DE LA RECHERCHE À L'INDUSTRIE

CEA, DAM, DIF
F91297 Arpajon, France

www.cea.fr

“MACHINE LEARNING” FOR INTERATOMIC POTENTIAL: EXAMPLES AND APPLICATIONS

Jean-Bernard Maillet

COLLABORATORS

Saaketh Desai

Gregoire Ferré

Alexandra Goryaeva

David Guzman

Cosmin Marinica

Sami Siraj-Dine

Gabriel Stoltz

Alejandro Strachan

CONTEXT

Condensed Matter Physics: Atomistic simulations

Molecular Dynamics

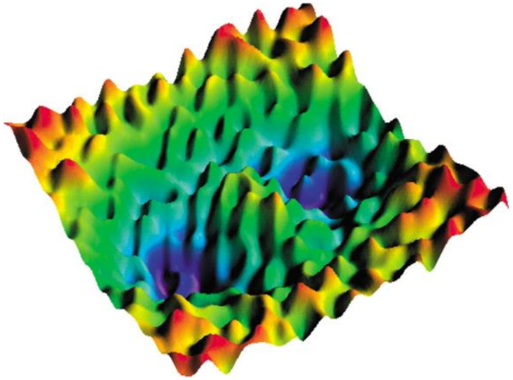
$$\boxed{\langle X \rangle = \lim_{t\to\infty} \frac{1}{t} \int_0^t X(t)dt}$$

Monte Carlo

$$\boxed{\langle X \rangle = \frac{1}{N} \sum_{n=1}^{N} X(n)}$$

Trajectory

$$\bar{F_i} = m_i \gamma_i = -\nabla_i E$$



Sampling

$$P_{Boltzmann} \propto e^{-\beta E}$$

Potential energy surface

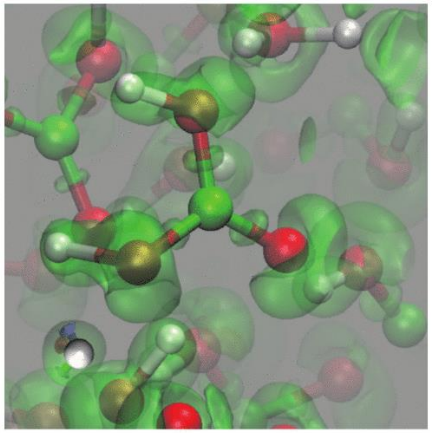
All the physics (and the chemistry) stands in the PES

CONTEXT

ab initio MD

no analytical form for E. The potential is computed at each timestep from quantum chemistry methods.

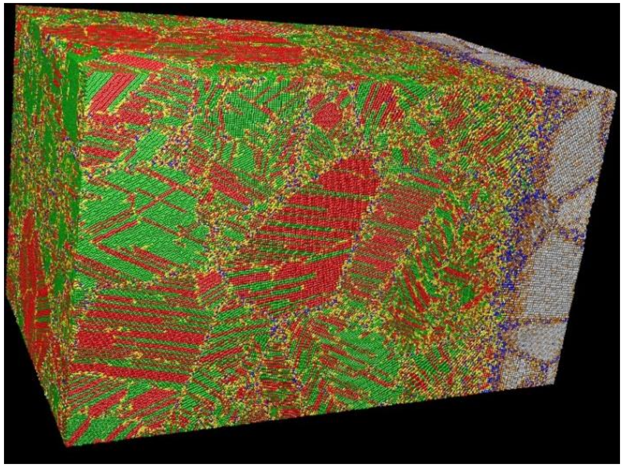
$$H\psi = E\psi$$



Classical MD

we postulate and analytical form of E.

$$E = f(\{\overline{r_i}\})$$



	ab initio MD	Classical MD
Avantages	**Accuracy**	**Scaling (space and time)**
Disadvantages	**Scaling (space and time)**	**Accuracy** **Transferability**

CONTEXT

Classical Potentials

$$E = f(\{\overline{r_i}\}) = V(\varepsilon, \alpha, \beta, \gamma, \delta, \dots)$$

Analytical forms are generally physically based. Example :

Pair potential:

$$V_{LJ} = \frac{1}{2}\sum_i \sum_j 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right]$$

Rare gas

Embedded Atom Method

$$V_{EAM} = \frac{1}{2}\sum_i \sum_j U(r_{ij}) + \sum_i F_i\left(\sum_j \rho(r_{ij})\right)$$

Metals

Bond Order Potential

$$V_{BOP} = \sum_j f_c(r_{ij})\left[V_R(r_{ij}) + \bar{B}_{ij} V_A(r_{ij})\right] + \sum_j V_{vdW}$$

$$B_{ij} = \left(1 + G\sum_k f_c(r_{ik}) e^{-m(r_{ij} - r_{ik})}\right)^{-\frac{1}{2}}$$

Chemistry

ReaxFF

$$V_{reax} = E_{bond} + E_{vdW} + E_{coulomb} + E_{val} + E_{tors} + E_{over} + E_{under} + E_{pen} + E_{conj}$$

All !

CONTEXT

Fitting Potentials

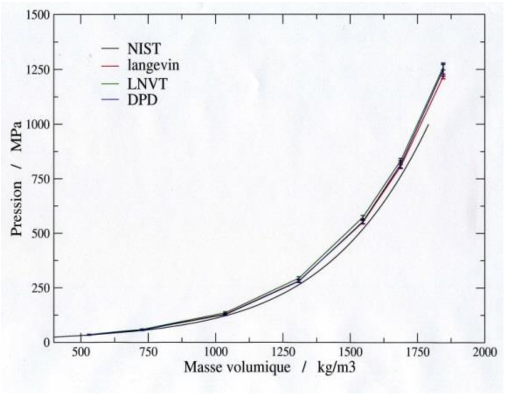
Fitting a potential reduces to minimizing a cost function relatively to a database.

The cost function defines the difference between the reference value and the computed value (MSE) :

$$C = \sum_i \left(E_i^{ref} - V_i\right)^2 \quad \Longleftrightarrow \quad \textbf{accuracy}$$

Ex: fitting a LJ potential on the EOS of argon

$$V_{LJ} = \frac{1}{2}\sum_i \sum_j 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right]$$



CONTEXT

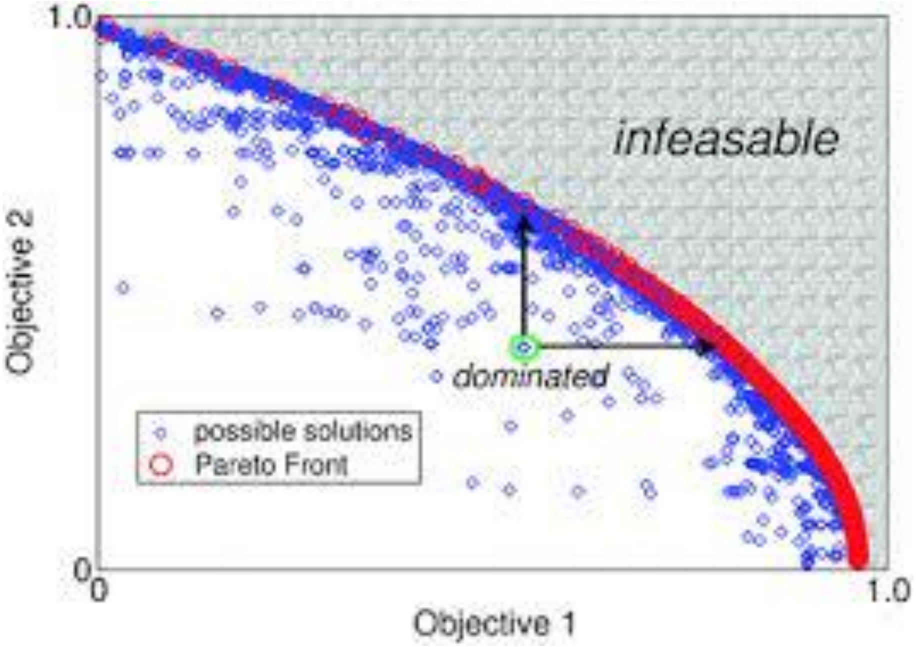
Multi objective functions and the Pareto surface

Ex: Fitting Reax parameters (n>600)

Multi objective function (noncommensurate quantities)

$$C = w_1 MSE(d_{C-C}) + w_2 MSE\left(\Delta H_f\right) + w_3 MSE\left(\overrightarrow{F_i}\right) + \cdots$$

The $\{w_i\}$ are arbitrary user-defined parameters. → Numerous "best parameters" sets



The Pareto surface: impossible to improve one objective function without making another one worse.

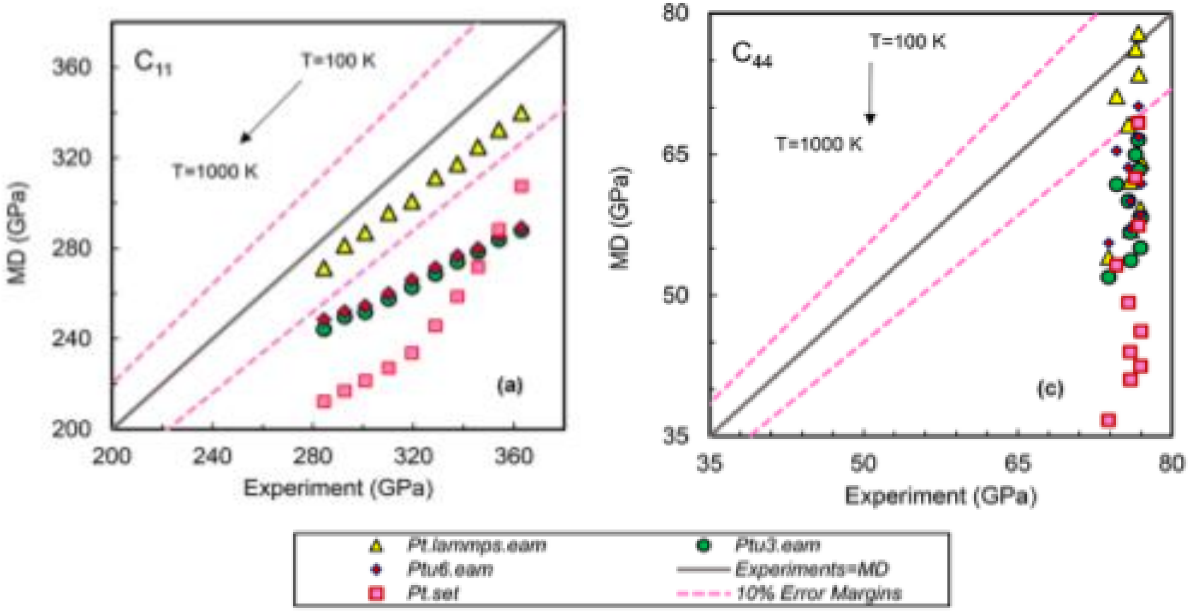
J.P. Larentzos, B.M. Rice, E.F.C. Byrd, N.S. Weingarten and J.V. Lill, JCTC, **11**, 381 (2015)

CONTEXT

Evaluation of transferability: measure of robustness at conditions other than those used in the fitting process

Example: EAM potential for Pt, from the NIST (National Institute of Standards and Technology) Interatomic Potentials Repository

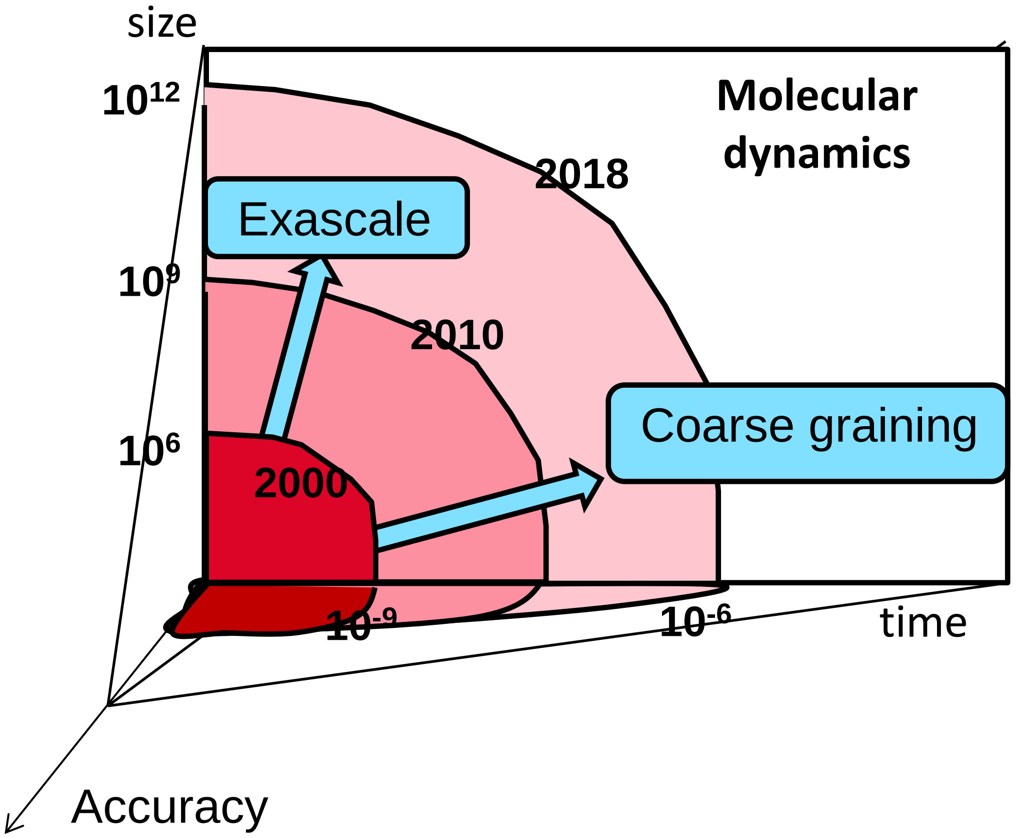
Evaluation of elastic constant at different temperatures



S.M. Rassoulinejad-Mousavi and Y. Zhang, Nature, **8**, 2424 (2018)

CONTEXT

Of course, size and time do matter, but the problem stands in the confidence of atomistic predictions

