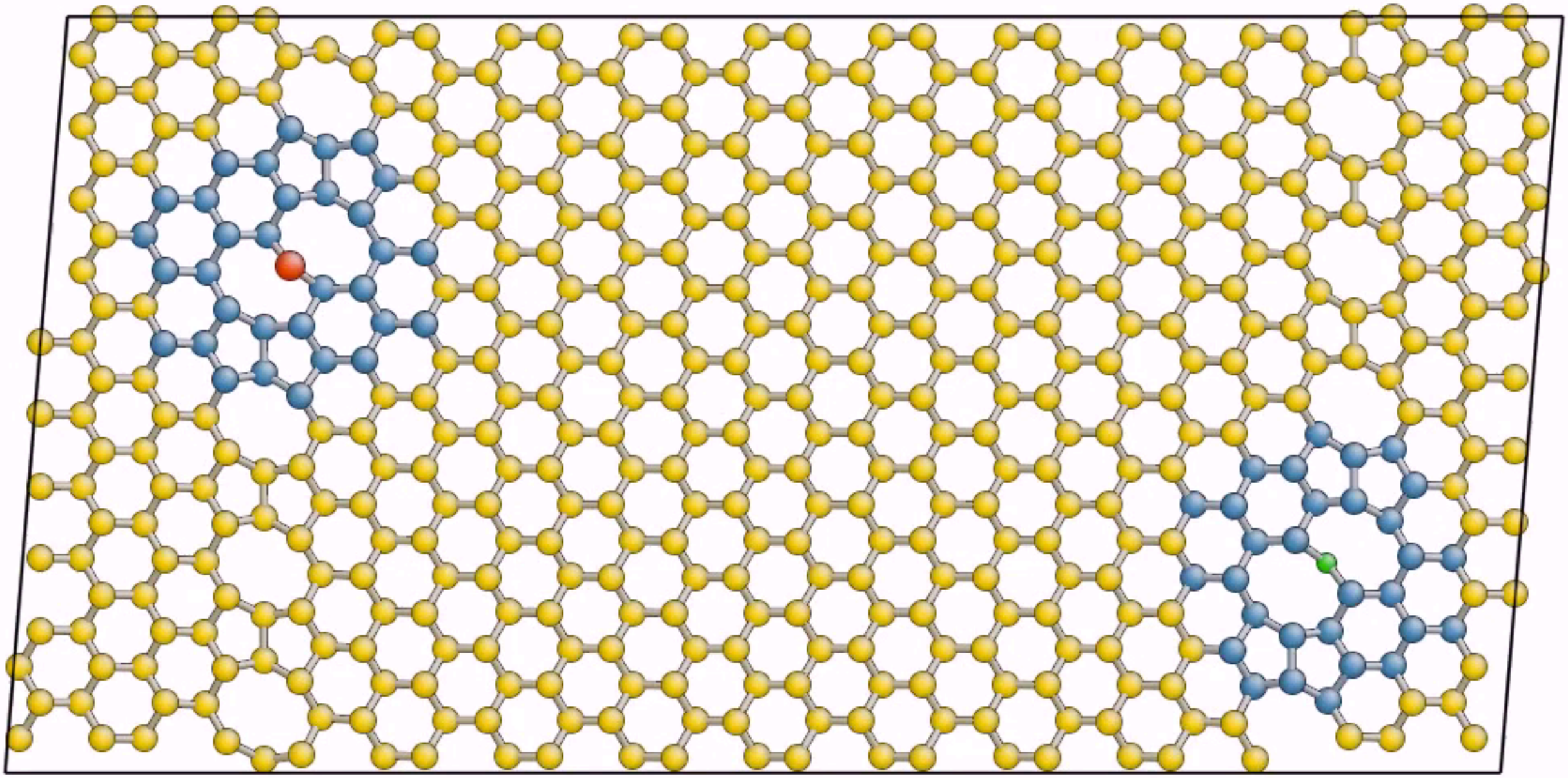
Correspondence principle: quantum system reduces to a classical system:

$$p_i = \langle\Psi|\nabla_{R_i}|\Psi\rangle$$
$$q_i = \langle\Psi|R_i|\Psi\rangle$$
$$H = \langle\Psi|\mathcal{H}|\Psi\rangle = \sum_i \frac{p_i^2}{2m_i} + \underbrace{\sum_{ij} \frac{Z_i Z_j}{|q_i - q_j|} + V_{\mathrm{el}}(q_1, q_2, \ldots)}$$

difficult electronic problem

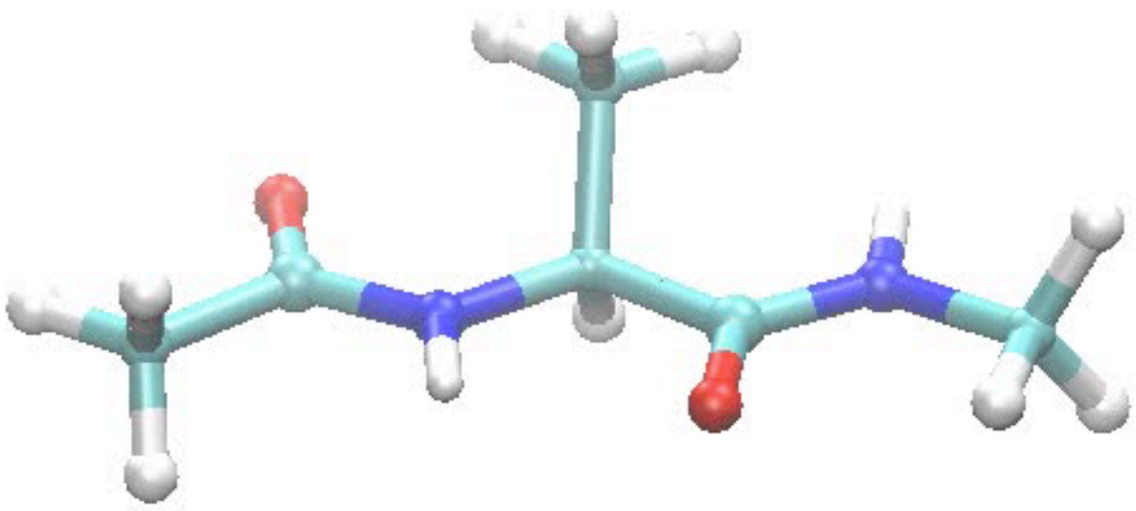
"Interatomic potential"
$V(q_1, q_2, \ldots)$

Typical questions



"What are the defects that control material response?"

Typical questions



"What are the typically observed conformations?"

Ingredients

- Representation of atomic neighbourhood → smoothness, faithfulness, continuity
- Interpolation of functions → flexible but smooth functional form, few sensible parameters
- Database of configurations → predictive power non-domain specific

Function fitting with basis functions

Fit a function $f(x)$ based on observations $\mathbf{y} \equiv \{y_i\}$ at $\{x_i\}$

$$f(x) = \sum_{i=1}^{N} \alpha_i k(x_i, x) \qquad \text{e.g. } k(x, x') = \sigma_w^2 e^{-|x-x'|^2/2\sigma^2}$$

$$y_j = \sum_{i=1}^{N} \alpha_i \left(k(x_i, x_j) + \sigma_\nu^2 \delta_{ij} \right)$$

regularised fit:
arbitrary σ,σ_w,σ_ν

$$\mathbf{y} = (\mathbf{K} + \sigma_\nu^2 \mathbf{I})\boldsymbol{\alpha}$$

$$\boldsymbol{\alpha} = \mathbf{C}^{-1}\mathbf{y}$$

$$f(x) = \mathbf{k}^T \mathbf{C}^{-1} \mathbf{y}$$

$$[\mathbf{K}]_{ij} \equiv k(x_i, x_j)$$

$$\mathbf{C} \equiv \mathbf{K} + \sigma_\nu^2 \mathbf{I}$$

$$\mathbf{k} \equiv k(x_i, x)$$

Machine learning framework: Kernel regression

$$\varepsilon(\mathbf{q}^{(i)}) = \sum_k^N \alpha_k K(\mathbf{q}^{(i)}, \mathbf{q}^{(k)})$$

- Linear regression:

$$K_{\mathrm{DP}}(\mathbf{q}^{(i)}, \mathbf{q}^{(k)}) = \mathbf{q}^{(i)} \cdot \mathbf{q}^{(k)} \quad \longrightarrow \quad \varepsilon(\mathbf{q}^{(i)}) = \sum_j q_j^{(i)} \sum_k^N \alpha_k q_j^{(k)} = \mathbf{q}^{(i)} \cdot \boldsymbol{\beta}$$

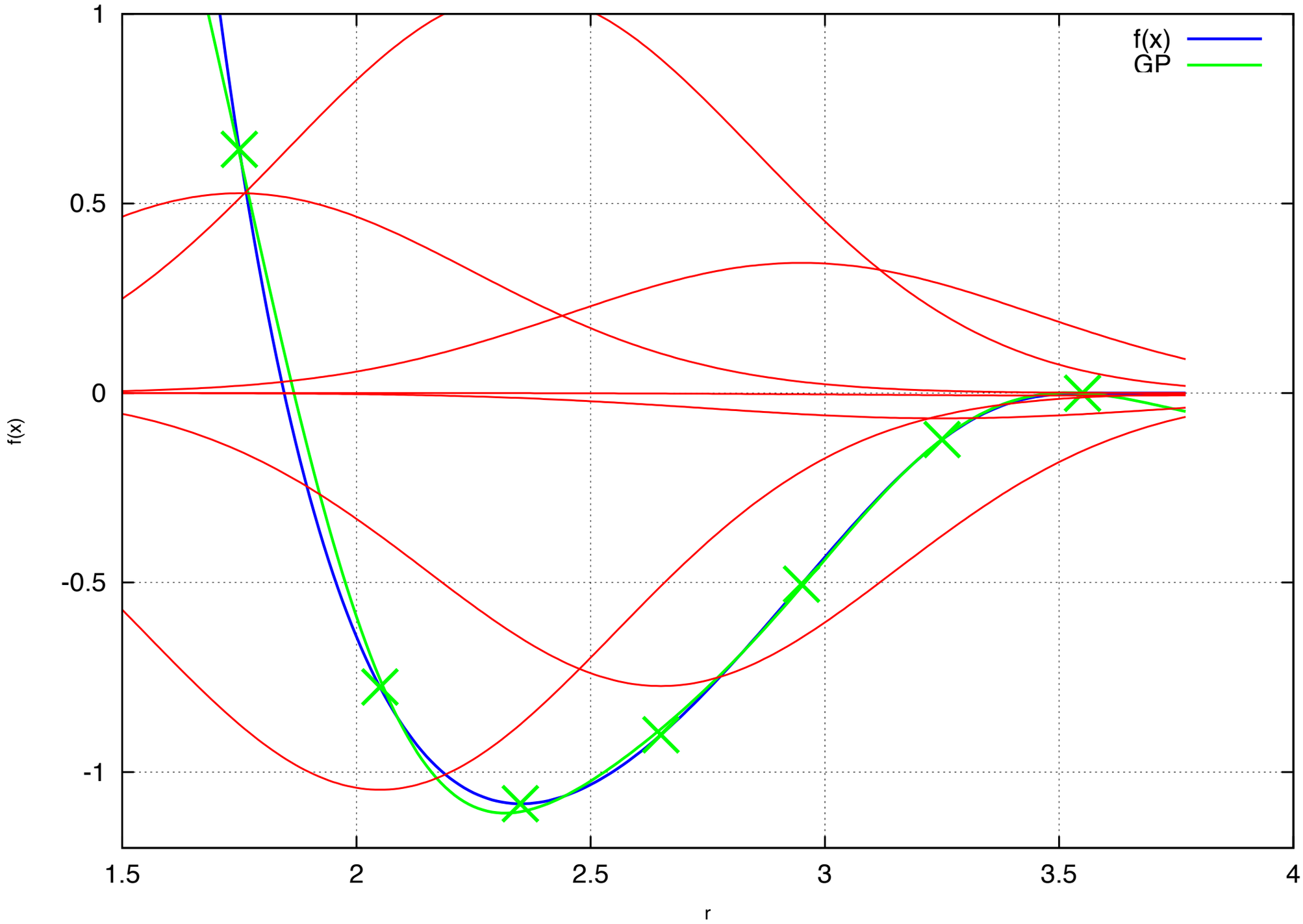
- Neural networks

$$K_{\mathrm{NN}}(\mathbf{q}^{(i)}, \mathbf{q}^{(k)}) = -|\mathbf{q}^{(i)} - \mathbf{q}^{(k)}|^2 + \mathrm{const.}$$