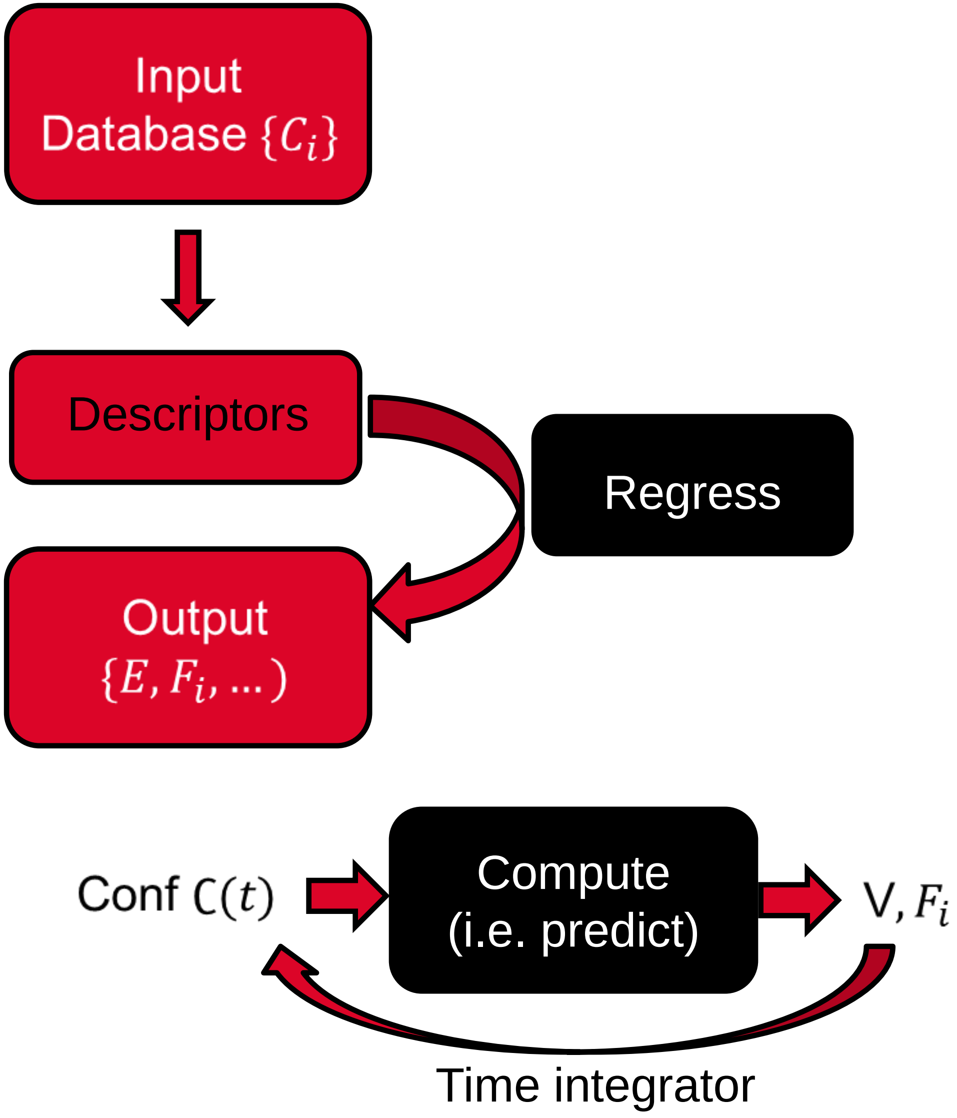


How to increase the accuracy and transferability of interaction potential ? ⇒ **Machine Learning Potential**

OUTLINE

LEARNING

COMPUTING



Input Database $\{C_i\}$

Descriptors

Regress

Output $\{E, F_i, \ldots)$

Conf $C(t)$

Compute (i.e. predict)

V, F_i

Time integrator

OUTLINE

Database

- Elements, transferability
- Representation (interpolation vs extrapolation)
- Sparsification

DATABASE

Elements: C configurations of N atoms

- ○ Total energy
- ○ Forces on each atom
- ○ Total virial

⟹

- ○ 1 data
- ○ 3 N data
- ○ 6 data

/ configuration

The database control the predictive capacities of the potential: one potential only knows what it has learnt !

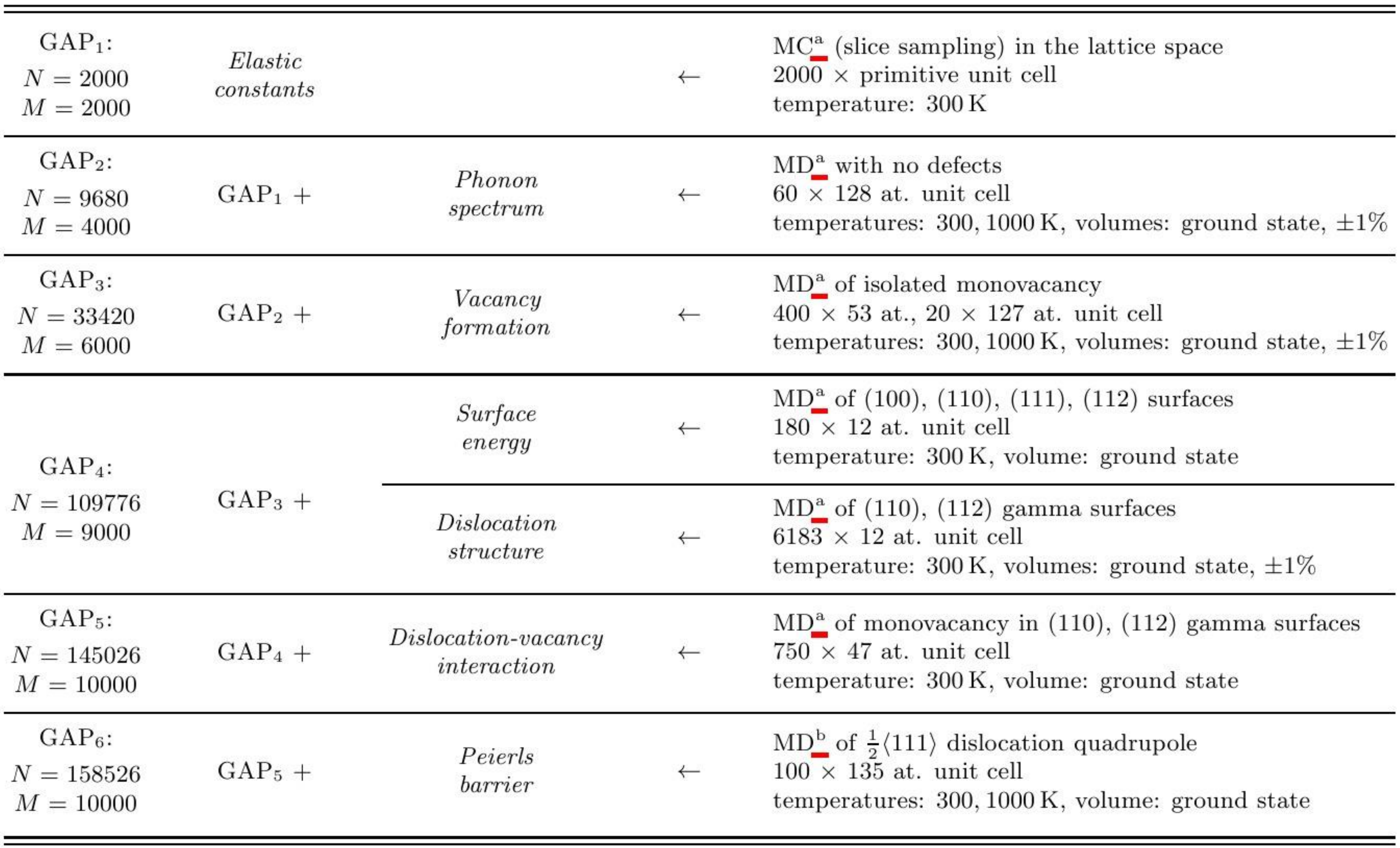
(very) large systems are envisioned but reference data (ab initio) are available only for small systems

Reference data should be representative

- Unit cell deformation→ EOS and elasticity
- Supercell → phonons
- Surface unit cell → surface energy
- γ-surface → screw dislocation
- …

DATABASE

Example: database for Tungsten

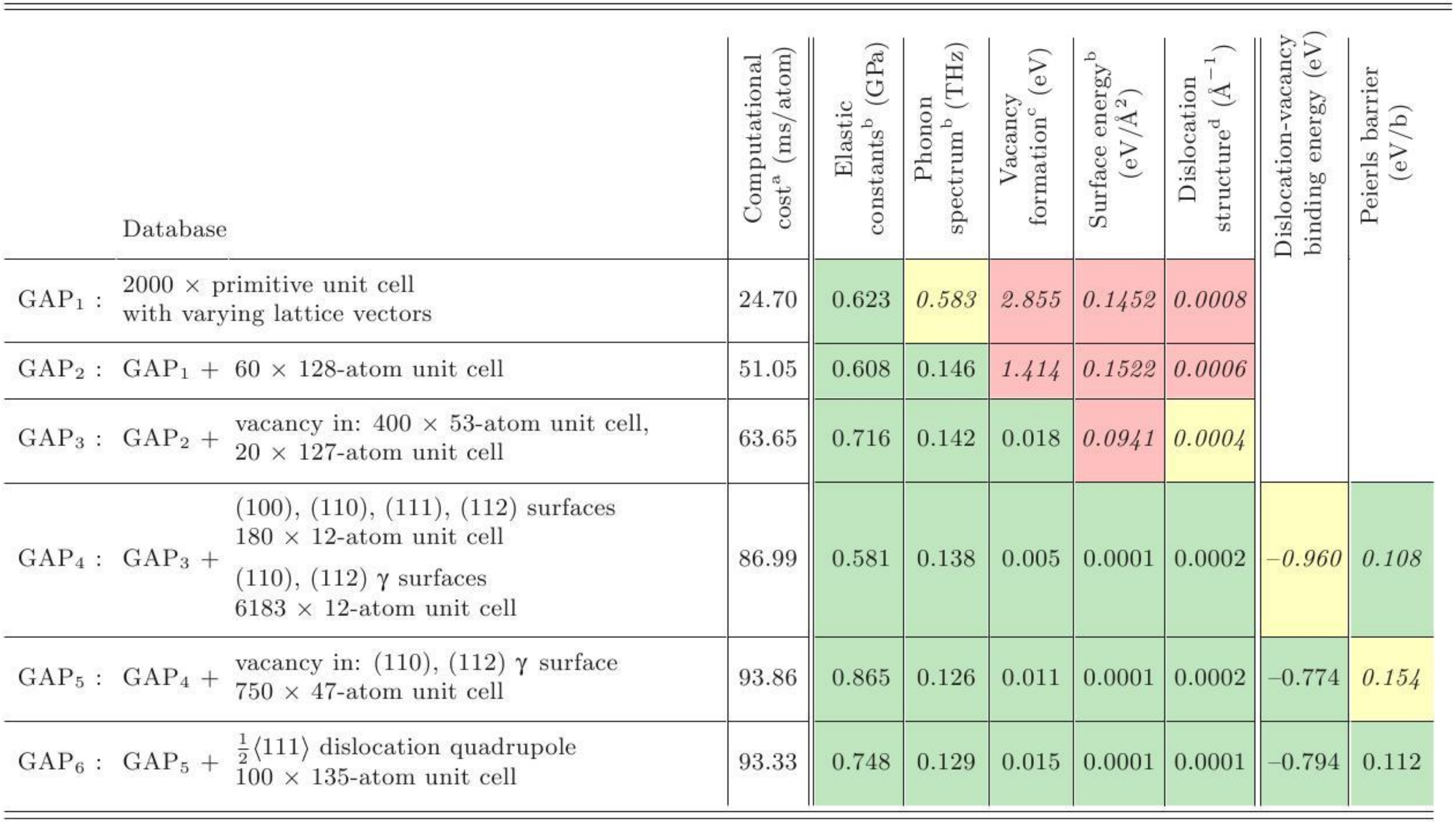


GAP$_1$: $N = 2000$ $M = 2000$	*Elastic constants*		←	MC[a] (slice sampling) in the lattice space 2000 × primitive unit cell temperature: 300 K
GAP$_2$: $N = 9680$ $M = 4000$	GAP$_1$ +	*Phonon spectrum*	←	MD[a] with no defects 60 × 128 at. unit cell temperatures: 300, 1000 K, volumes: ground state, ±1%
GAP$_3$: $N = 33420$ $M = 6000$	GAP$_2$ +	*Vacancy formation*	←	MD[a] of isolated monovacancy 400 × 53 at., 20 × 127 at. unit cell temperatures: 300, 1000 K, volumes: ground state, ±1%
GAP$_4$: $N = 109776$ $M = 9000$	GAP$_3$ +	*Surface energy*	←	MD[a] of (100), (110), (111), (112) surfaces 180 × 12 at. unit cell temperature: 300 K, volume: ground state
		Dislocation structure	←	MD[a] of (110), (112) gamma surfaces 6183 × 12 at. unit cell temperature: 300 K, volumes: ground state, ±1%
GAP$_5$: $N = 145026$ $M = 10000$	GAP$_4$ +	*Dislocation-vacancy interaction*	←	MD[a] of monovacancy in (110), (112) gamma surfaces 750 × 47 at. unit cell temperature: 300 K, volume: ground state
GAP$_6$: $N = 158526$ $M = 10000$	GAP$_5$ +	*Peierls barrier*	←	MD[b] of $\frac{1}{2}\langle 111 \rangle$ dislocation quadrupole 100 × 135 at. unit cell temperatures: 300, 1000 K, volume: ground state

W.J. Szlachta, A.P. Bartok and G. Csanyi, Phys. Rev. B., **90**, 104108, 2014.