- Gaussian kernel

$$K_{\mathrm{SE}}(\mathbf{q}^{(i)}, \mathbf{q}^{(k)}) = \exp\left(-\sum_j \frac{(q^{(i)} - q^{(k)})^2}{2\sigma_j^2}\right)$$

1D example



Gaussians basis functions are wide!

$$f(x) = \mathbf{k}^T \mathbf{C}^{-1} \mathbf{y}$$

Bayesian function fitting: Gaussian Process Regression

Fit a function $f(x)$ based on observations $\mathbf{y} \equiv \{y_i\}$ at $\{x_i\}$

$$f(x) = \sum_{h=1}^{H} w_h \phi_h(x)$$

basis functions $\phi_h(x)$

$$f_i = \sum_h R_{ih} w_h$$

$$R_{ih} \equiv \phi_h(x_i)$$

Gaussian prior: $P(\mathbf{w}) = \text{Normal}(0, \sigma_w^2 \mathbf{I})$

Expected function Var

$$P(\mathbf{f}) = \text{Normal}(0, \sigma_w^2 \mathbf{R}\mathbf{R}^T)$$

Var of data

Gaussian likelihood: $\mathbf{y} = \mathbf{f} + \varepsilon \qquad \varepsilon \sim \text{Normal}(0, \sigma_\nu^2 \mathbf{I})$

Joint of observations also Gaussian:

$$P(\mathbf{y}) = \text{Normal}(0, \sigma_w^2 \mathbf{R}\mathbf{R}^T + \sigma_\nu^2 \mathbf{I})$$

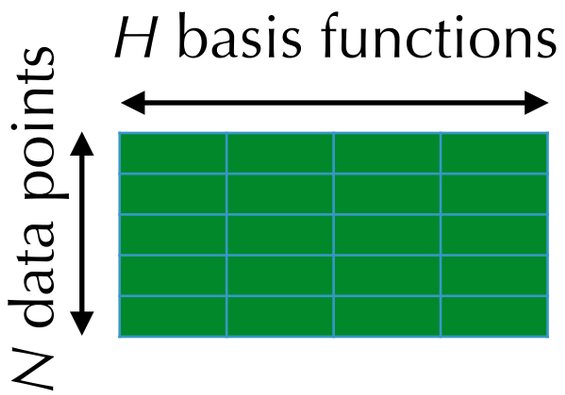
Predict the next observation using the conditional

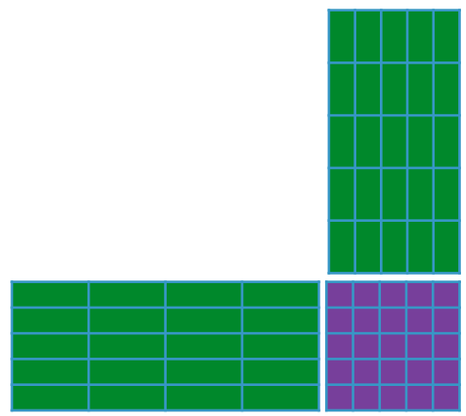
$$P(y_{N+1}|\mathbf{y}_N) = P(y_{N+1}, \mathbf{y}_N)/P(\mathbf{y}_N)$$

$$f(x) = \mathbf{k}^T \mathbf{C}^{-1} \mathbf{y}$$

Limit of ∞ basis functions

$$R_{ih} \equiv \phi_h(x_i)$$





$\mathbf{RR}^{\mathrm{T}}$

Use Gaussian basis functions, take $H \to \infty$, $\mathbf{RR}^{\mathrm{T}}$ is finite

Prior is a Gaussian Process with

$$\mathrm{Cov}[f(x_i), f(x_j)] \propto \exp\left[-(x_i - x_j)^2/\sigma^2\right]$$

Similarly for the observations:

$$\mathrm{Cov}[y(x_i), y(x_j)] \propto \exp\left[-(x_i - x_j)^2/\sigma^2\right] + \sigma_\nu^2 \delta_{ij}$$

$$f(x) = \mathbf{k}^T \mathbf{C}^{-1} \mathbf{y}$$

Gaussian Process connection to linear fit

$$P(y_{N+1}|\mathbf{y}_N) = P(y_{N+1}, \mathbf{y}_N)/P(\mathbf{y}_N) \propto P(y_{N+1}, \mathbf{y}_N)$$

$$P(y_1 \dots y_N) \propto \text{Normal}(0, \mathbf{C}_N) \propto \exp(-\mathbf{y}_N^T \mathbf{C}_N^{-1} \mathbf{y}_N)$$

$$P(y_{N+1}, \mathbf{y}_N) \propto \text{Normal}(0, \mathbf{C}^{N+1}) \propto \exp(-[\mathbf{y}_N^T\, y_{N+1}] \mathbf{C}_{N+1}^{-1} [\mathbf{y}_N\, y_{N+1}])$$

$$\mathbf{C}_{N+1} \equiv \begin{bmatrix} \begin{bmatrix} & \mathbf{C}_N & \end{bmatrix} & \begin{bmatrix} \mathbf{k} \end{bmatrix} \\ \begin{bmatrix} \mathbf{k}^T \end{bmatrix} & [\kappa] \end{bmatrix} \quad \mathbf{C}_{N+1}^{-1} \equiv \begin{bmatrix} \begin{bmatrix} & \mathbf{M} & \end{bmatrix} & \begin{bmatrix} \mathbf{m} \end{bmatrix} \\ \begin{bmatrix} \mathbf{m}^T \end{bmatrix} & [\mu] \end{bmatrix}$$

$$[\mathbf{k}]_i \equiv C(x_i, x_{N+1})$$

$$\kappa \equiv C(x_{N+1}, x_{N+1})$$

$$\begin{aligned} \mu &= \left(\kappa - \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{k}\right)^{-1} \\ \mathbf{m} &= -\mu\, \mathbf{C}_N^{-1} \mathbf{k} \\ \mathbf{M} &= \mathbf{C}_N^{-1} + \frac{1}{\mu} \mathbf{m}\mathbf{m}^T \end{aligned}$$

$$[\mathbf{y}_N^T\, y_{N+1}] \mathbf{C}_{N+1}^{-1} [\mathbf{y}_N\, y_{N+1}] = \mathbf{y}_N^T \mathbf{M} \mathbf{y} + 2\mathbf{y}^T \mathbf{m} y_{N+1} + \mu y_{N+1}^2$$
$$= \mu(y_{N+1} + \mathbf{y}^T \mathbf{m}/\mu)^2 + \dots$$

$$P(y_{N+1}|\mathbf{y}_N) \propto \exp(-(y_{N+1} - \bar{y})^2 / 2\hat{\sigma}^2)$$

$$\bar{y} = \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{y}_N$$

$$\hat{\sigma} = \kappa - \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{k}$$

$$\arg\max_{y_{N+1}} P(y_{N+1}|\mathbf{y}_N) = \bar{y} = \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{y}_N$$

$$f(x) = \mathbf{k}^T \mathbf{C}^{-1} \mathbf{y}$$

Gaussian Process Regression summary

- Choose Gaussian covariance:

$$K(x_i, x_j) = \exp(-(x_i - x_j)^2/2\sigma^2)$$ Prior assumption

$$f(x) = \arg\max_f P(f|\text{data}) = \sum_i \alpha_i K(x, x_{(i)})$$

Maximum of posterior

$$\alpha = \mathbf{C}^{-1}\mathbf{y} \equiv (\sigma_w^2\mathbf{K} + \sigma_\nu^2\mathbf{I})^{-1}\mathbf{y}$$

$$C_{ii'} = \sigma_w^2 K(x_i, x_{i'}) + \sigma_\nu^2 \delta_{ii'}$$

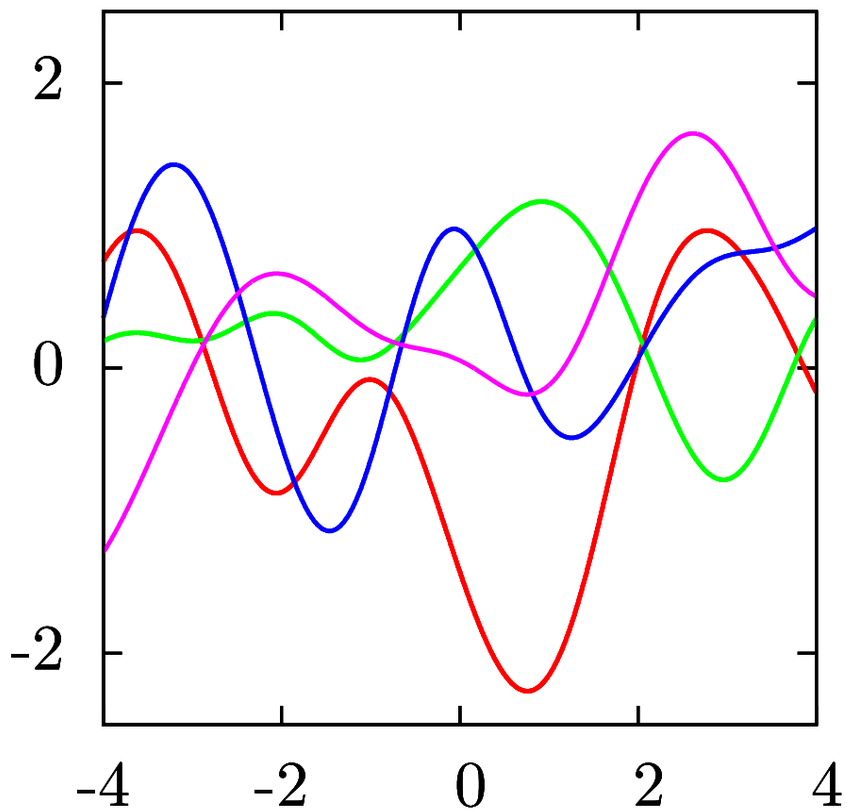
- Meaningful hyper-parameters:
  - σ: smoothness (*x*-scale) of f
  - σ_w: y-scale of f
  - σ_ν: variance (noise) of input data

- This is *not an optimised fit*, but a closed form estimate!

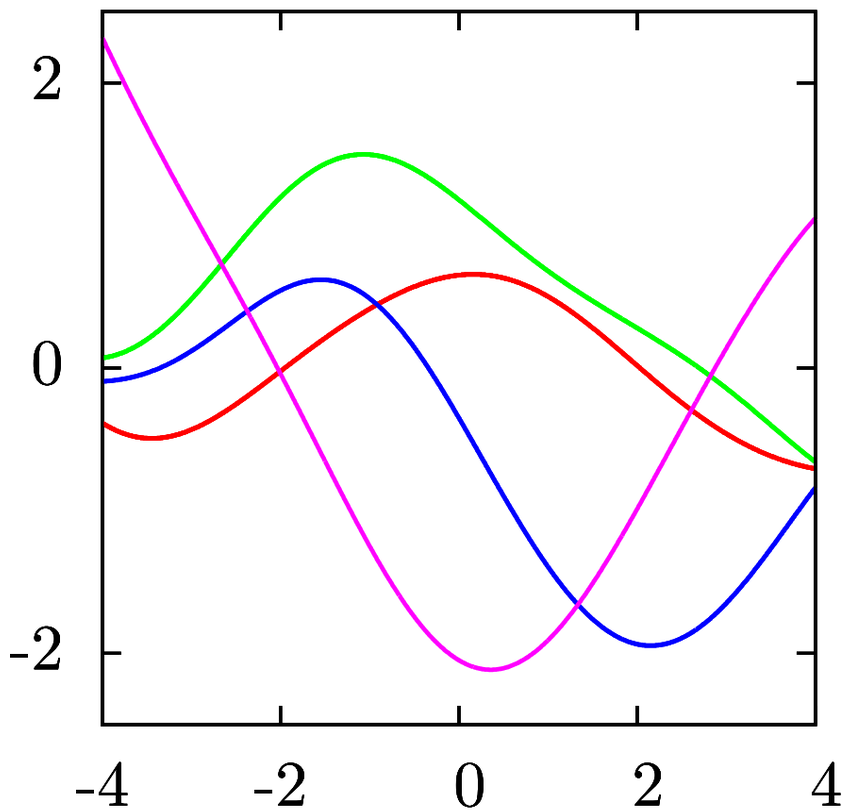
Samples from the Gaussian prior

$$P(f) \propto \text{Normal}(0, \mathbf{C})$$

$\sigma_f = 1.0$

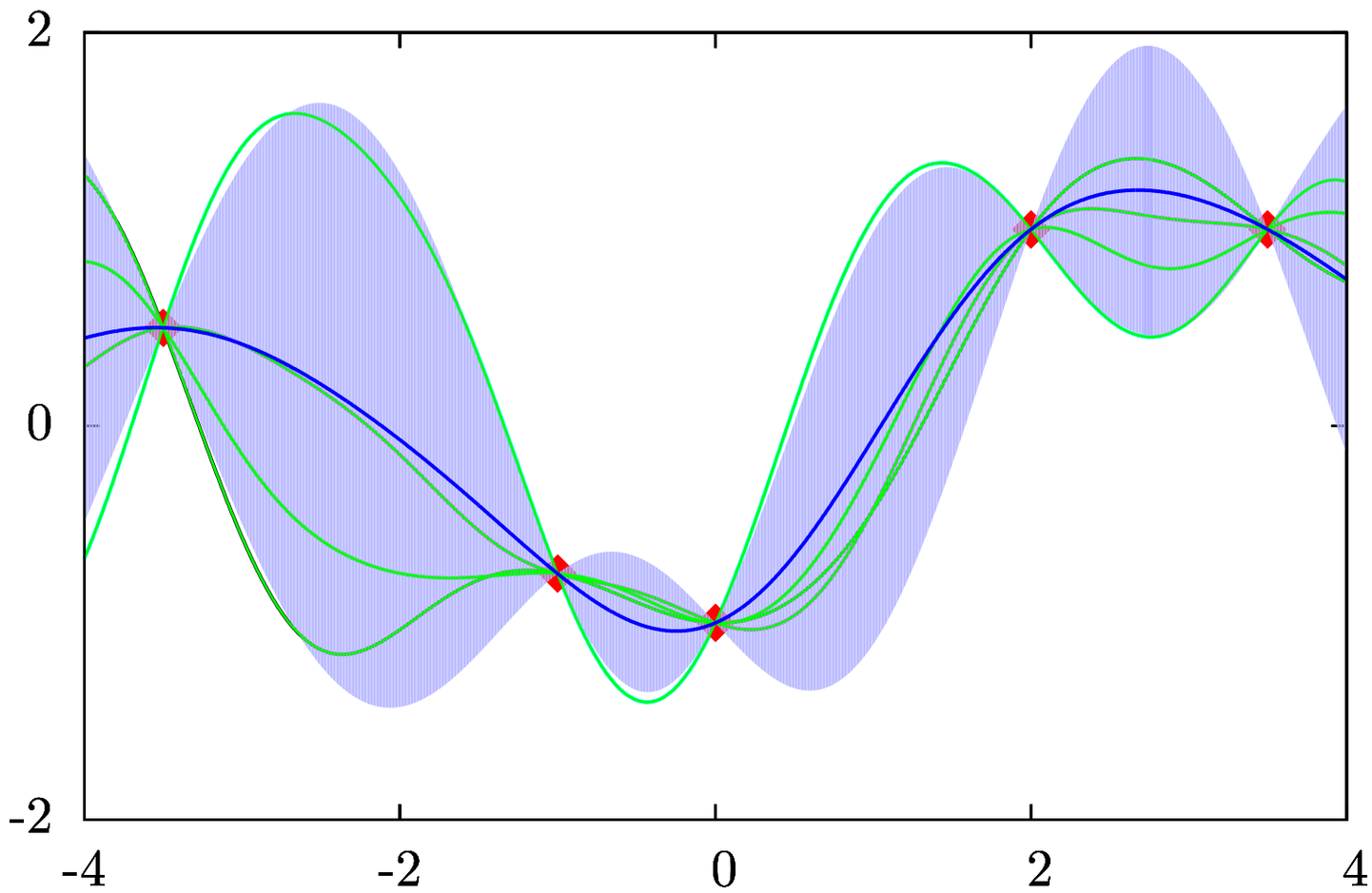


$\sigma_f = 2.0$



Samples from Gaussian posterior

$$P(y_{N+1}|\mathbf{y}) \propto \text{Normal}(\mathbf{k}^T\mathbf{C}^{-1}\mathbf{y}, \ldots)$$

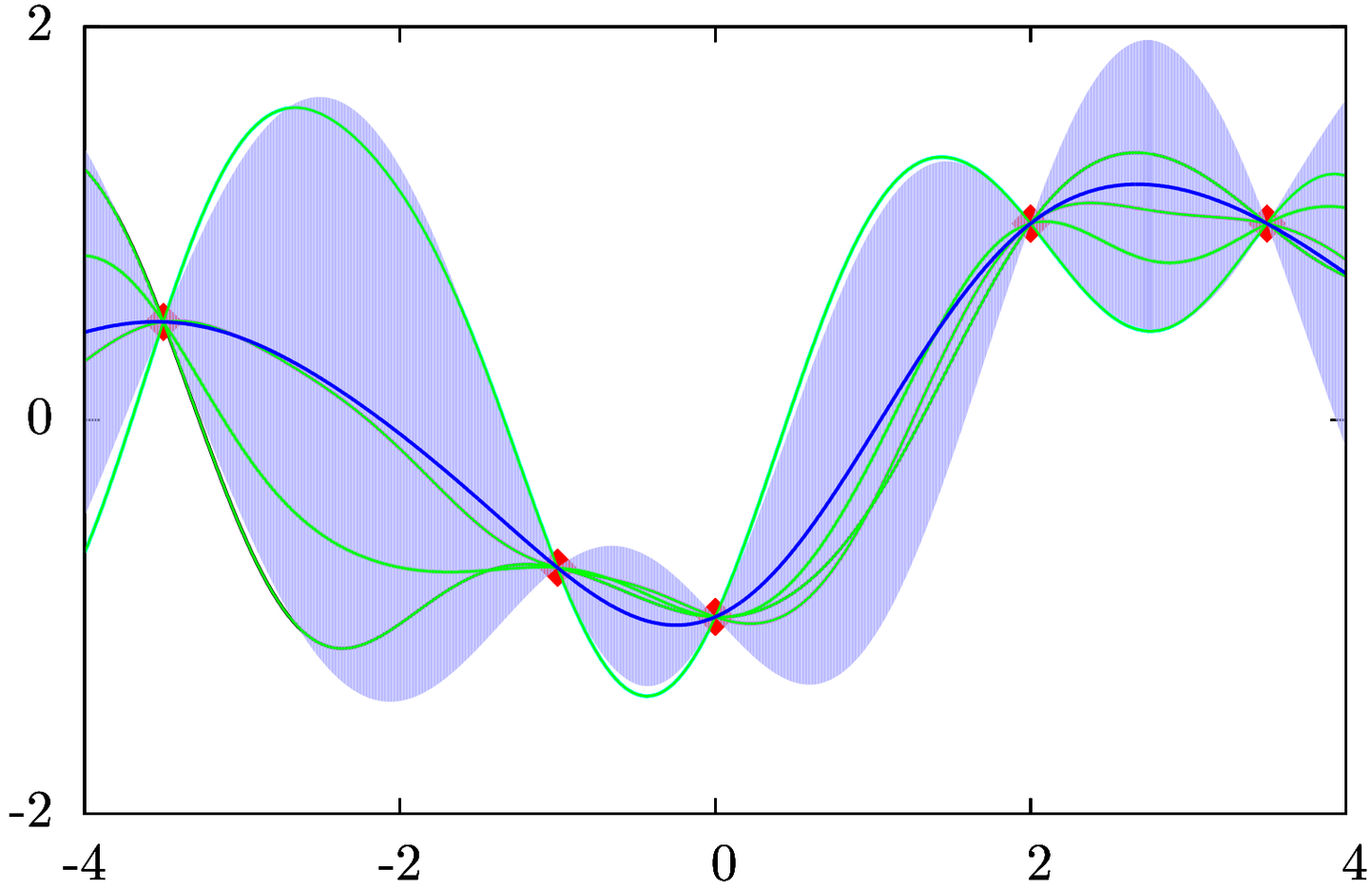


Mean of posterior distribution

$$\bar{y}(x) = \mathbf{k}^T(x)(\mathbf{K} + \sigma_\nu \mathbf{I})^{-1}\mathbf{y}$$

Kernel: $k(x_i, x_j) = e^{-|x_i - x'_j|^2/\sigma}$

$$[\mathbf{k}(x)]_i = k(x, x_i) \quad K_{ij} = k(x_i, x_j)$$



Generalised input data

- Often direct measurements of $\mathbf{y}_i$ not available
- Consider any linear operator $L(\mathbf{y})$
  Compute $P(y_{N+1}(x_{N+1})|L(\mathbf{y}))$

  L can be built using Σ , $\partial/\partial x$ etc
- Can deal with gradient data, sums, etc.