DATABASE

Example : GAP potential for Tungsten



	Database	Computational cost[a] (ms/atom)	Elastic constants[b] (GPa)	Phonon spectrum[b] (THz)	Vacancy formation[c] (eV)	Surface energy[b] (eV/Å^2)	Dislocation structure[d] (Å^{-1})	Dislocation-vacancy binding energy (eV)	Peierls barrier (eV/b)
GAP_1 :	2000 × primitive unit cell with varying lattice vectors	24.70	0.623	*0.583*	*2.855*	*0.1452*	*0.0008*		
GAP_2 :	GAP_1 + 60 × 128-atom unit cell	51.05	0.608	0.146	*1.414*	*0.1522*	*0.0006*		
GAP_3 :	GAP_2 + vacancy in: 400 × 53-atom unit cell, 20 × 127-atom unit cell	63.65	0.716	0.142	0.018	*0.0941*	*0.0004*		
GAP_4 :	GAP_3 + (100), (110), (111), (112) surfaces 180 × 12-atom unit cell (110), (112) γ surfaces 6183 × 12-atom unit cell	86.99	0.581	0.138	0.005	0.0001	0.0002	*−0.960*	*0.108*
GAP_5 :	GAP_4 + vacancy in: (110), (112) γ surface 750 × 47-atom unit cell	93.86	0.865	0.126	0.011	0.0001	0.0002	−0.774	*0.154*
GAP_6 :	GAP_5 + $\frac{1}{2}\langle 111\rangle$ dislocation quadrupole 100 × 135-atom unit cell	93.33	0.748	0.129	0.015	0.0001	0.0001	−0.794	0.112

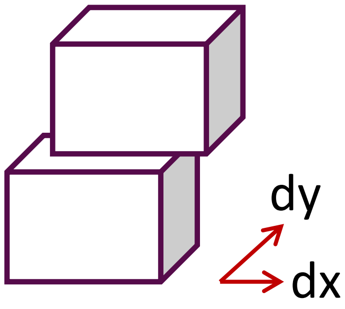
DATABASE

Numerical potential for the bcc-hcp transition in Iron

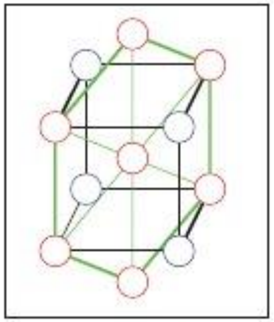
Database

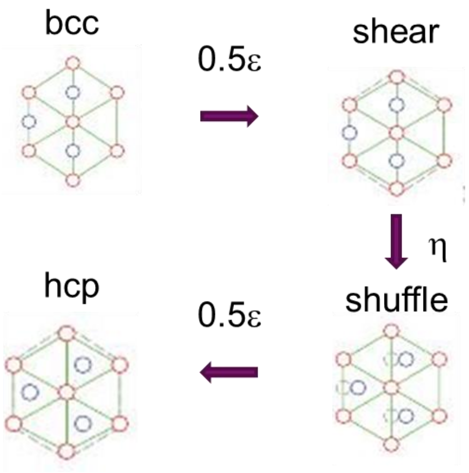
- BCC
  - ✓ 100 confs MC / 1000
  - ✓ 24 confs deformations / 42
  - ✓ 60 confs γ-surface/100
- HCP
  - ✓ 20 confs cold curve / 20
  - ✓ 100 confs MC / 1000
  - ✓ 24 confs deformations / 42
  - ✓ 60 confs γ-surface/100
- BCC-HCP
  - ✓ 20 confs in ε
  - ✓ 20 confs in η

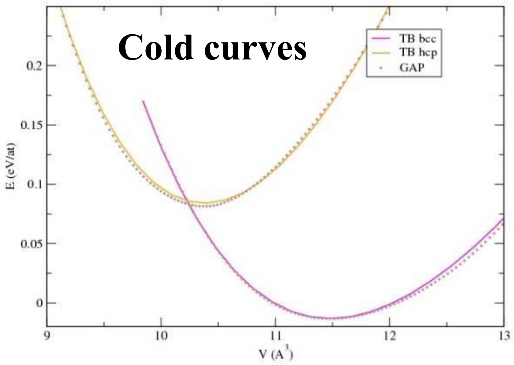
γ-surface



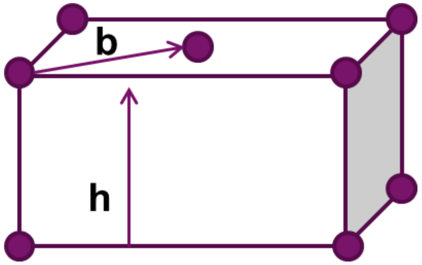
bcc-hcp transition







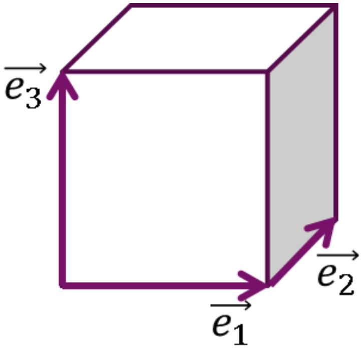
Large deformation path along dense planes



DATABASE

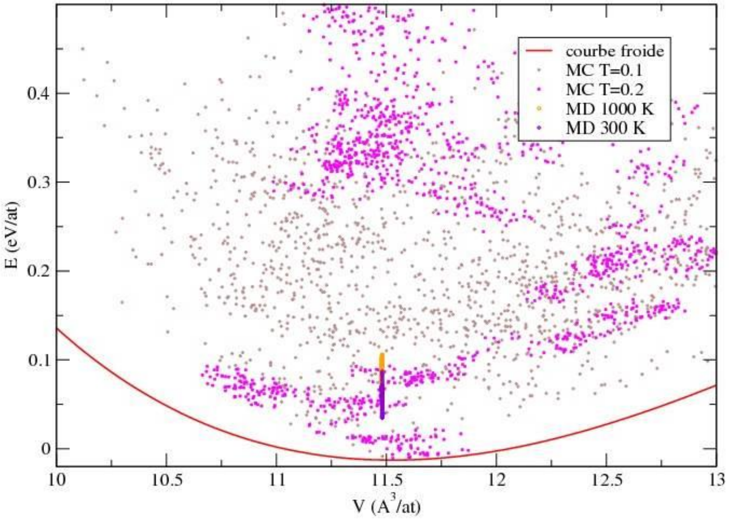
Transferability: predictive capacity within the envelope defined by the database (interpolation)

Monte Carlo in parameter space (a,b,c,α,β,γ)



Random deformation of the unitary cell with acceptance probability :

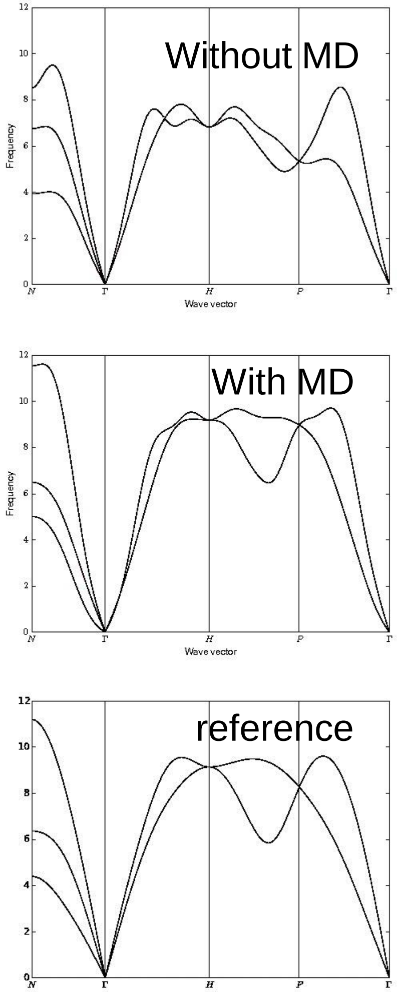
$$e^{-\frac{\Delta E}{kT}}$$



10 % ⇨ learning

90 % ⇨ testing

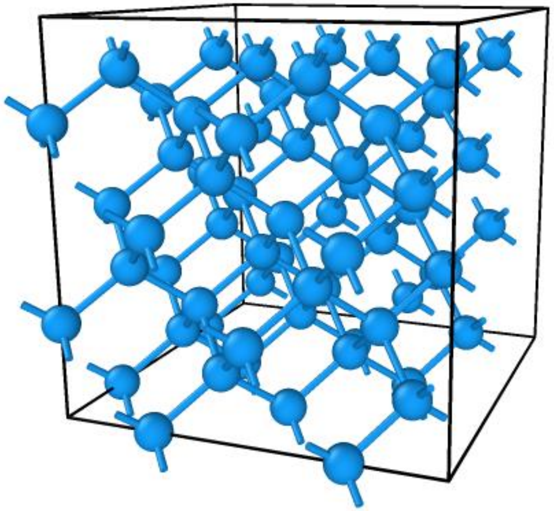
Phonon spectrum



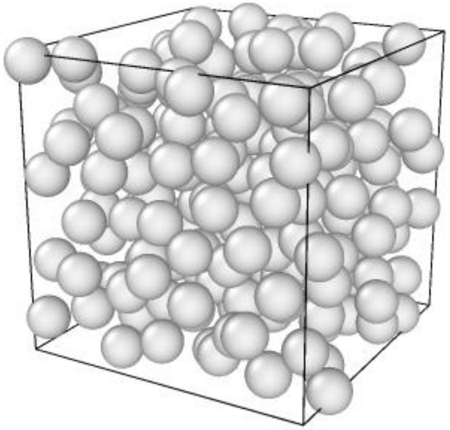
Extended database ⟺ Increased transferability

DATABASE

Ge database



Diamond cubic phase
Volumetric deformation
Shear deformation
Normal deformation



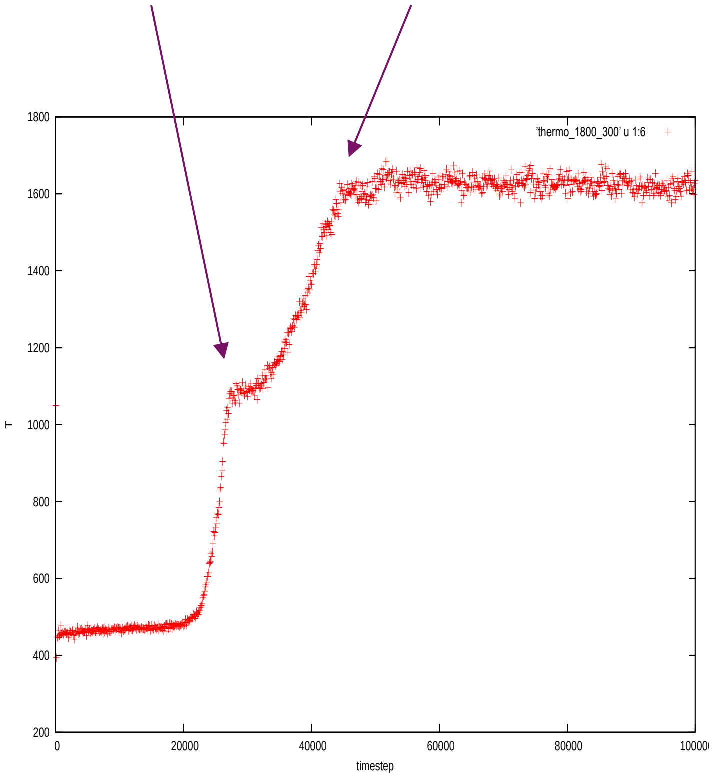
Liquid phase
Ab initio simulations @ 1500 K, 3000 K

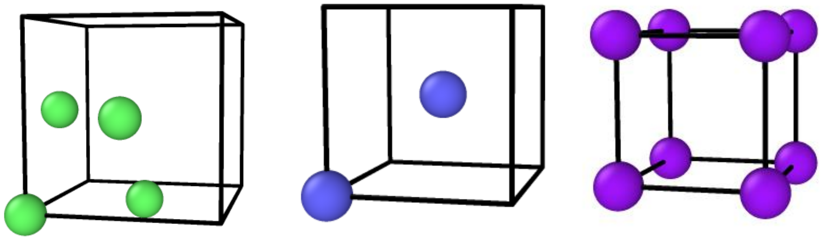
Get a very good potential for diamond and liquid, but not for the melting temperature.

DATABASE

Melting of Ge diamond

Melting of Ge hexagonal





Metastable phases

Volumetric deformation for FCC, BCC, Simple cubic and A5 phases

Hot crystal

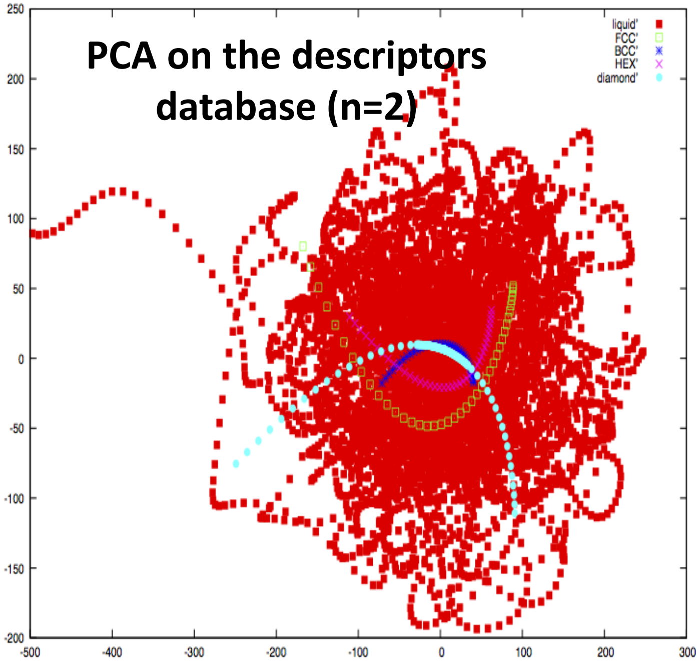
Ab initio simulations @ 800 K, 1000 K

Better agreement on the melting temperature

DATABASE

Representation: database for Ge

Each cartesian configurations is transformed in a feature space where atomic environment are represented by descriptors (here 55 Bispectrum components)



PCA on the descriptors database (n=2)

Results are less separable.

Non parametric regression method are usually successful in interpolation, but can be awful in extrapolation. The representation of the database could serve as a first diagnostic to prevent extrapolation.