Interpolating the solution of an eigenproblem

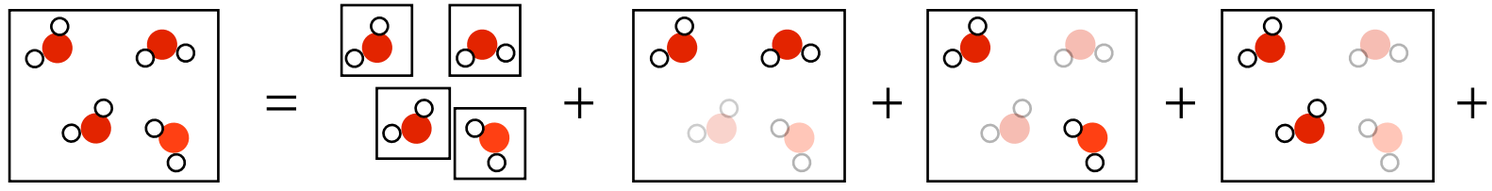
$$-i\hbar\frac{\partial\Psi}{\partial t} = \mathcal{H}\Psi \qquad \mathcal{H}_{\mathrm{el}} = \frac{1}{2}\sum_i \nabla_{r_i}^2 + \sum_{ij}\frac{Z_j(-1)}{|R_j - r_i|} + \sum_{ij}\frac{1}{|r_i - r_j|}$$

Lowest eigenstate: $E_0 \equiv V_{\mathrm{el}}(R_1, R_2, \ldots)$ has spatial locality

---

Weak

Many-body
(cluster) expansion



+ higher multi-body terms... + analytic long range terms

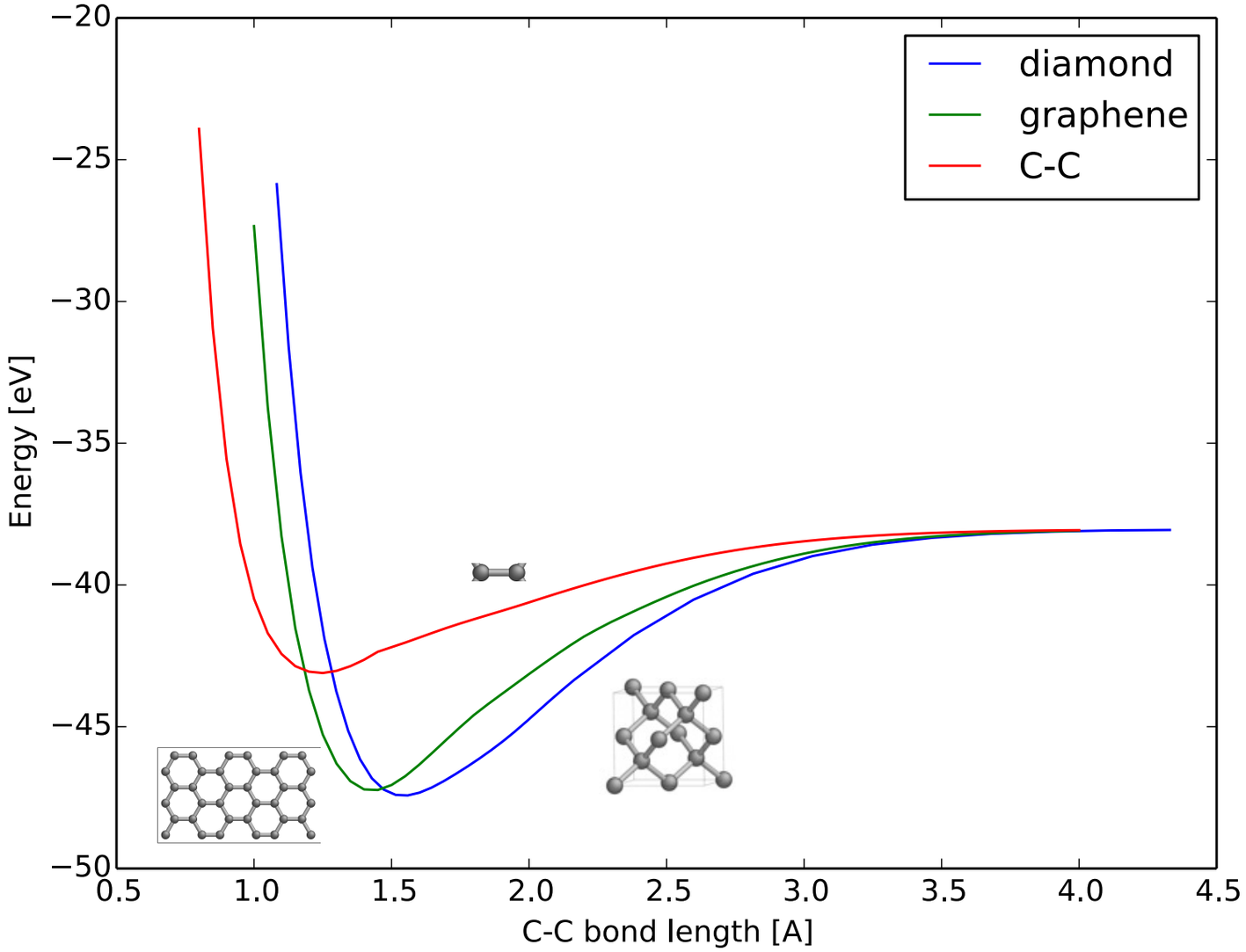
Strong

$$E = \sum_i \varepsilon(q_1^{(i)}, q_2,^{(i)} \ldots, q_M^{(i)})$$ + analytic long range terms

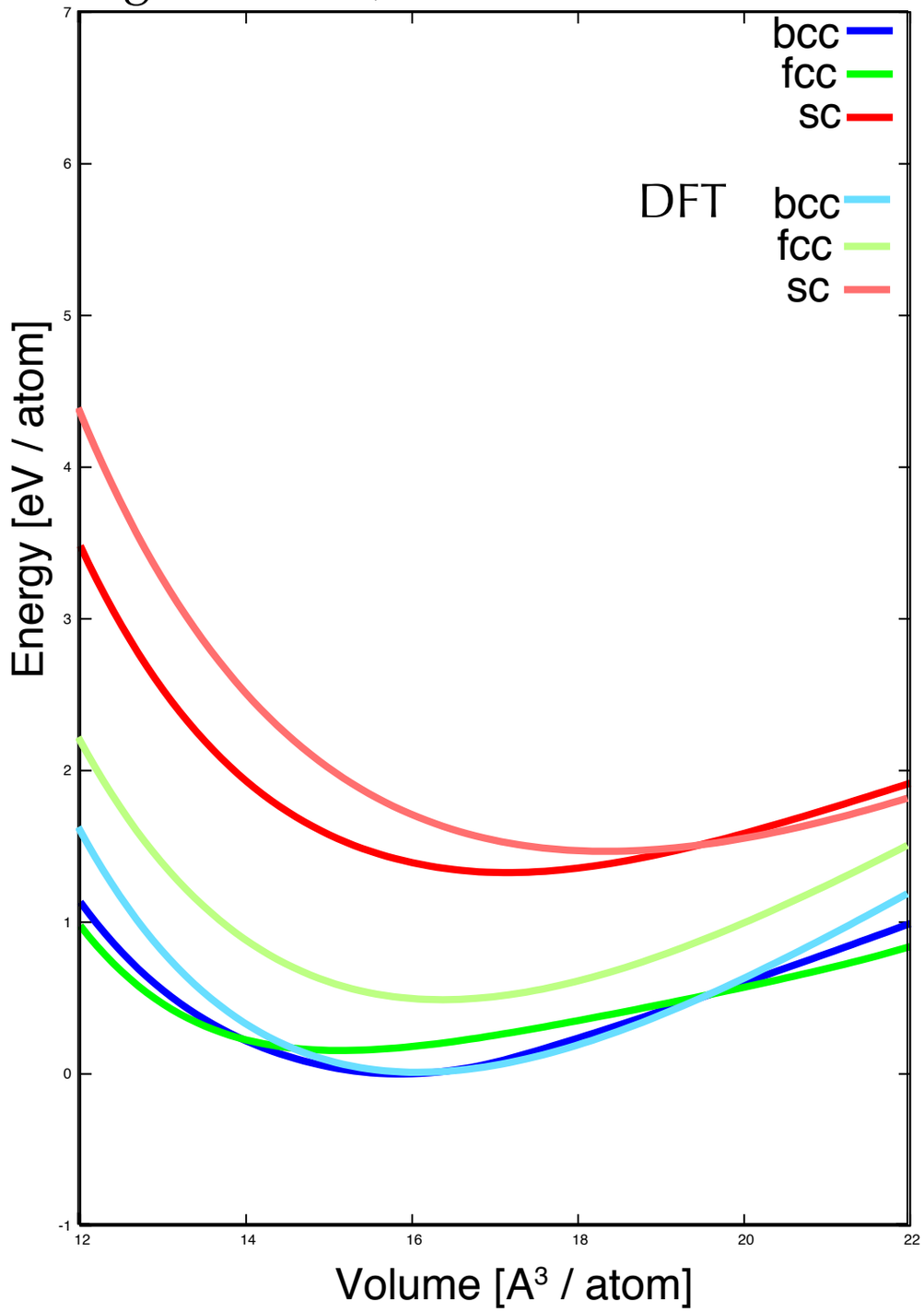
q_i are descriptors of local atomic neighbourhood

Quantum mechanics is many-body

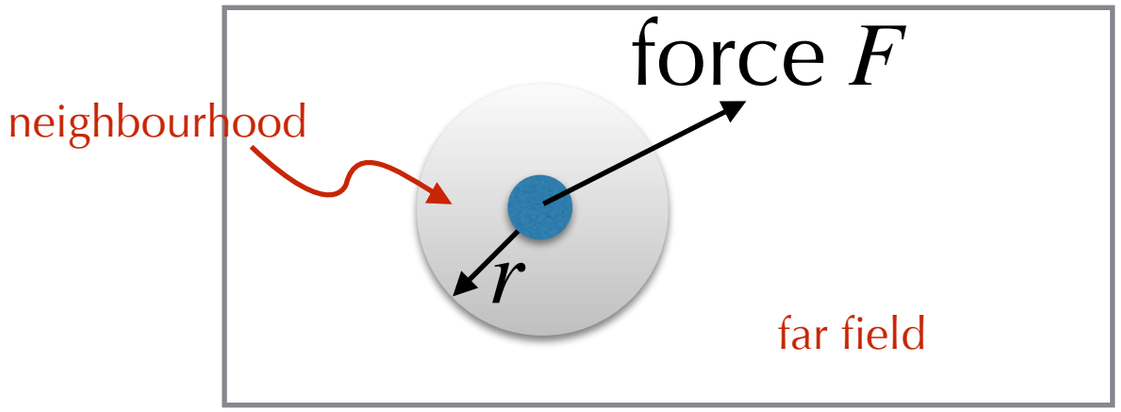
Carbon: tight binding



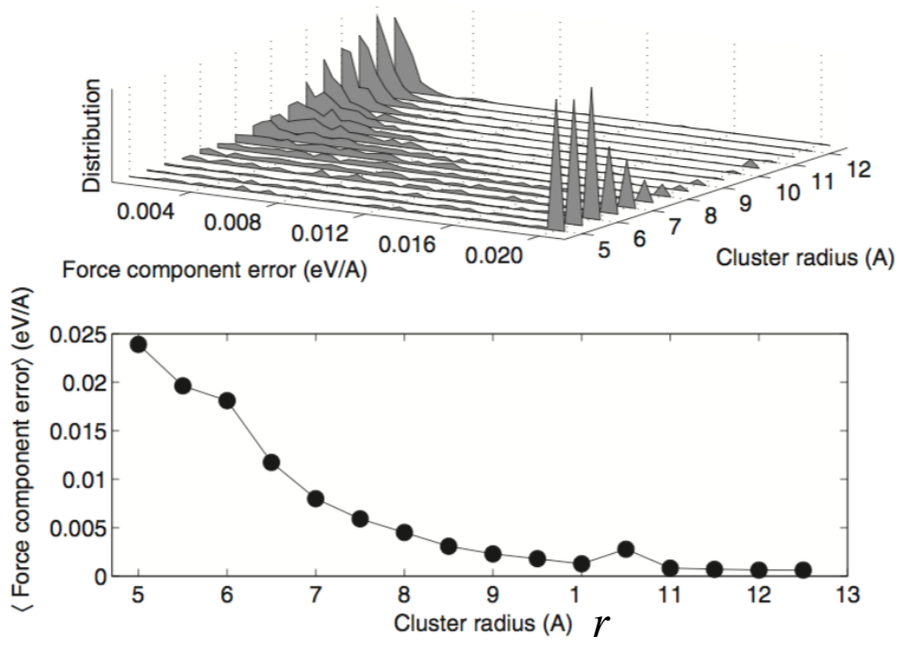
Tungsten: DFT, Embedded Atom Model



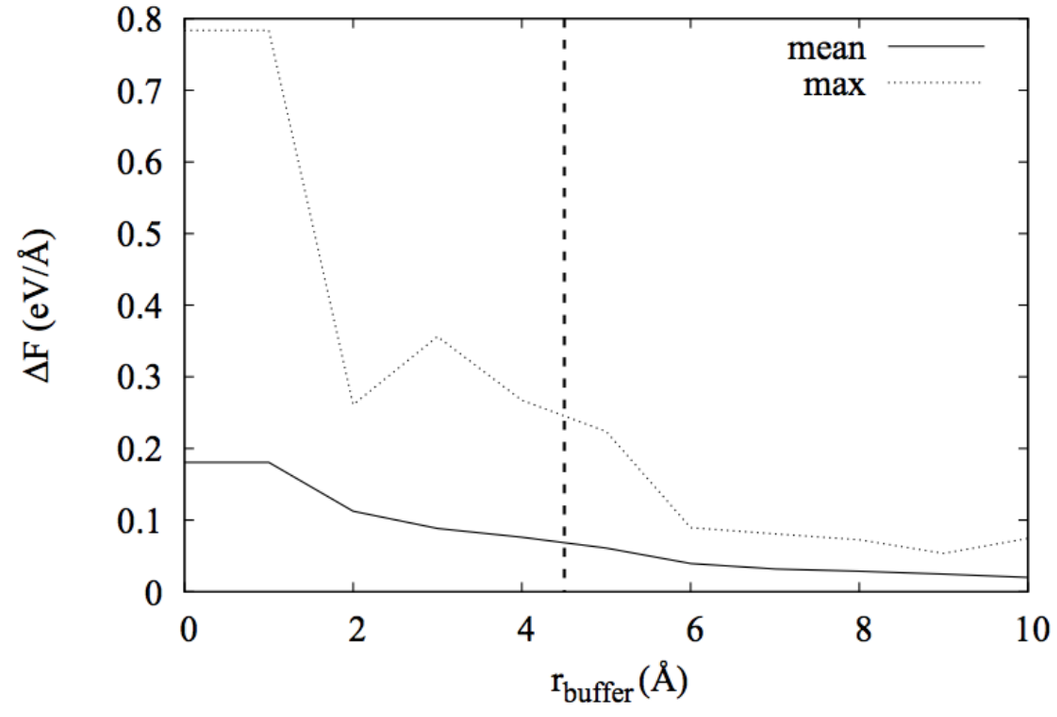
Quantum mechanics has some locality



Force errors around
Si self interstitial



Force errors around
O in water with QM/MM



Is there a finite range atomic energy function that gives the correct total energy and forces?

Traditional ideas for functional forms

- Pair potentials: Lennard-Jones, RDF-derived, etc.
- Three-body terms: Stillinger-Weber, MEAM, etc.
- Embedded Atom (no angular dependence)
- Bond Order Potential (BOP)
  Tight-binding-derived attractive term with pair-potential repulsion
- ReaxFF: kitchen-sink + hundreds of parameters

$$\varepsilon_i = \frac{1}{2}\sum_j V_2(|r_{ij}|) + \sum_{jk} k(\theta_{ijk} - \theta_0)^2$$

$$\varepsilon_i = \Phi\left(\sum_j \rho(|r_{ij}|)\right)$$

Representation is implicit

These are NOT THE CORRECT functions.
Limited accuracy, not systematic

given by

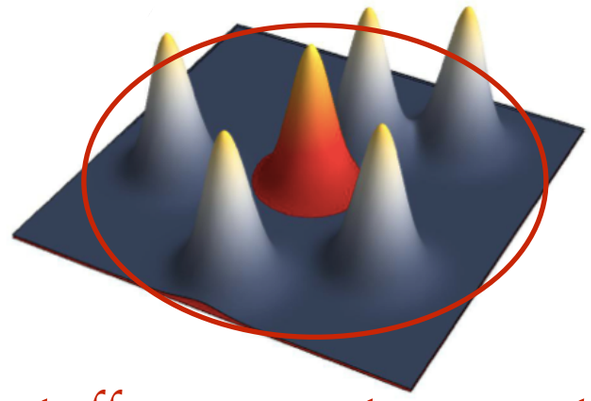
GOAL: potentials ~~based on~~ quantum mechanics

Construct smooth similarity kernel directly

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

- Overlap integral

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



cutoff: compact support

- Integrate over all 3D rotations:

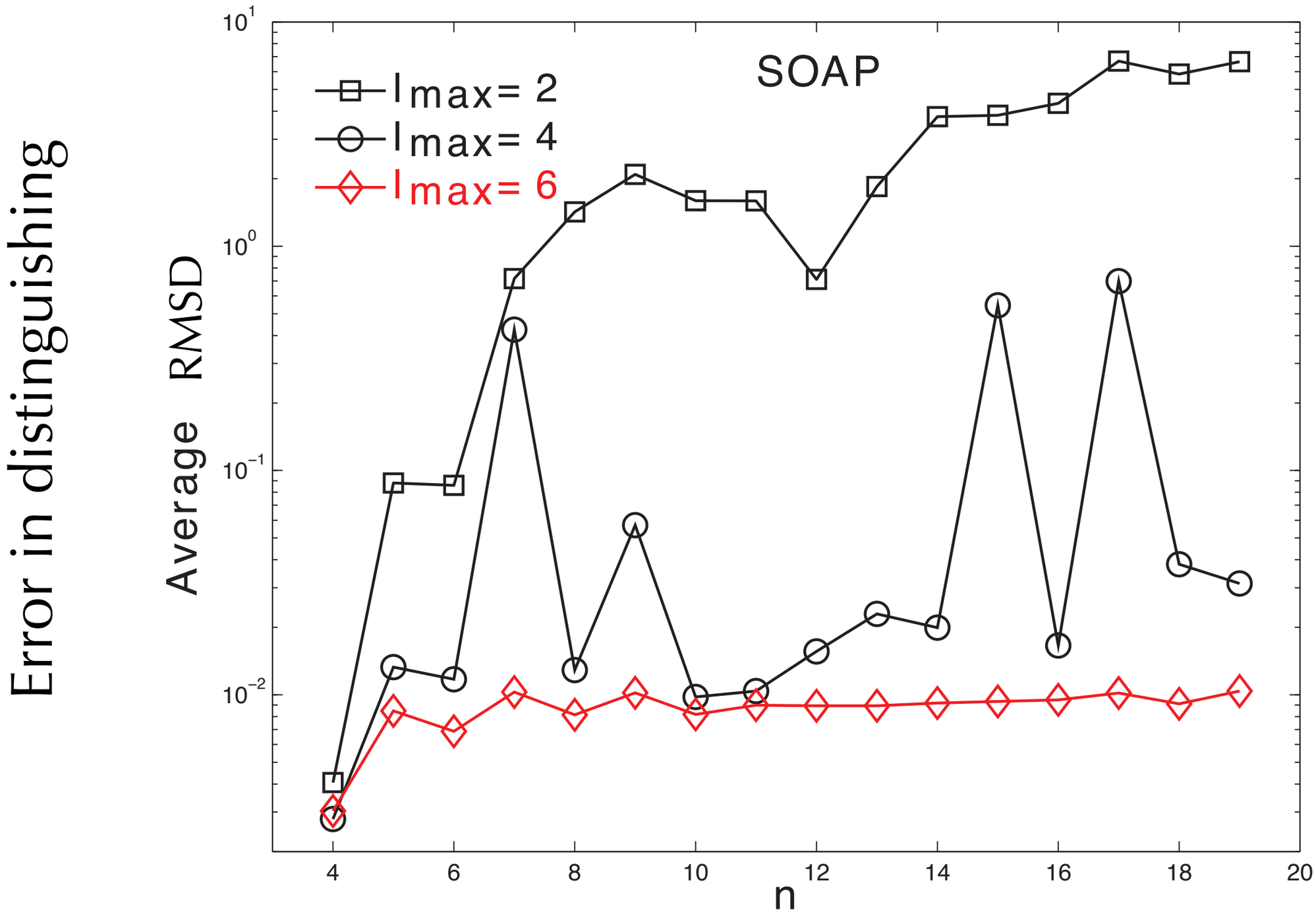
$$k(\rho_i, \rho_{i'}) = \int \left|S(\rho_i, \hat{R}\rho_{i'})\right|^2 d\hat{R} = \int d\hat{R} \left|\int \rho_i(\mathbf{r})\rho_{i'}(\hat{R}\mathbf{r})d\mathbf{r}\right|^2$$

- After LOTS of algebra: SOAP kernel

$$k(\rho_i, \rho_{i'}) = \sum_{n,n',l} p_{nn'l}^{(i)} p_{nn'l}^{(i')} \qquad p_{nn'l} = \mathbf{c}_{nl}^{\dagger}\mathbf{c}_{n'l}$$

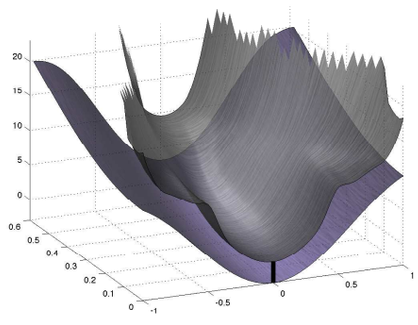
$$\boxed{K(\mathbf{q}, \mathbf{q}') \propto |k(\rho, \rho')|^{\xi}}$$

Smooth Overlap of Atomic Positions (SOAP)