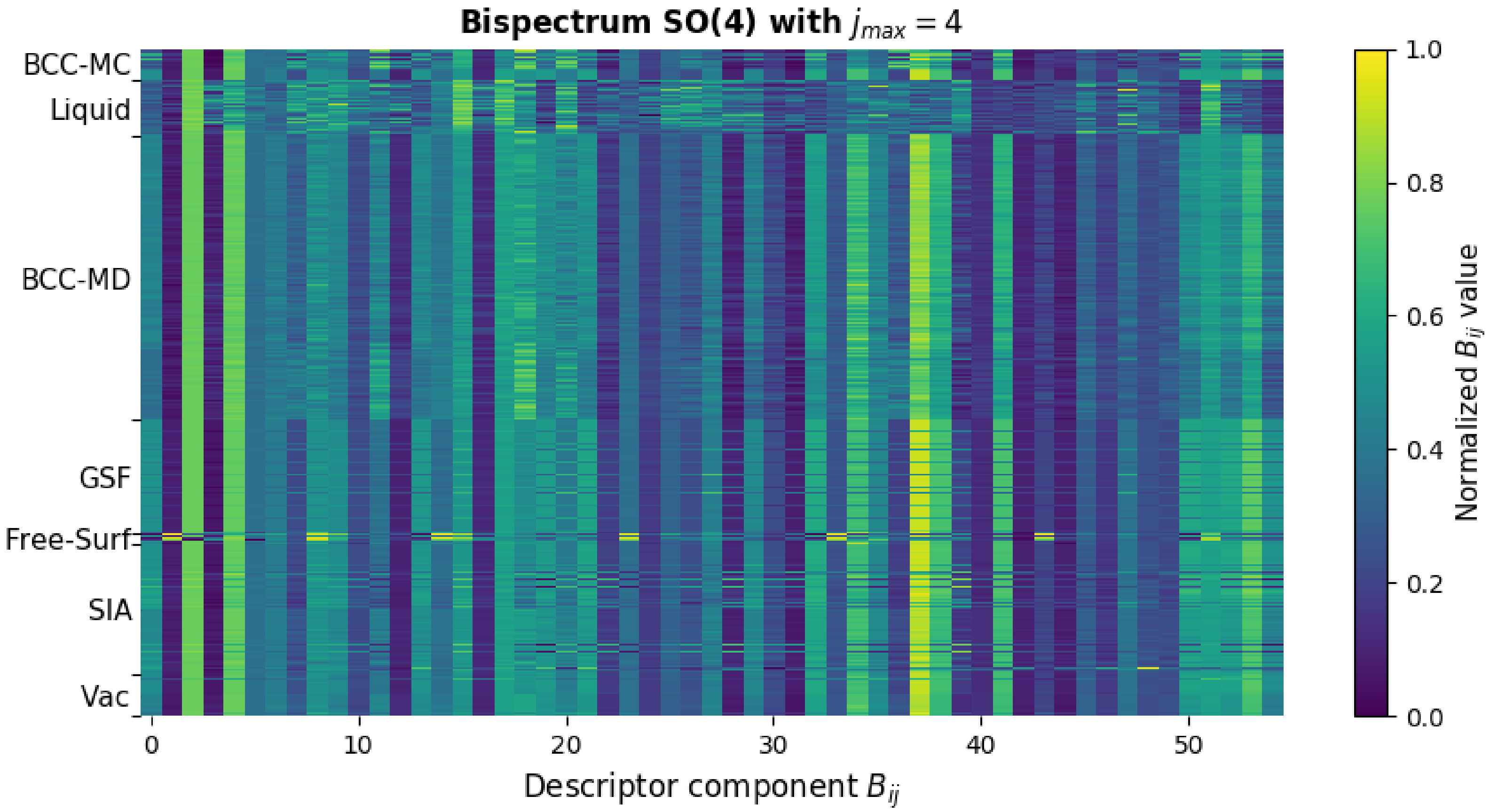
Representation of the database

↕

Interpolation vs extrapolation

DATABASE

Representation: database for Fe



➡ Sparse matrix

DE LA RECHERCHE À L'INDUSTRIE

DATABASE

Sparsification: reduce the size of the database ($D \times N$ matrix)

Random sparsification : randomly delete configurations from the database until the desired size is achieved.

SVD: compute the entropy of the matrix.
Normalized spectrum (for k non-zero eigenvalues):

$$V_j = \frac{u_j^2}{\sum_{i=1}^{k} u_i^2}$$

Compute the entropy of the matrix as:

$$S(A) = -\frac{1}{\log k}\sum_{j=1}^{k} V_j \log V_j$$

Then compute the entropy of the matrix A^{-i}
Sort the configuration by entropy, and delete the smaller contributors.

CUR decomposition: decomposition of $A = CUR$, where C is composed of columns of A and R of lines. Use SVD to obtain a statistical weight associated to each column, used as delete probability.

Max variance:

- Start with a very small database $a \in A$
- Compute the distance between each elements of a and $C_i \in A$
- Select the most distant element $C_k$and add it to a

DATABASE

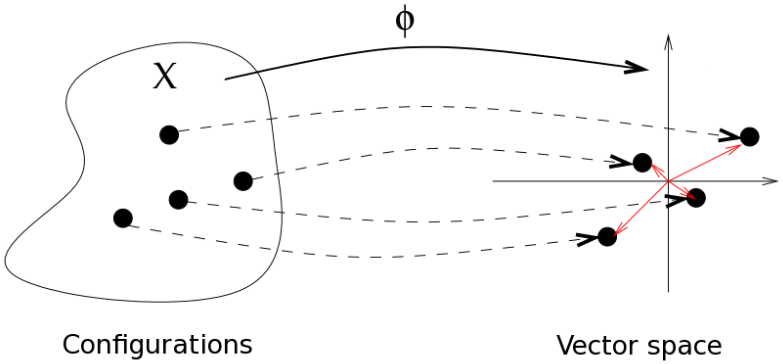
The richer the database the better the potential

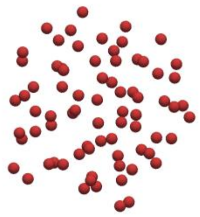
BUT

- Elements should be selected with a “physically” relevant criterion
- Adding distinct elements increases the transferability of the potential but decreases its accuracy
- Representation of the database: identify relevant (separated) domains, control extrapolation.
- Scaling of the potential with the size of the database: optimize the ratio information/size

DESCRIPTORS

Cartesian space





N atoms
(1 environment)

3N Degrees of freedom

One vector $\Phi(C)$

Constant dimension

But physics imposes some invariance on the energy: translation, permutation, rotation

$\Phi(C)$ should have these invariances

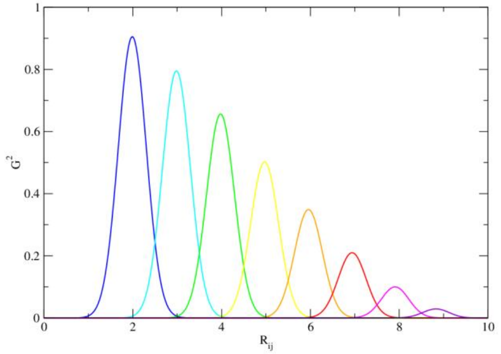
- Symmetry functions
- Matrices
- Spectral expansions
- Graphs
- …

DESCRIPTORS

Example 1 : Symmetry functions

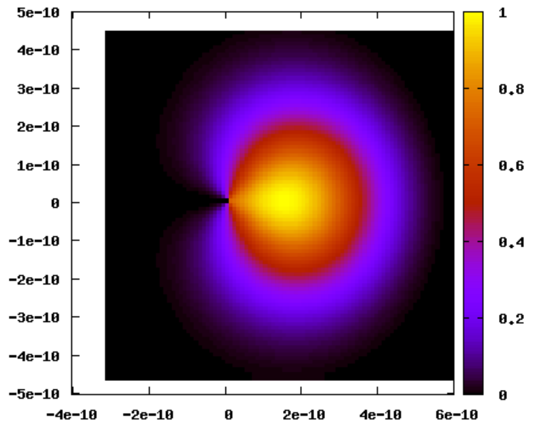
Radial symetry functions

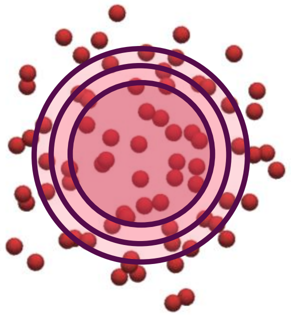
$$G_i^2 = \sum_j e^{-\eta\left(R_{ij}-R_S\right)^2} f_c(R_{ij})$$



Angular symetry functions

$$G_i^4 = 2^{1-\zeta}\sum_{j,k\neq j}\left(1+\lambda\cos\theta_{ijk}\right)^{\zeta} e^{-\eta\left(R_{ij}^2+R_{ik}^2+R_{jk}^2\right)} f_c(R_{ij})f_c(R_{ik})f_c(R_{jk})$$





$$D = \begin{pmatrix} G_1^2 \\ G_2^2 \\ \dots \\ G_n^4 \end{pmatrix}$$

Comparison of configurations ⟷ Comparison of descriptors vectors

Problem: ambuiguity of representations

J. Behler, , J. Chem. Phys., **134**, 074106, 2011.

DESCRIPTORS

Example 2 : Gram matrix, coulomb matrix

Weyl theorem (1939) → a configuration can be represented by the scalar products of its radial vectors.

Gram (or Weyl) matrix (symmetric) :

$$G = \begin{pmatrix} r_1 r_1 & \cdots & r_1 r_N \\ \vdots & \ddots & \vdots \\ r_N r_1 & \cdots & r_N r_N \end{pmatrix}$$

Over complete descriptor

G is invariant by rotation and reflection, but not by permutation

To retain permutation invariance, one can retain the spectrum of this matrix (but it is then under complete)

M. Rupp, A. Tkatchenko, K-R. Muller and O.A. Lilienfelf, Phys. Rev. Lett., **108**, 058301, 2012.

DESCRIPTORS

The coulomb matrix can be seen as an entrywise non linear transformation of the Weyl matrix

$$M_{ij} = \begin{cases} 0.5 {Z_i}^2 \; if \; i = j \\ \dfrac{Z_i Z_j}{\left\| r_i - r_j \right\|} \end{cases}$$

Similarity measure:

$$D(M, M') = \sqrt{\sum_i |\varepsilon_i - {\varepsilon'}_i|^2}$$

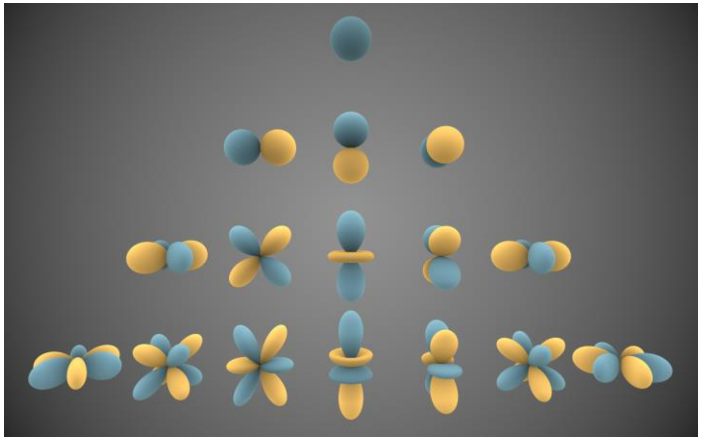
ε_i are the ordered eigenvalues of M.

D has the correct invariances but remains under complete, and the ordering of its eigenvalues hinders regularity (problem for derivative)

DESCRIPTORS

Example 3 : Spherical power spectrum

Expansion of the density function on a basis of spherical harmonics and radial functions.



$$\rho_i(r) = \sum_j \sum_{l,m} c_{nlm}^{(i)j} g_n(r) Y_{lm}(\hat{r})$$

It can be shown that the product $p_{nn'l} = c_{nlm}^{\dagger} c_{n'lm}$ is rotational invariant

spherical power spectrum

$$p_{nn'l} = \sqrt{\frac{8\pi^2}{2l+1}} \sum_m c_{nlm}^{\dagger} c_{n'lm}$$

A.P. Bartók, M.C. Payne, and G. Csányi, Phys. Rev. B., **87**, 184115, 2013.

DESCRIPTORS

Example 4 : SO4 bispectrum

Density of neighbor atoms at location r relative to a central atom i

ω_j are dimensionless weight chosen to distinguish atoms of different types

$$\rho_i(r) = \delta(r) + \sum_{r_j < R_{cut}} f_c(r_j)\omega_j \delta(r - r_j)$$

The radial coordinate r is mapped on to a third angular coordinate $\theta_0 = \theta_0^{max} r / R_{cut}$. Each neighbor position (r, θ, ϕ) is mapped to (θ_0, ϕ, θ), a point on the unit 3-sphere.

The natural basis for functions on the 3-sphere is formed by the 4D hyperspherical harmonics $U_{m,m'}$ (θ_0, θ, ϕ), defined for j = 0, 1/2,1,... and m,m′ = −j,−j+1,...,j−1,j .

The density function defined on the 3-sphere can then be expanded using 4D hyperspherical harmonics:

$$\rho(r) = \sum_{j=0}^{\infty} \sum_{m=-j}^{j} \sum_{m'=-j}^{j} u^j_{m,m'} U^j_{m,m'}(\theta_0, \phi, \theta)$$

A.P. Bartók, M.C. Payne, and G. Csányi, Phys. Rev. B., **87**, 184115, 2013.

DESCRIPTORS

Because the neighbor density is a weighted sum of δ-functions, each expansion coefficient is a sum over discrete values of the corresponding basis function evaluated at each neighbor position :

$$u^{j}_{m,m'} = U^{j}_{m,m'}(0) + \sum_{r_j < R_{cut}} f_c(r_j)\omega_j U^{j}_{m,m'}(\theta_0, \phi, \theta)$$

The bispectrum components are formed as the scalar triple products of the expansion coefficients (to insure rotational invariances)

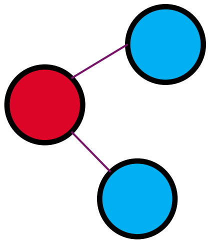
$$B_{j_1,j_2,j} = \sum_{m,m'} {u^{j}_{m,m'}}^{*} \sum_{\substack{m_1,m_1' \\ m_2,m_2'}} H^{\substack{j,m,m' \\ j,m_1,m_1'}}_{j,m_2,m_2'} u^{j_1}_{m_1,m_1'} u^{j_2}_{m_2,m_2'}$$

$H^{\substack{j,m,m' \\ j,m_1,m_1'}}_{j,m_2,m_2'}$: Clebsch-Gordan coupling coefficients for the hyperspherical harmonics

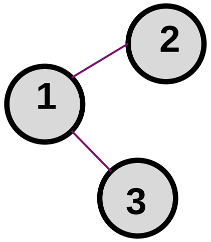
DESCRIPTORS

Example 5 : Graphs

Molecule



Graph



Adjacent matrix

$$\begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{matrix}$$

Graphs are translational and rotational invariant

Laplacian of a weighted graph

$$-w_{ij}$$

$$L = \sum_{j} w_{ij}\ if\ (i = j)$$

$$0\ otherwise$$

The spectrum of the Laplacian (and adjacency) matrix is a graph invariant, that is, it is invariant to permutations in the indices of the vertices.