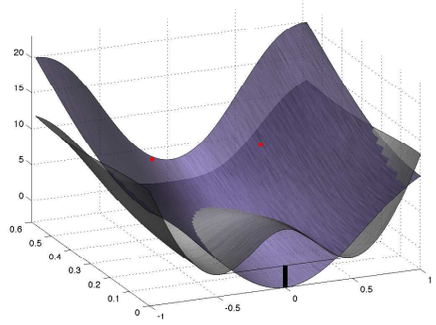


Number of neighbours

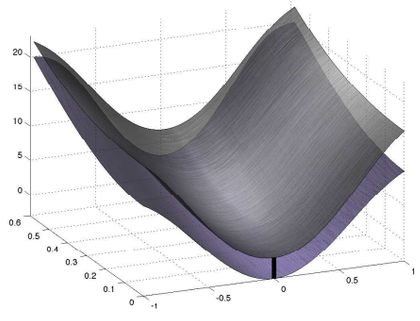
Diamond



Brenner potential



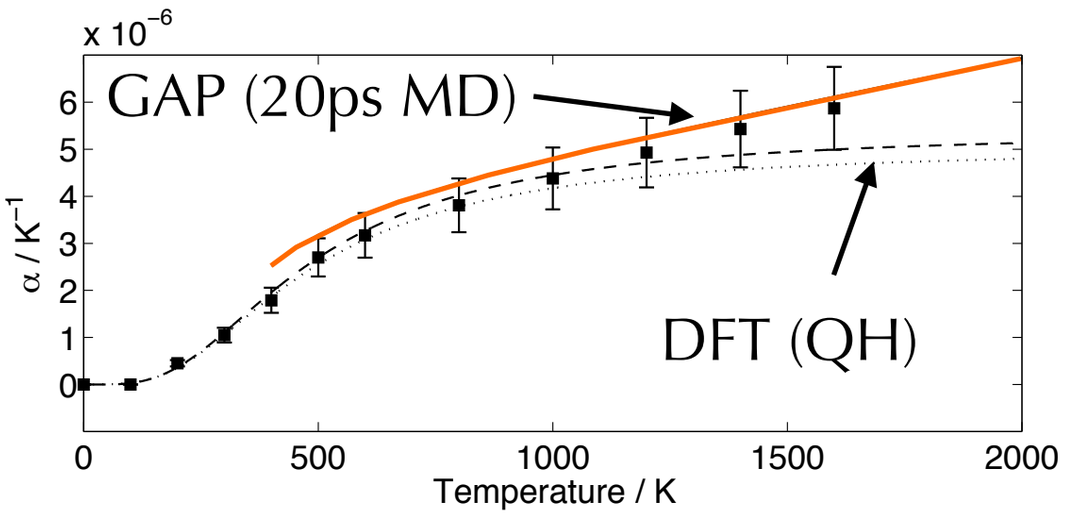
1 data point



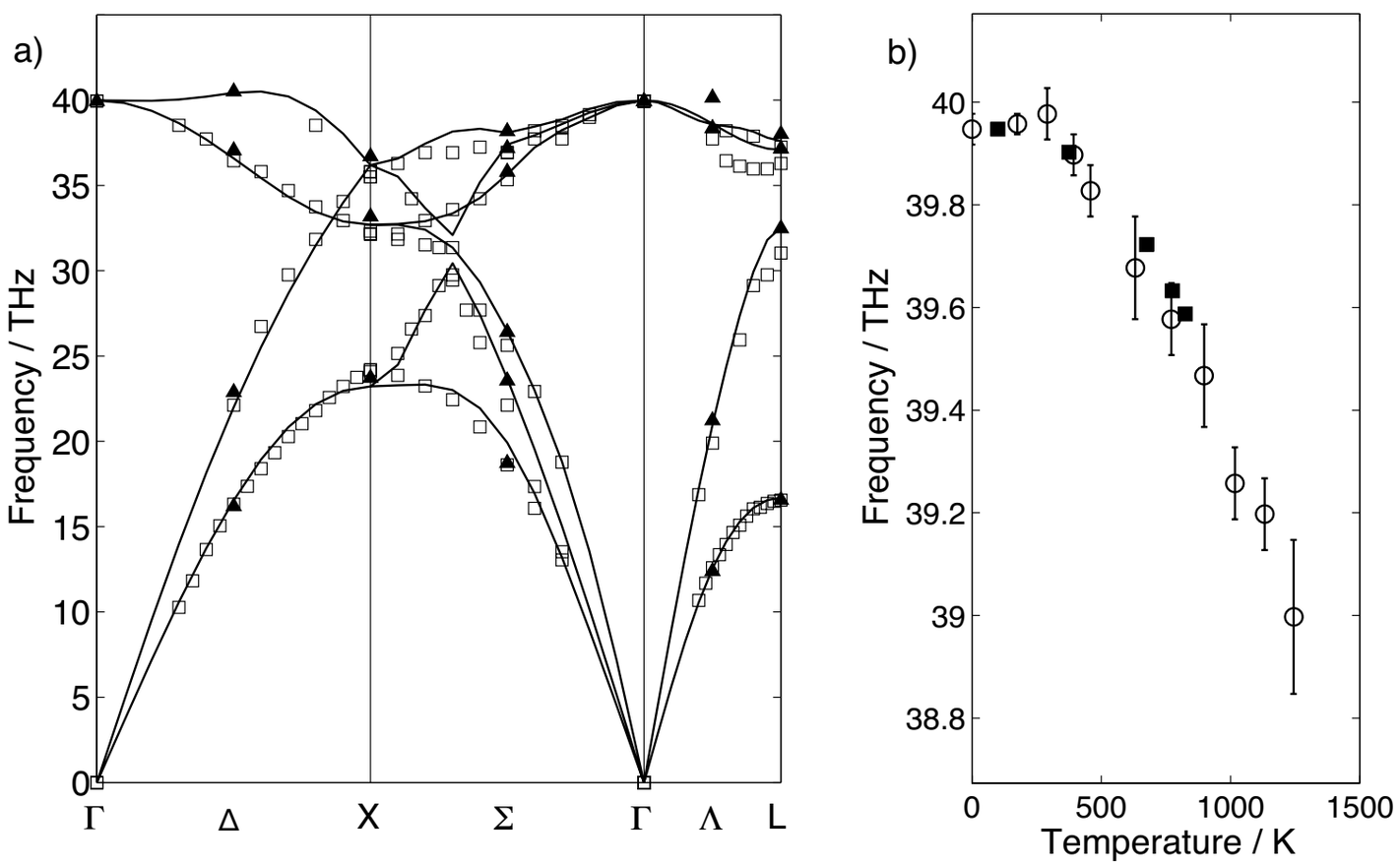
50 random data points

300 configurations from ab initio MD:

	DFT	GAP	Brenner
C_{11}	1118	1081	1061
C_{12}	151	157	133
C^0_{44}	610	608	736
C_{44}	603	601	717



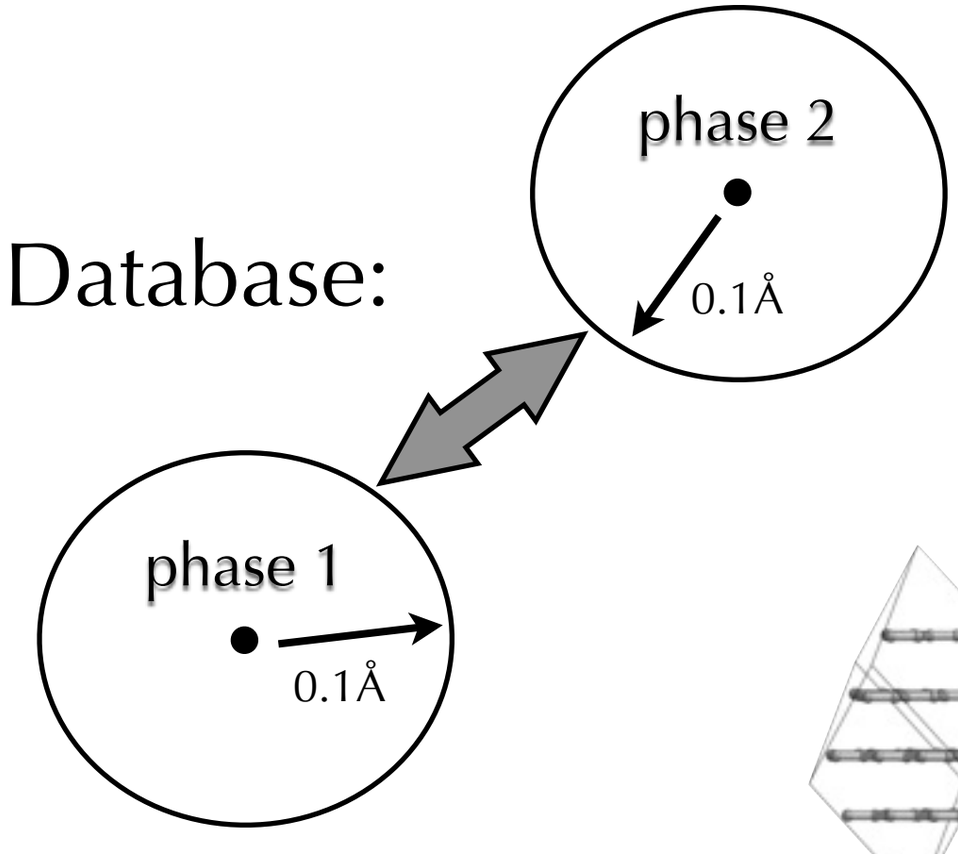
Thermal expansion

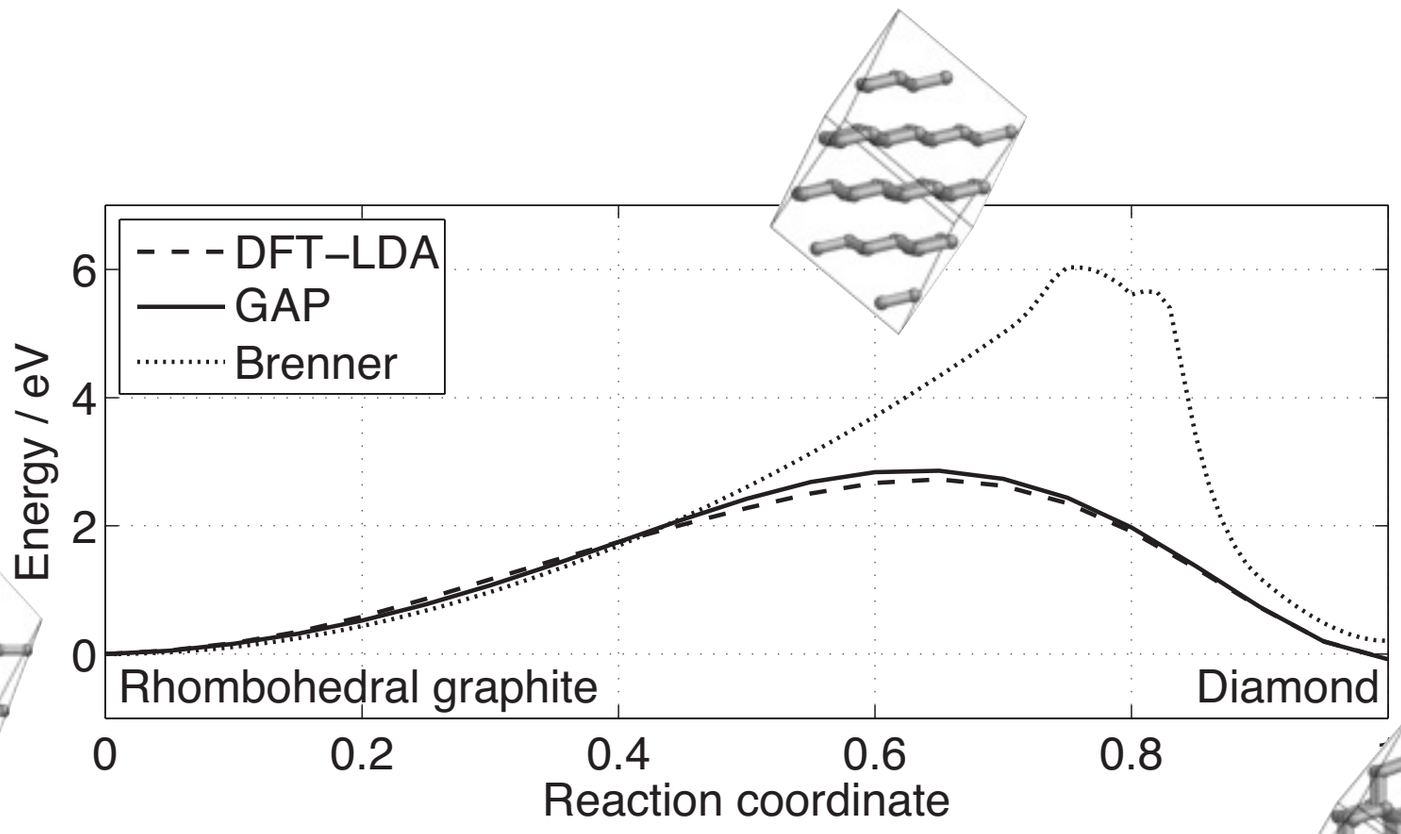


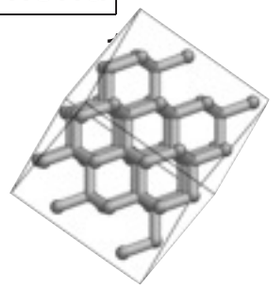
Phonon spectrum

Reactive potentials

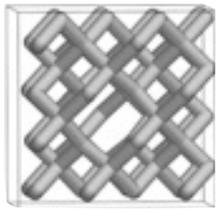
Database:

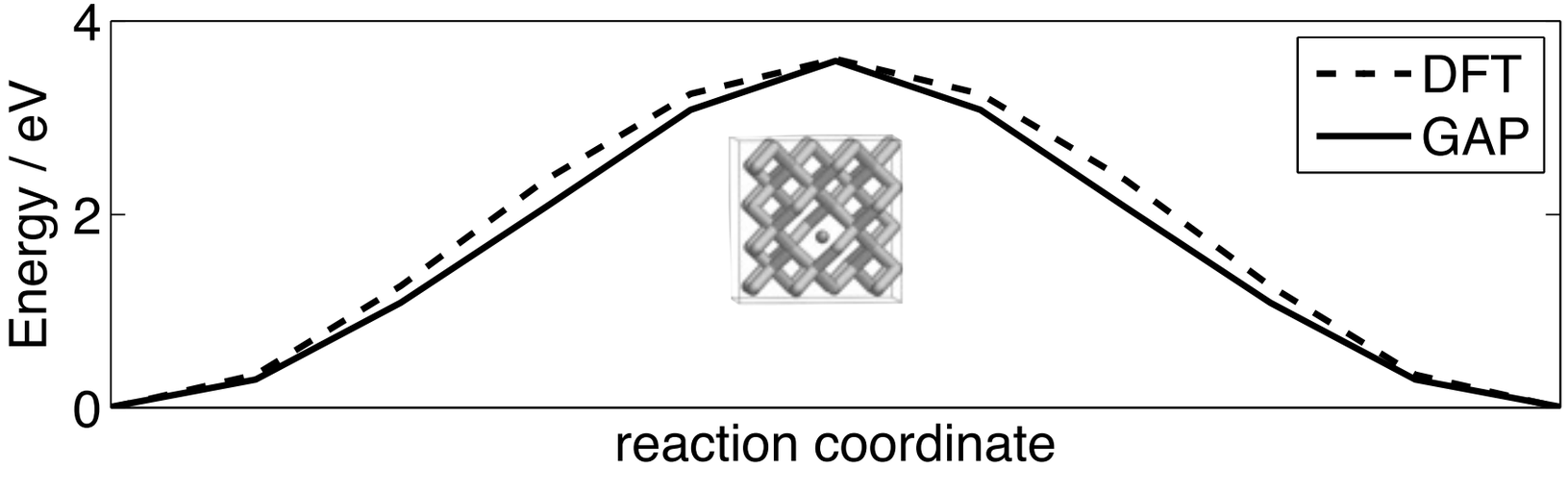


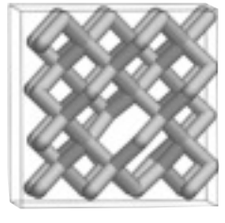




Vacancy migration







How to generate databases?

- Target applications: large systems
- Capability of full quantum mechanics (QM): small systems

QM MD on "representative" small systems:

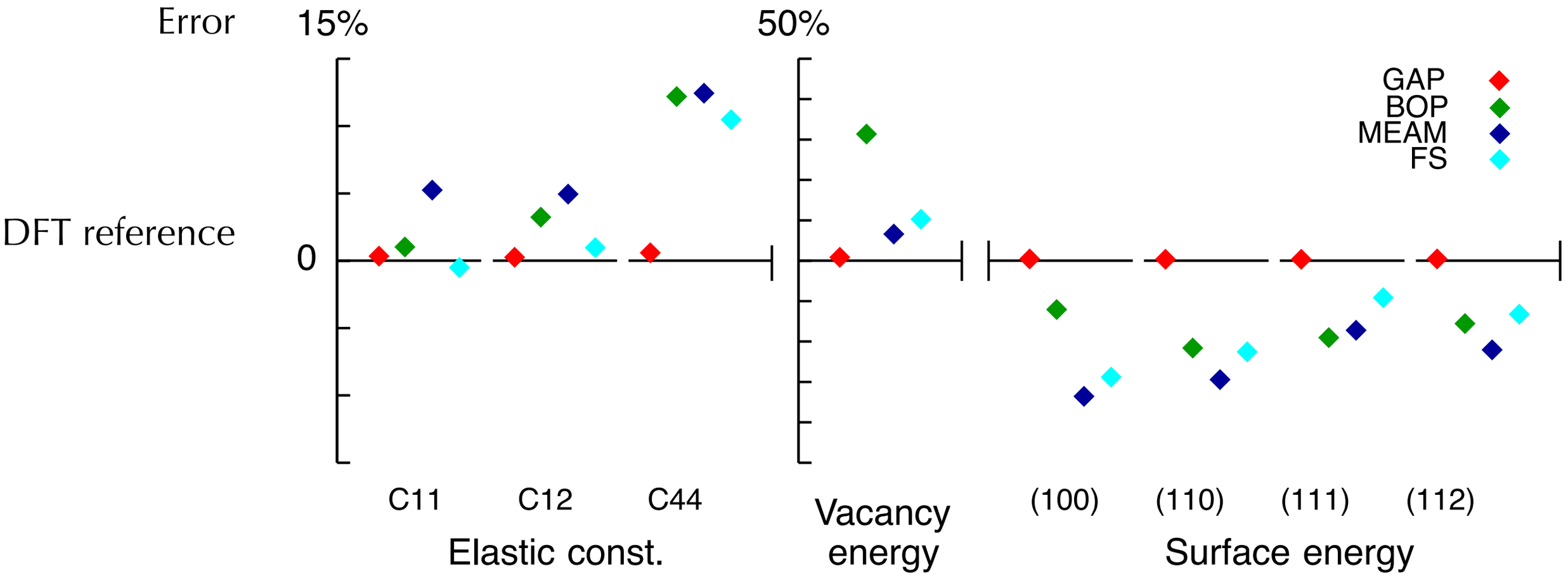
sheared primitive cell → elasticity
large unit cell → phonons
surface unit cells → surface energy
gamma surfaces → screw dislocation
vacancy in small cell → vacancy
vacancy @ gamma surface → vacancy near dislocation

Iterative refinement

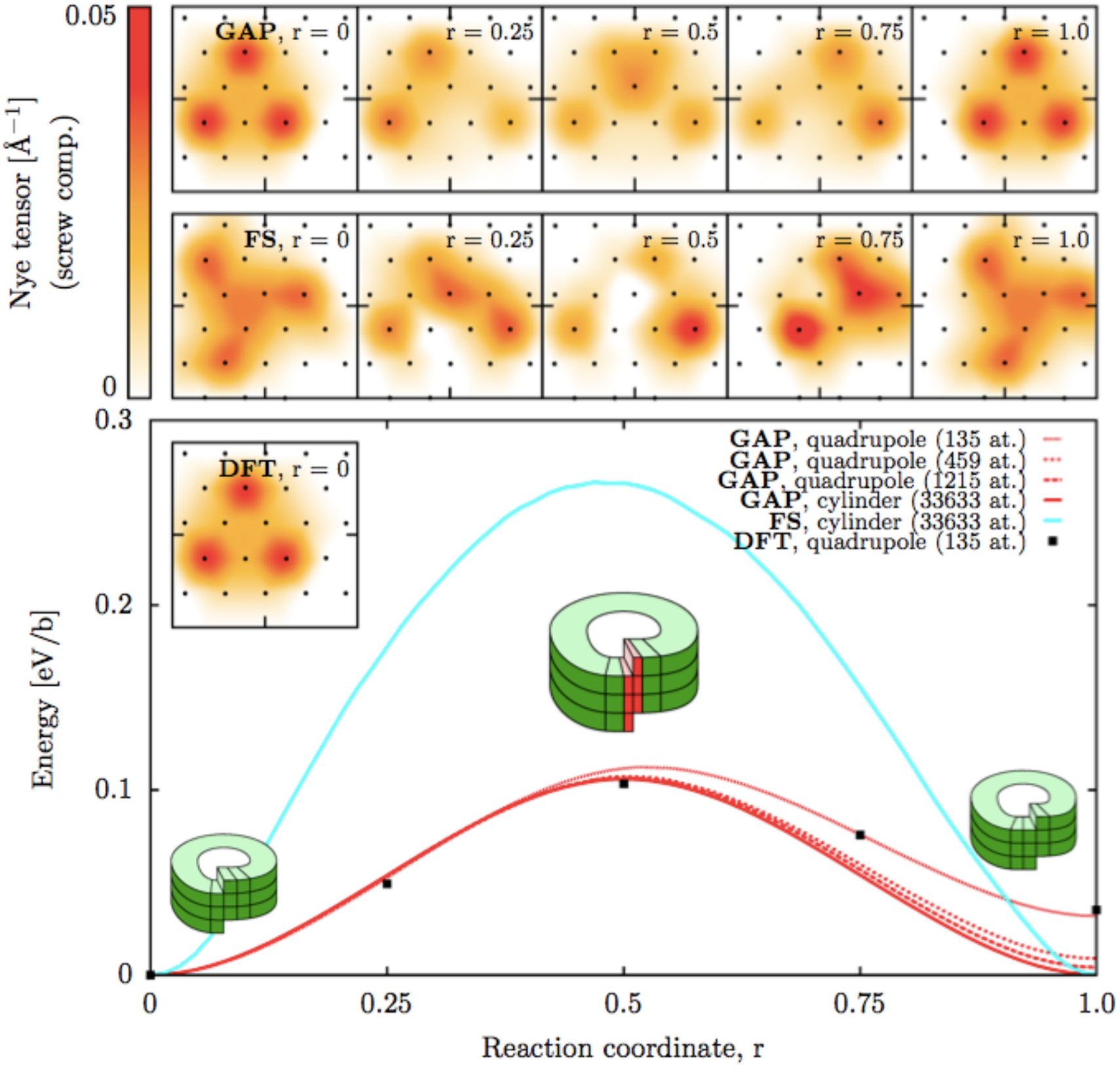
1. QM MD → Initial database
2. Model MD
3. QM → Revised database

What is the acceptable validation protocol?
How far can the domain of validity be extended?