A weight graph built with $\left\|r_i - r_j\right\|^2$ leads to an adjacency matrix close to Weyl's one. Its spectrum has the required invariances.

G. Ferré, T. Haut, and K. Barros, J. Chem. Phys., **146**, 114107, 2017.

DESCRIPTORS

Questions : accuracy vs CPU: mixing descriptors ?

Example:

$$D = G_2 \oplus B$$

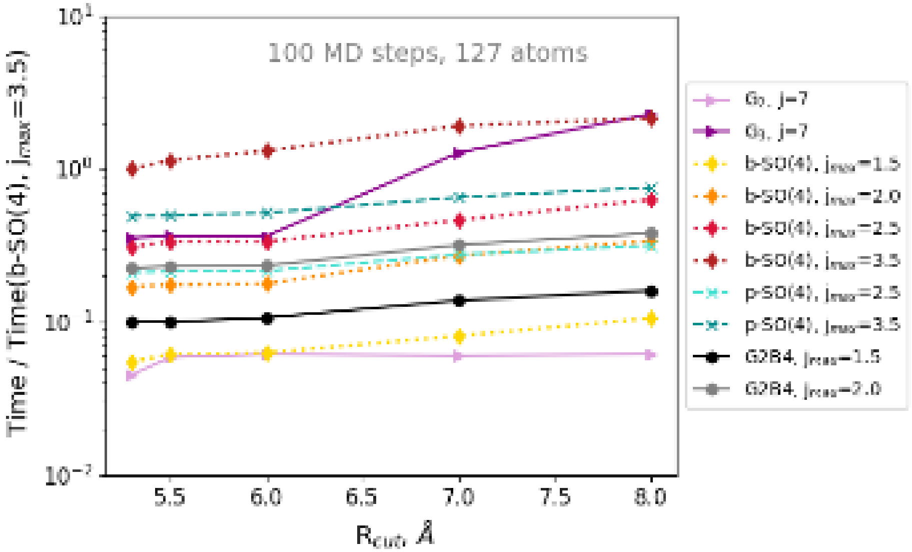


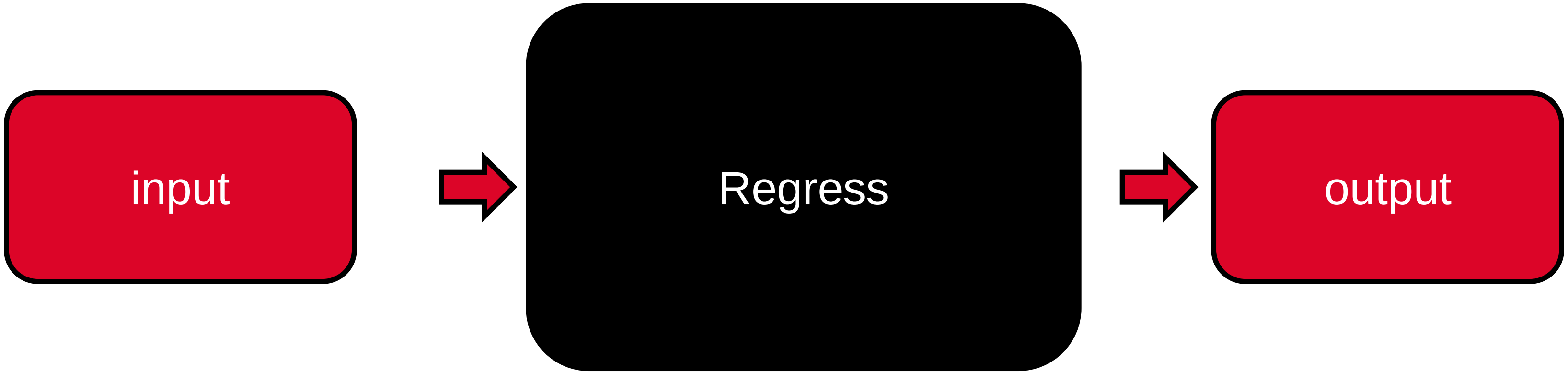
Figure 1: Computational cost of the atomic descriptors (on a single Intel Broadwell core) normalized by that of bispectrum SO(4) with j_{max}=3.5 and R_{cut}=5 as a function of the cutoff distance R_{cut}. The timings do not comprise evaluation of the nearest neighbor environment, *i.e.* it is independent of the employed neighbor list algorithm.

Perspective/questions :

- Coarse grained descriptor ?
- Hybrid descriptor (adding invariant force field data into the descriptor) ?
- ….

A. Goryaeva, C. Marinica, and J-B Maillet Proc for NIMB, 2018

REGRESSION METHOD



input

Regress

output

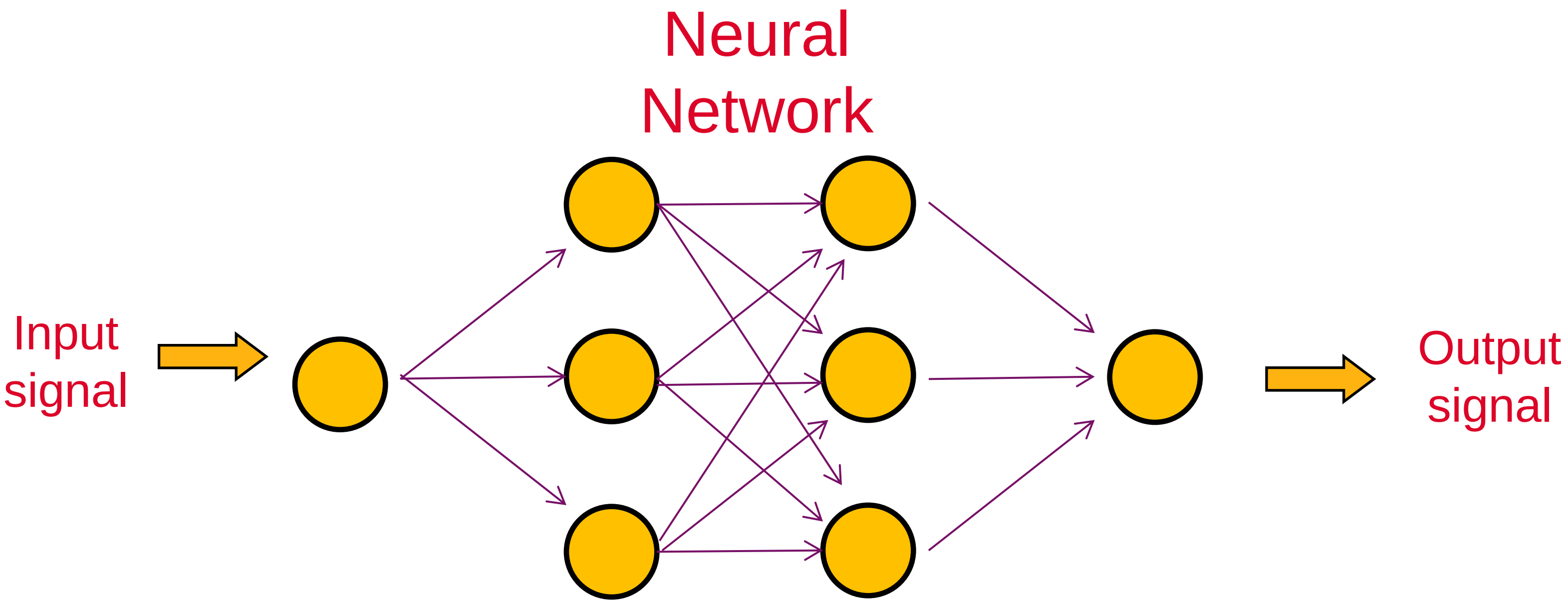
Ensemble of vectors of dimension D

- Neural network
- Kernel method
- linear

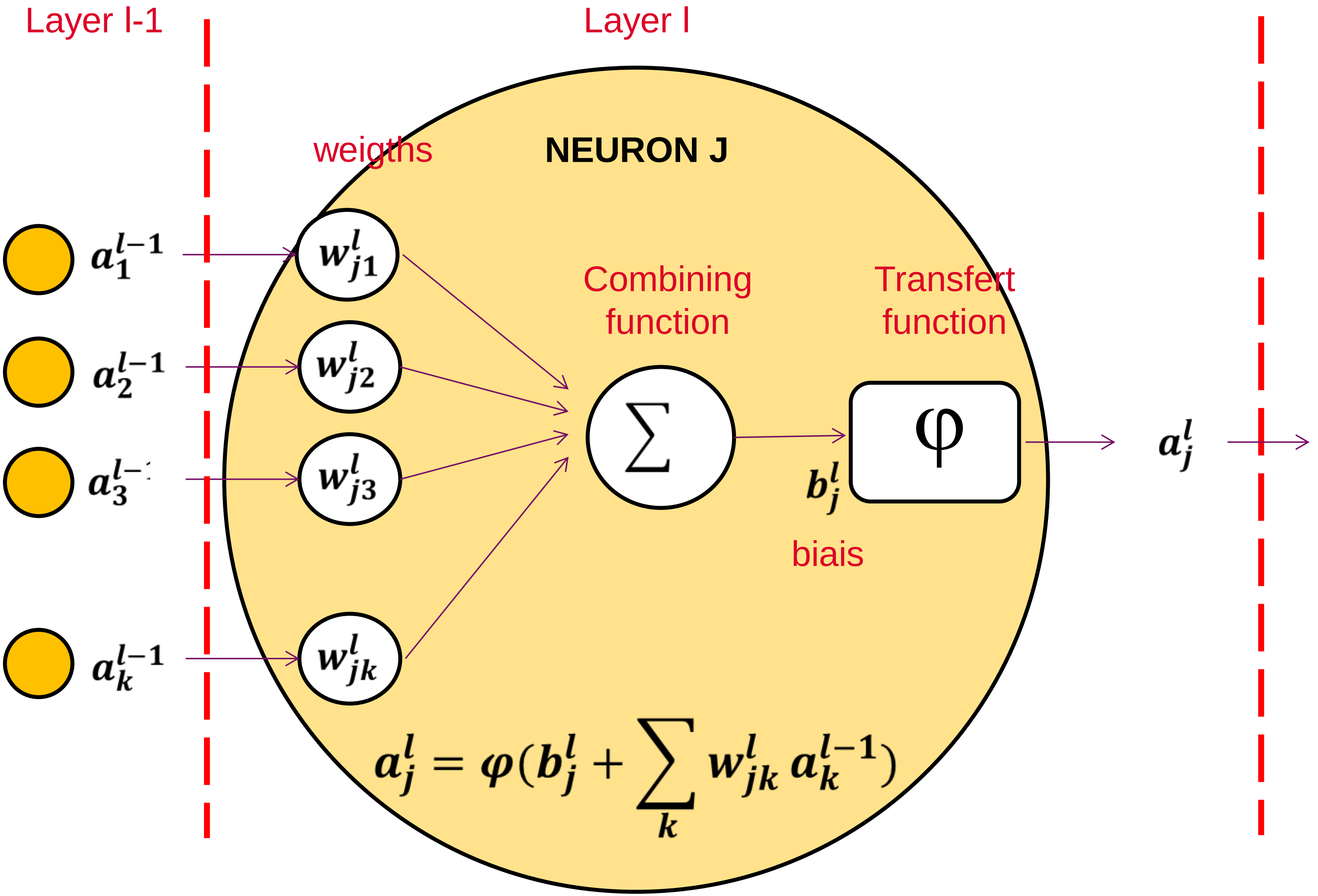
- energy
- forces
- virial

NEURAL NETWORK

Artificial neural networks are massively parallel interconnected networks of simple (usually adaptive) elements and their hierarchical organizations which are intended to interact with the objects of the real world in the same way as biological nervous system do.



NEURAL NETWORK



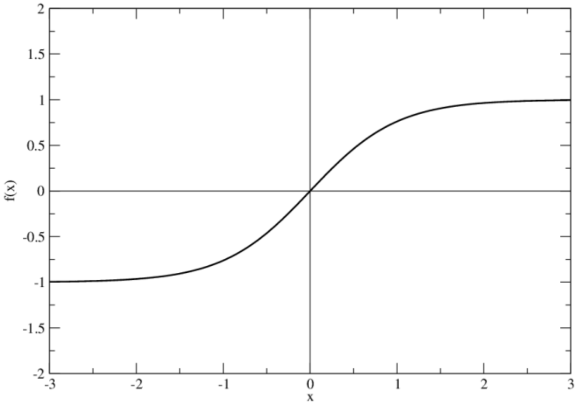
$$a_j^l = \varphi(b_j^l + \sum_k w_{jk}^l \, a_k^{l-1})$$

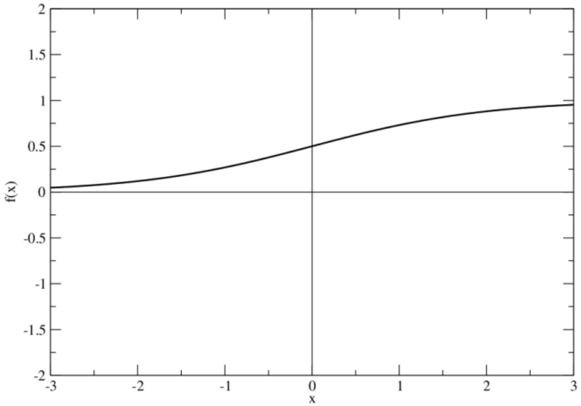
NEURAL NETWORK

Transfert function guarantees the non linear behavior of the NN

Hyperbolic tangent

$$f(x) = \tanh(x)$$



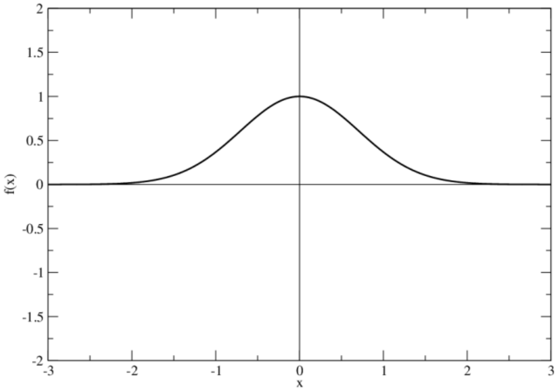


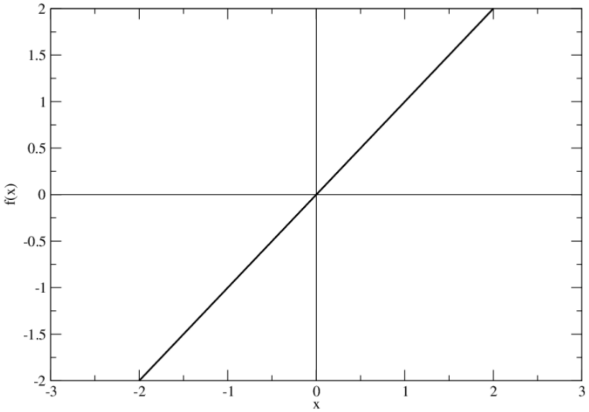
Sigmoid

$$f(x) = \frac{1}{1 + e^{-x}}$$

Gaussian

$$f(x) = e^{-\eta x^2}$$





linear

$$f(x) = x$$

DE LA RECHERCHE À L'INDUSTRIE

NEURAL NETWORK

LEARNING: BACKPROPAGATION

For each input (i.e. each configuration of the training database) a forward pass over the NN leads to the output y. The cost function is generally defined as :

$$C = \frac{1}{2}\sum_i (E - y)^2$$

We want to minimize the cost function relatively to the parameters of the NN, i.e. to compute $\frac{\partial C}{\partial w_{ij}}$

Apply the chain rule:

$$\frac{\partial C}{\partial w_{ij}} = \frac{\partial C}{\partial O_j}\frac{\partial O_j}{\partial I_j}\frac{\partial I_j}{\partial w_{ij}}$$

Output of (previous layer) neuron i

Derivative of activation function

$y - E$

$$\frac{\partial C}{\partial O_j} = \sum_l \left(\frac{\partial C}{\partial O_l}\frac{\partial O_l}{\partial I_l}\frac{\partial I_l}{\partial O_j}\right) = \sum_l \left(\frac{\partial C}{\partial O_l}\frac{\partial O_l}{\partial I_l} w_{jl}\right) \Longleftarrow$$

Recursion relation = backpropagation

NEURAL NETWORK

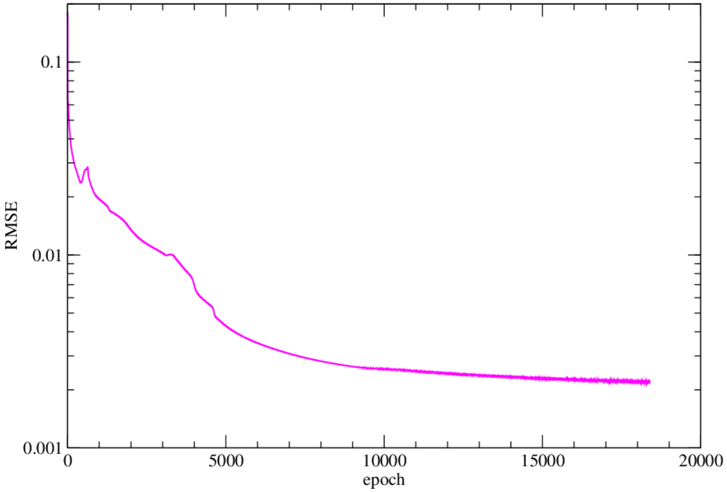
NN: A 1D EXAMPLE

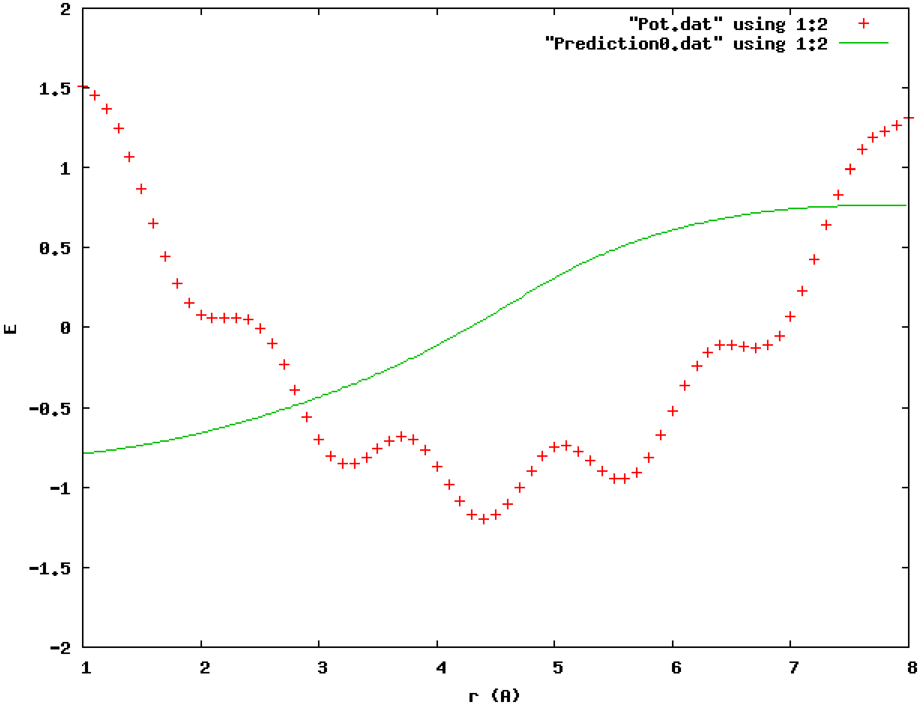
$$E = \frac{\cos(5R) + (R - 4.5)^2}{5} - 1$$