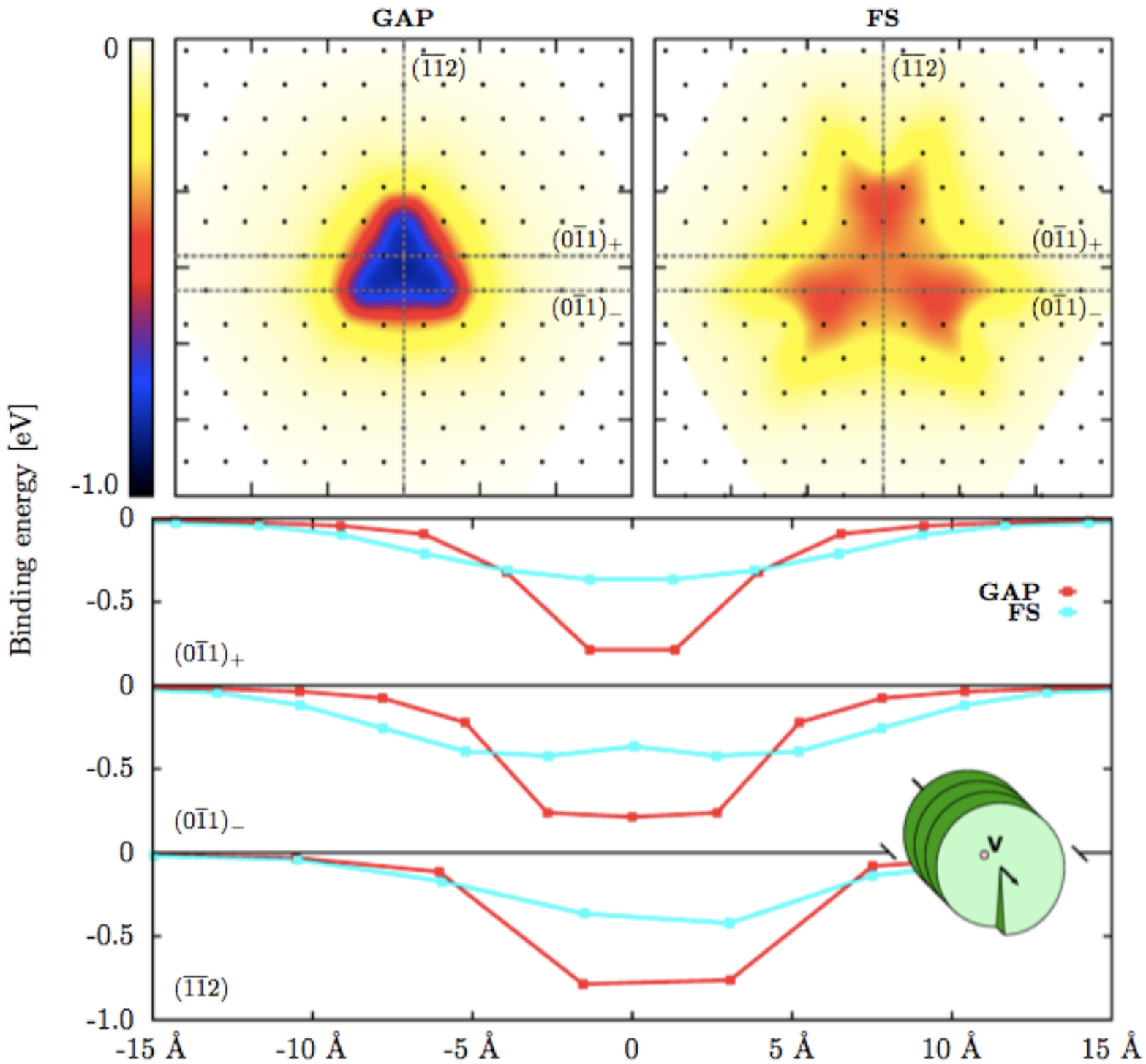
Existing potentials for tungsten



Peierls barrier for screw dislocation glide



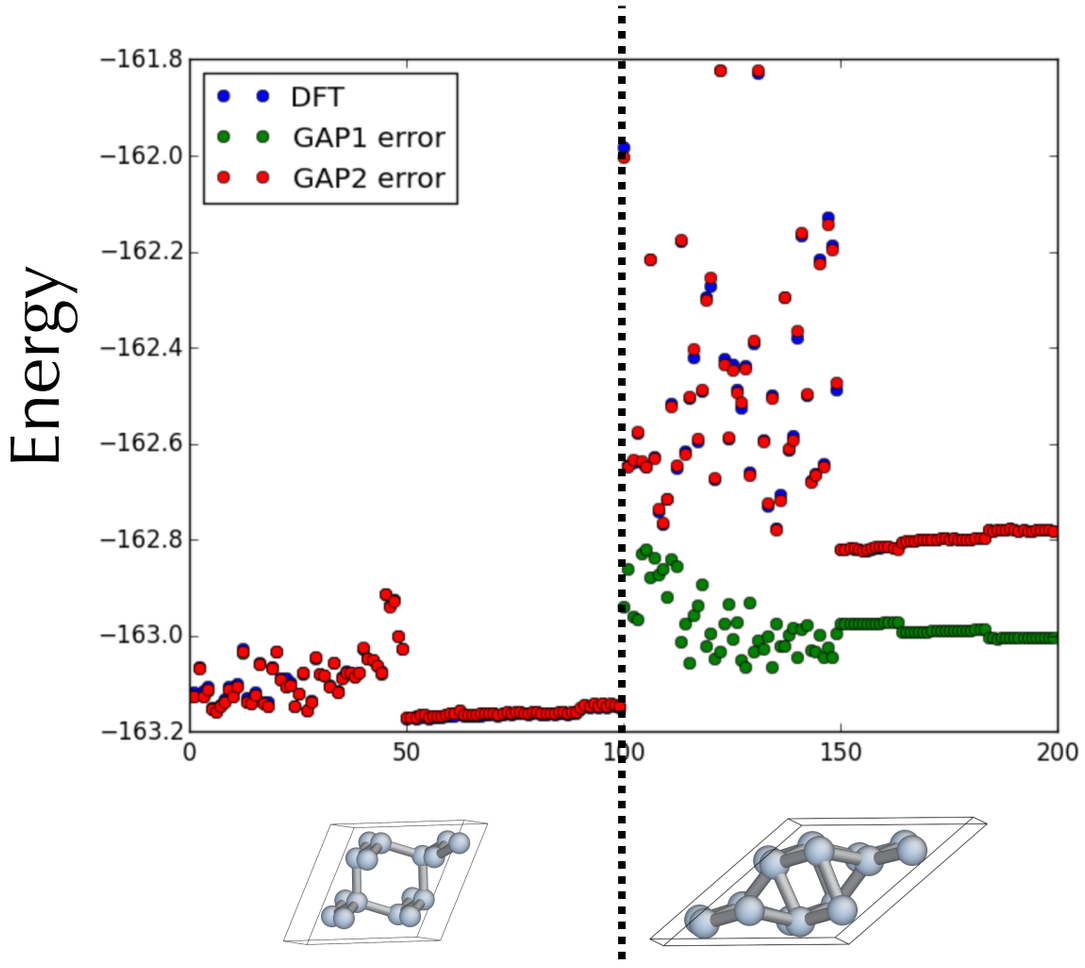
Vacancy-dislocation binding energy

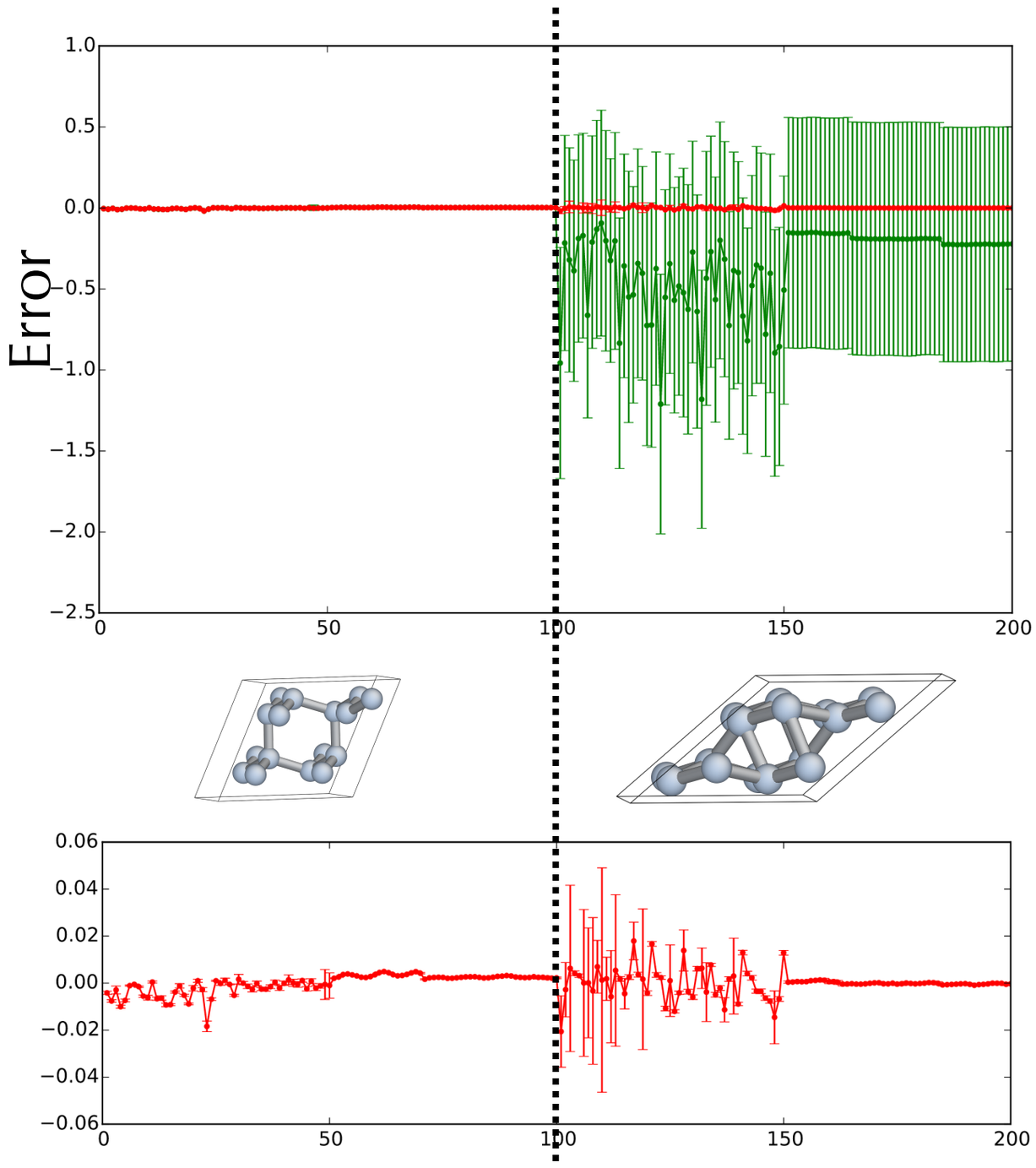


(~100,000 atoms in 3D simulation box)

Self-aware potentials: predicting the error

- Bulk silicon: Diamond and Beta-tin phases
- Predicted error correctly signals where GAP is unreliable





Outstanding problems

- Accuracy on database ⟶ accuracy in properties?
- Database contents ⟶ region of validity ?
- Alloys - permutational complexity? Chemical variability?
- Systematic treatment of long range effects
- Electronic temperature