Network: S-25H-25H-L

Learning sample: 70

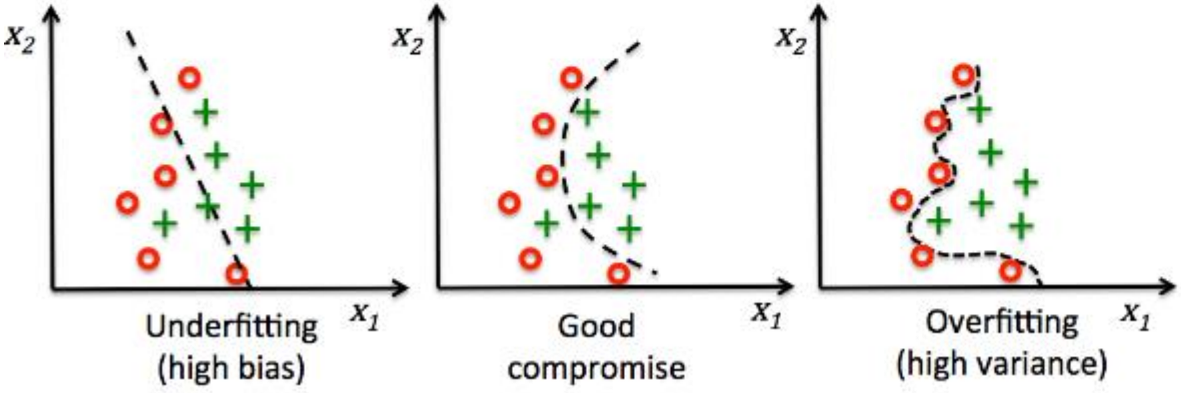
Testing sample: 300

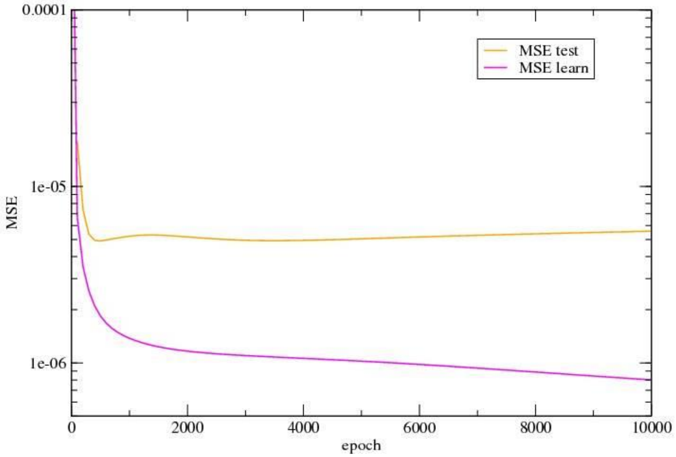


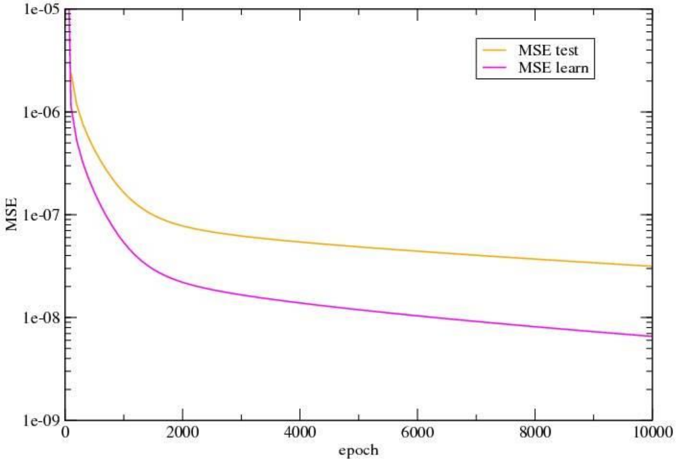


NEURAL NETWORK

Risk: under and overfitting

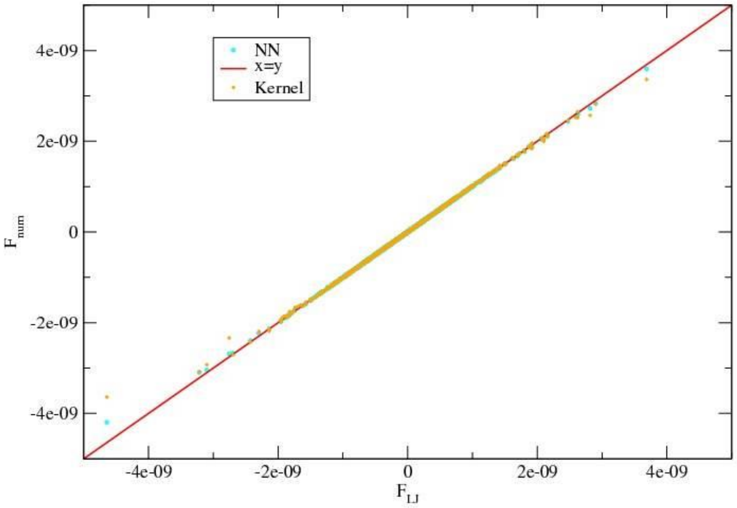




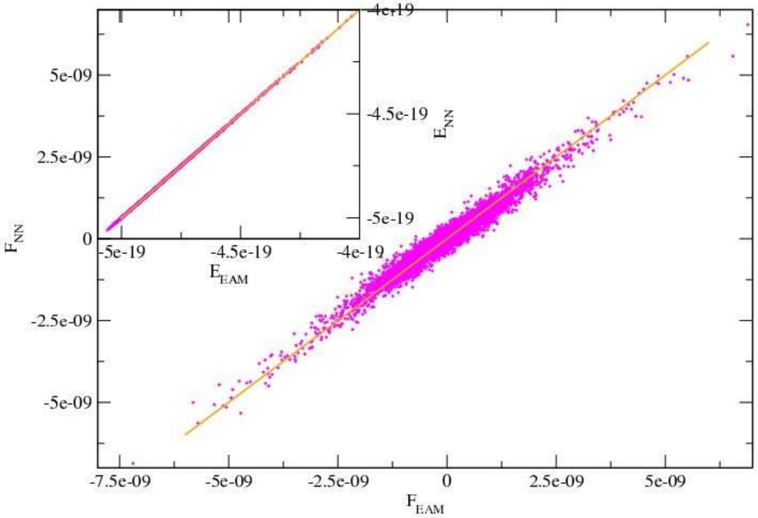


NEURAL NETWORK

LJ potential

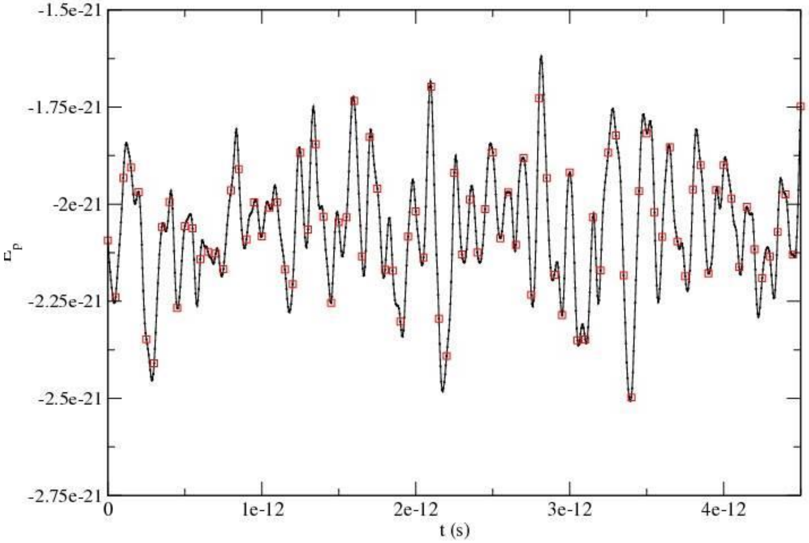


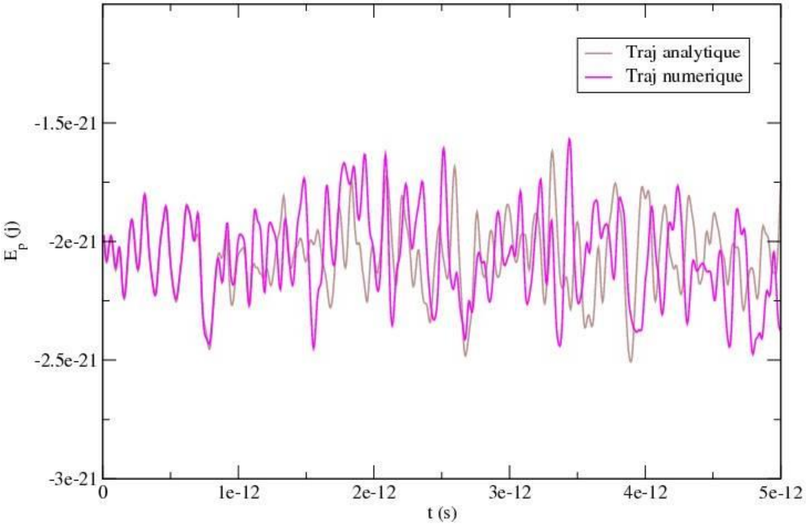
EAM potential



NEURAL NETWORK

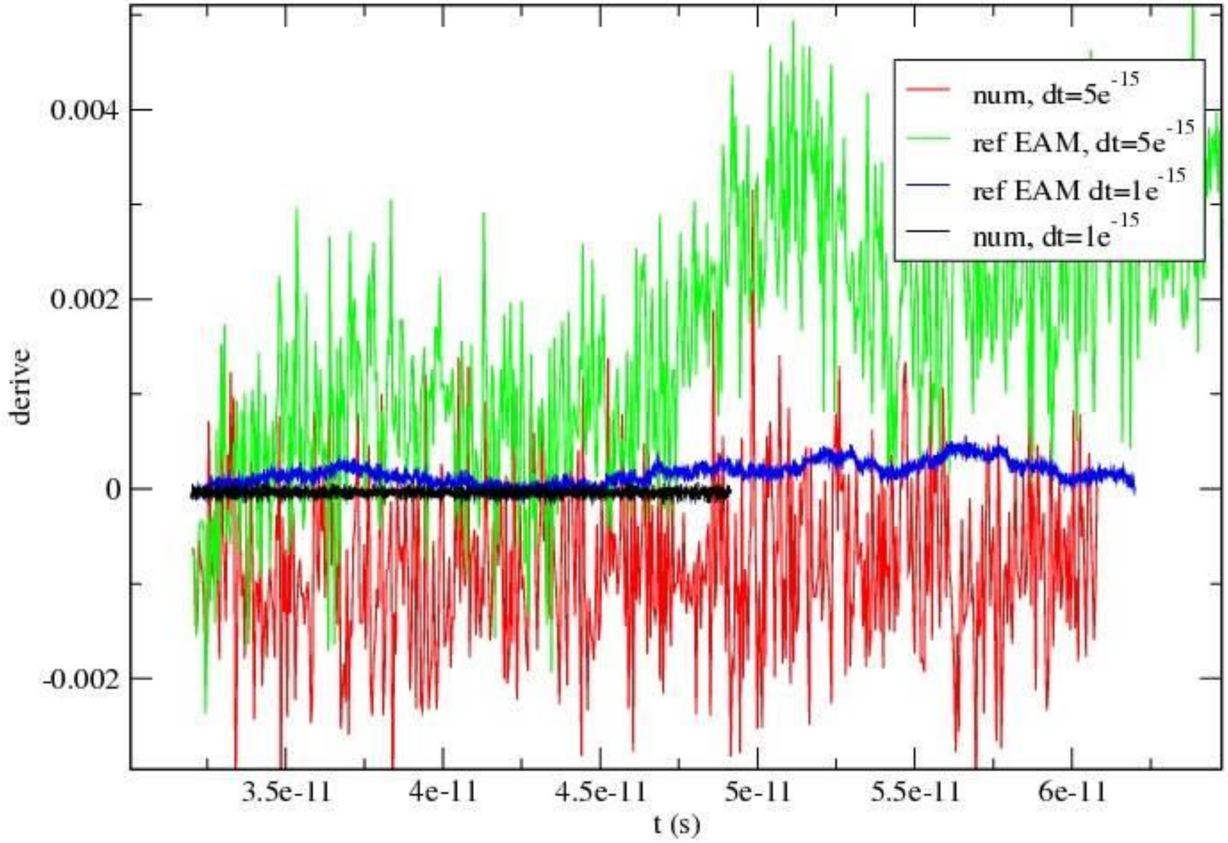
Validation in molecular dynamics





NEURAL NETWORK

Validation: energy conservation



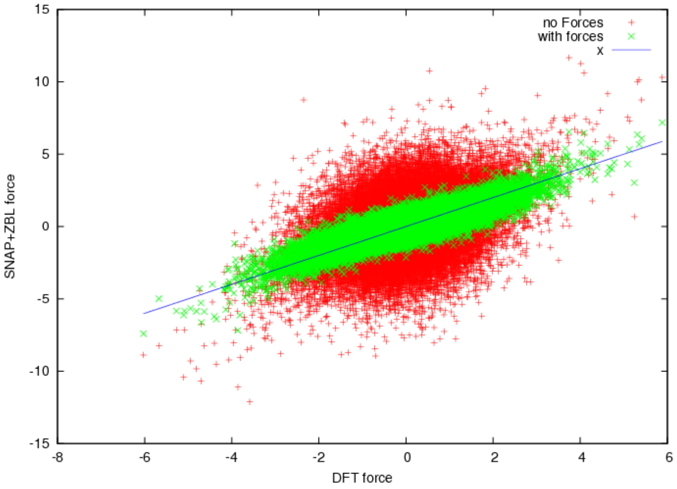
NEURAL NETWORK

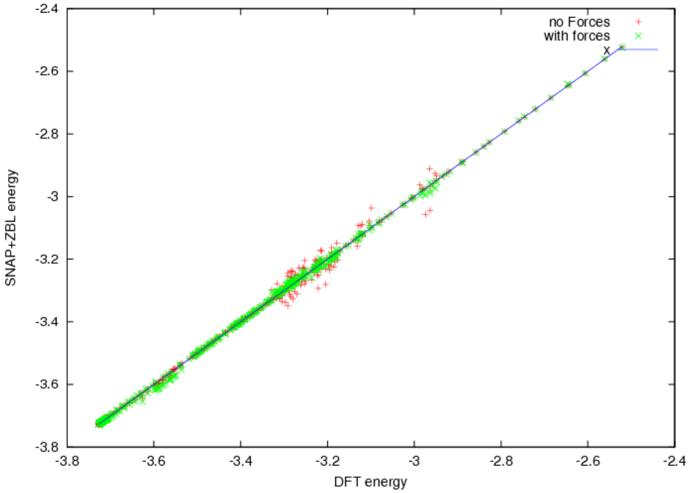
Characteristics:

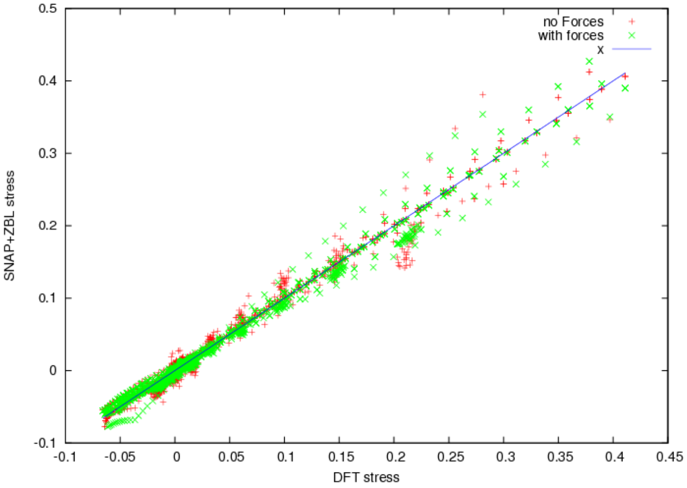
- numerically stable
- efficient parallelization
- compatible GPU …

Questions :

- **fitting on forces ?** No backpropagation for derivatives, optimization in large dimensions ?
- control of the error in prediction : interpolation vs extrapolation (see database)







KERNEL METHOD

Building an approximation of f from a database (x_i, y_i)

Kernel estimation: $$f(x) = \sum_i \alpha_i K(x, x_i)$$

A similarity measure K defines an approximation space H

↓

Defining a distance between configuration

↕

use a kernel method

For a new x, we want to choose y such that (x,y) be similar to the elements of the database. For the similarity measure we take K(x,x’).

KERNEL METHOD

K is a similarity measure between x and x': it measures the correlation

Ingredient for RKHS :

- Kernel K
- Reproducing Kernel Hilbert Space $\mathcal{H}$
- Database $(x_i, y_i)^n$, $y_i = f(x_i)$

We define $J(f)$ as a cost to minimize:

$$J(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}(x_i) - y_i\right)^2 + \lambda\|\hat{f}\|_{\mathcal{H}}$$

Representer theorem: the minimum of $J(\hat{f})_{\hat{f}\in\mathcal{H}}$ can be written as :

$$f(x) = \sum_i \alpha_i K(x, x_i) \qquad \text{with} \qquad \alpha = (K + n\lambda I)^{-1} y$$

Note: the CPU scales with the number of elements in the database

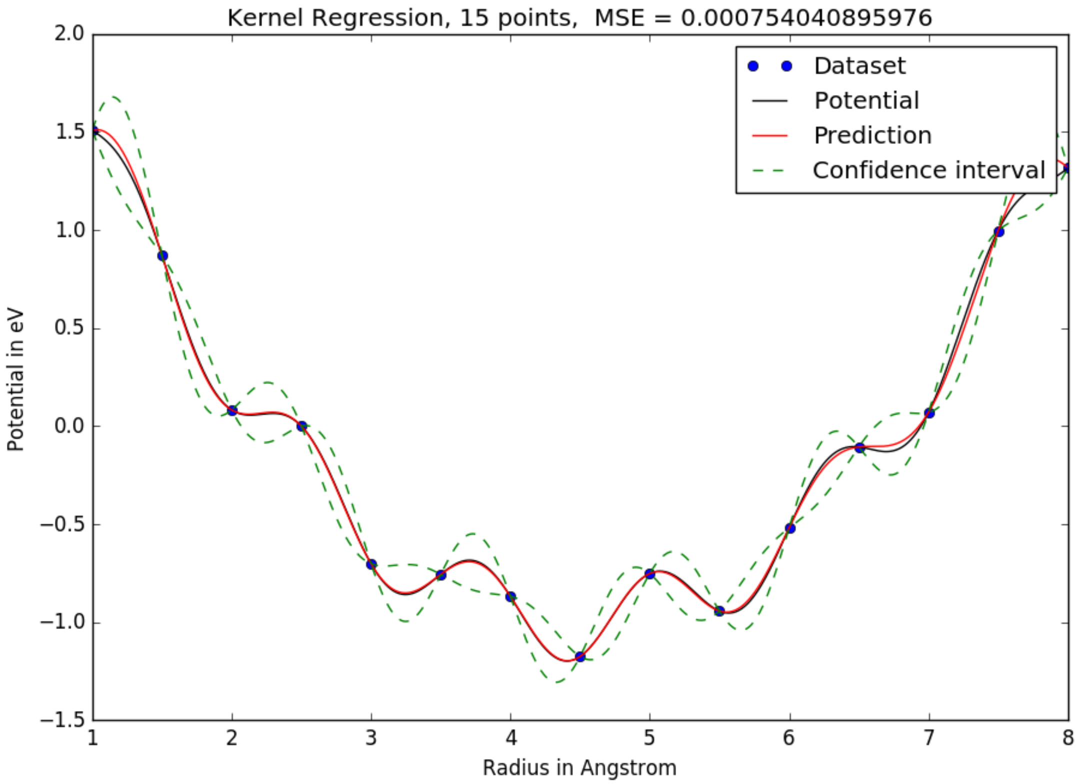
KERNEL METHOD

A bayesian approach allows to compute the variance associated with the prediction

$$\sigma^2(x) = \sigma_0^2 - v^T K^{-1} v$$

Example $f(x) = \frac{\cos 5x + (x-4.5)^2}{5} - 1$

$$v = \begin{pmatrix} v_1 \\ \cdots \\ K(x, x_i) \\ \cdots \\ v_n \end{pmatrix}$$



KERNEL METHOD

Bypassing the use of descriptor: SOAP

The key point in fitting a PES is the similarity measure $K(C_1, C_2)$ → *$K(C_1, C_2)$ should have propert invariances, and smoothness*

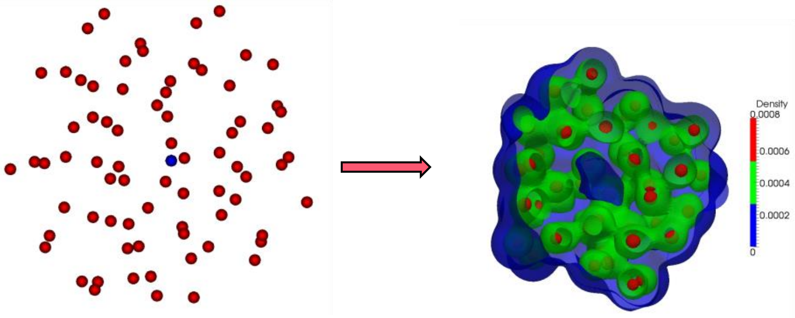
Similarity:

$$\mathbf{S}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \int_{\mathbb{R}^3} \boldsymbol{\rho}_1(\mathbf{r})\boldsymbol{\rho}_2(\mathbf{r})\mathbf{dr}$$

Rotationally invariant kernel:

$$\mathbf{k}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \int \left|\boldsymbol{S}(\boldsymbol{\rho}_1, \hat{\boldsymbol{R}}\boldsymbol{\rho}_2)\right|^{\boldsymbol{n}} \mathbf{d}\hat{\boldsymbol{R}}$$

Expansion of the neighbor density:



$$\rho(r) = \sum_i e^{-\alpha|r-r_i|^2} = \sum_i \sum_{l,m} c_{l,m}(r) Y_{l,m}(\hat{r})$$

SOAP kernel:

$$\mathbf{K}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \left(\frac{\boldsymbol{k}(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)}{\sqrt{\boldsymbol{k}(\boldsymbol{\rho}_1,\boldsymbol{\rho}_1)\boldsymbol{k}(\boldsymbol{\rho}_2,\boldsymbol{\rho}_2)}}\right)^{\xi}$$

A.P. Bartók, M.C. Payne, and G. Csányi, Phys. Rev. B., **87**, 184115, 2013.

KERNEL METHOD

Functional Representation of Atomic Configuration

$$d_2(C_1, C_2)^2 = \|\rho_1 - \rho_2\|_{L^2}^2 = \int_{\mathbb{R}^3} (\rho_1 - \rho_2)^2$$

Similarity: $S(\rho_1, \rho_2) = \int_{\mathbb{R}^3} \rho_1 \rho_2$

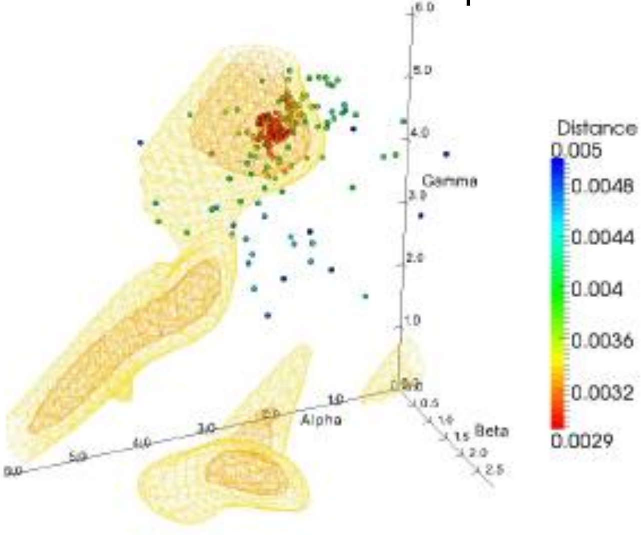
Distance between 2 configurations: $\|\rho_1 - \rho_2\|_{L^2}^2 = S(\rho_1, \rho_1) - 2S(\rho_1, \rho_2) + S(\rho_2, \rho_2)$

Gaussian case

$$S(\rho_1, \rho_2) = \frac{8(\pi\sigma^2)^{\frac{3}{2}}}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \exp\left(-\frac{(q_i - q'_j)^2}{4\sigma^2}\right)$$

Finding the best rotation: simulated anneling in the space defined by the rotation angles vs systematic exploration. The minimum corresponds to the shortest distance

Distance between 2 configurations in a LJ liquid



⇒ Too costly for MD applications

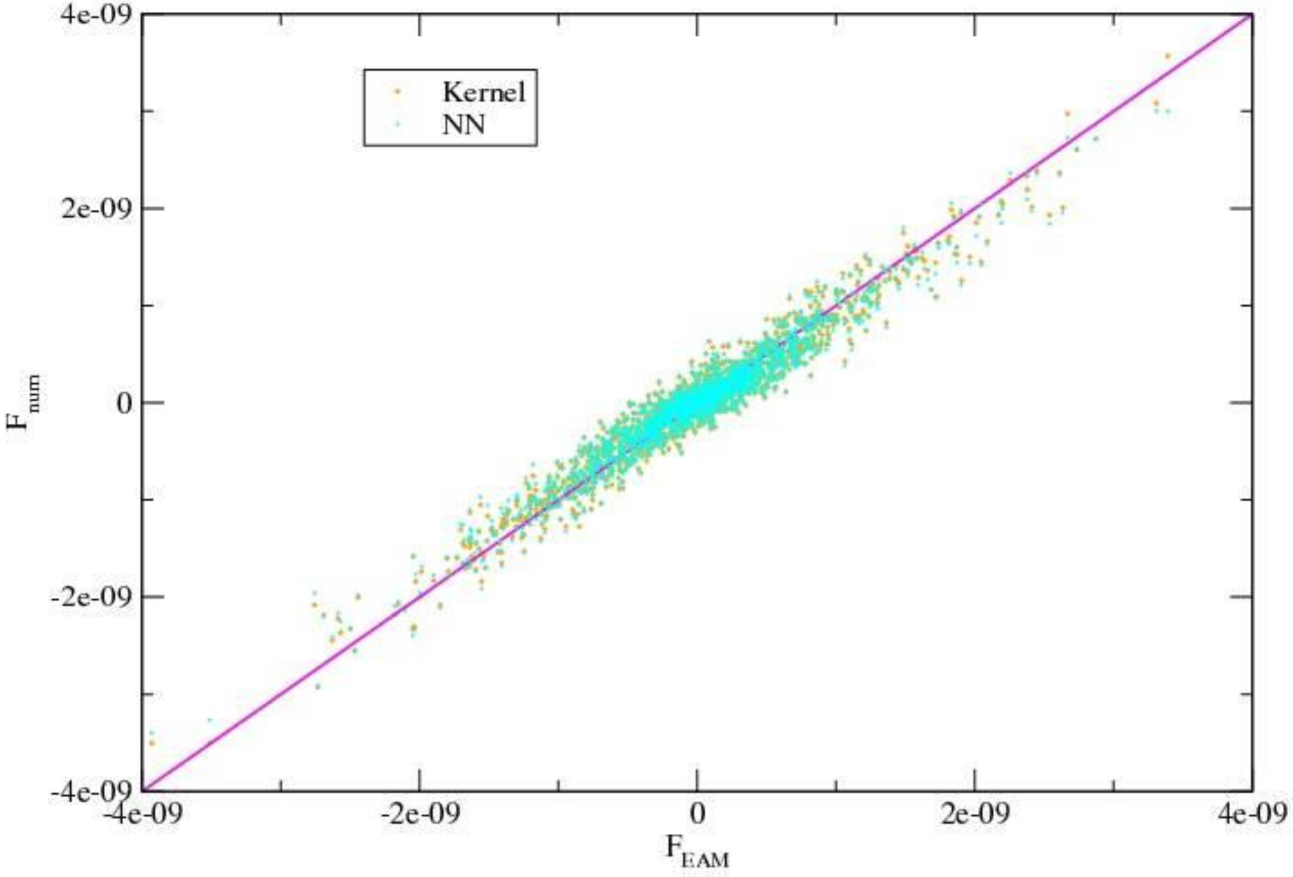
G. Ferré, J.-B. Maillet, and G. Stoltz, J. Chem. Phys., **143**, 104114, 2015.

KERNEL METHOD

Comparison kernel-NN for EAM potential database (descriptor: symmetry functions)

Kernel function:

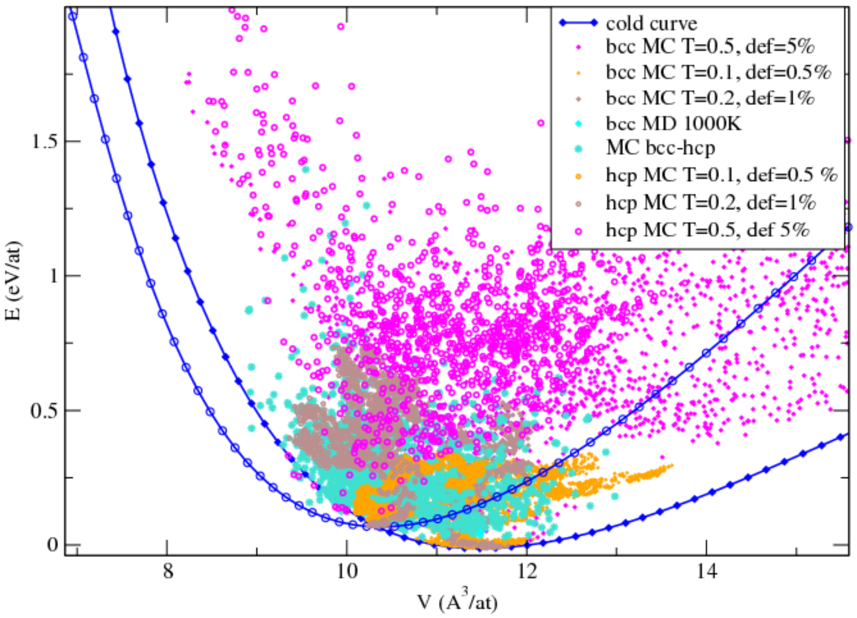
$$k(x,y) = e^{-\frac{(x-y)^2}{2\sigma^2}}$$

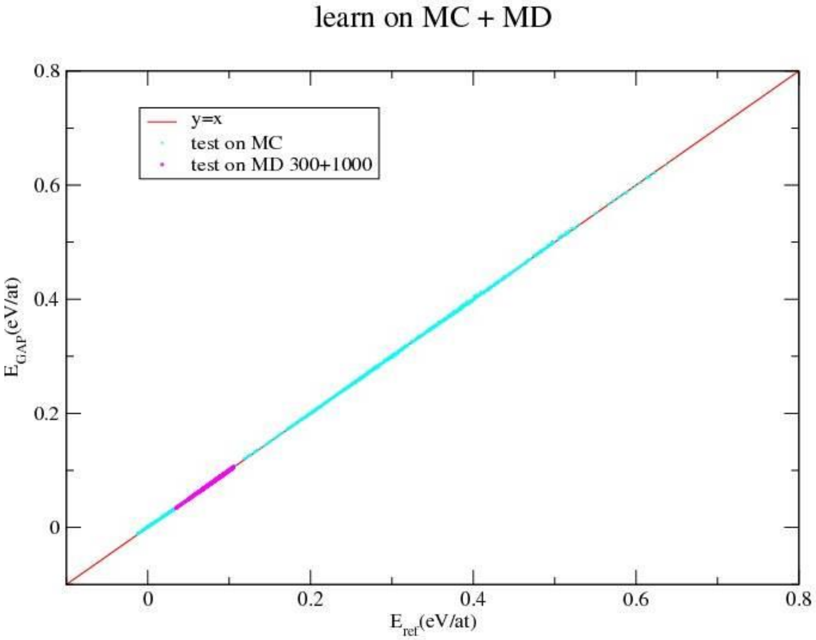


KERNEL METHOD

Iron : there is no classical potential that describes both plasticity and phase transition

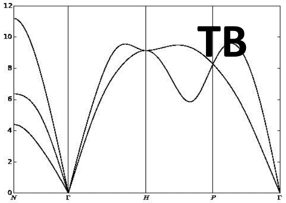
SOAP kernel

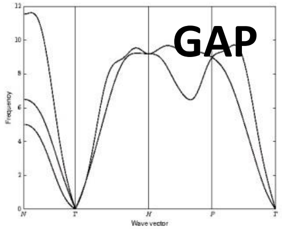



learn on MC + MD

KERNEL METHOD

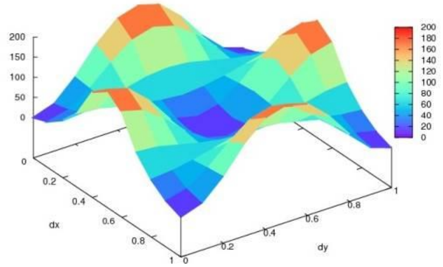
Numerical potential for the bcc-hcp transition in Iron

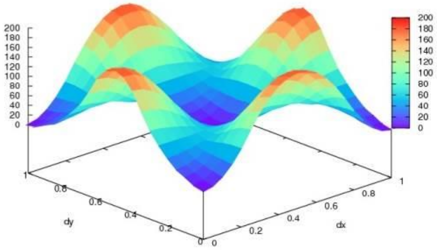




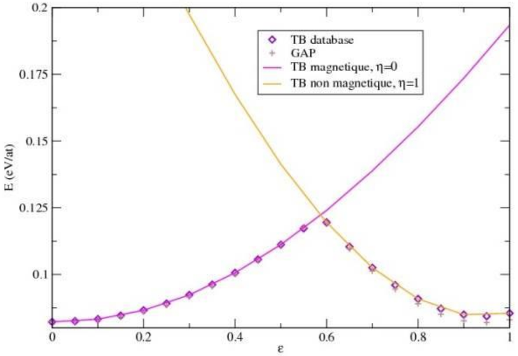
	C_{11}	C_{12}	C_{44}
ref	259.8	154.2	141.7
Pot Num	226.1	138.3	143.4

bcc



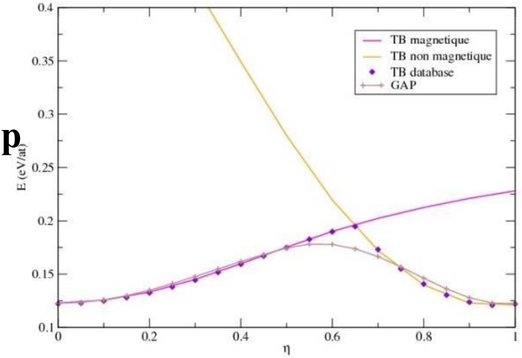


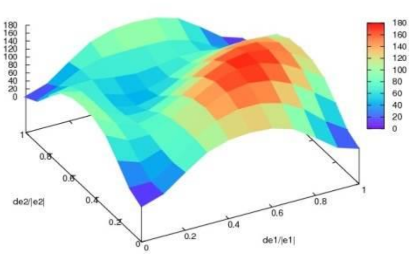
bcc-hcp transition

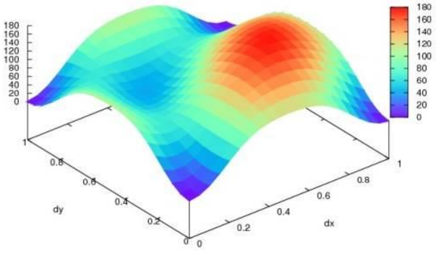


	C_{11}	C_{12}	C_{33}	C_{44}	C_{66}
ref	611.6	149.5	566.4	200.9	231.1
Pot Num	646.8	159.7	593.5	206.8	243.6

hcp







J.-B. Maillet, C. Denoual andG. Csányi, APS-SCCM Proc, 2017.

LINEAR REGRESSION

SNAP POTENTIAL: descriptor : Bispectrum SO4 ($\boldsymbol{B_k}$)

atom energy $$E_i = \beta_0 + \sum_{k=1}^{K} \beta_k B_k^i$$

total energy $$E_{tot} = N\beta_0 + \sum_{i=1}^{N} \sum_{k=1}^{K} \beta_k B_k^i = N\beta_0 + \sum_{k=1}^{N} \beta_k \sum_{i=1}^{N} B_k^i$$

force $$\boldsymbol{F}_i = -\boldsymbol{\beta} \cdot \sum_{j=1}^{N} \frac{d\boldsymbol{B}^j}{d\boldsymbol{r}_i}$$

virial $$\boldsymbol{W} = -\boldsymbol{\beta} \cdot \sum_{i=1}^{N} \boldsymbol{r}_i \otimes \sum_{j=1}^{N} \frac{d\boldsymbol{B}^j}{d\boldsymbol{r}_i}$$

A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, and G.J. Tucker. J. Comput. Phys., **285** :316, 2015.

LINEAR REGRESSION

FITTING SNAP POTENTIAL

For each configuration in the database, we want:

1

$$E_{SNAP} = N\beta_0 + \sum_{k=1} \beta_k \sum_{i=1}^{N} B_k^i = E^{DFT}$$

$$\vdots$$

3N

$$\boldsymbol{F}_{SNAP}^{i} = -\sum_{k=1} \beta_k \sum_{j=1}^{N} \frac{d\boldsymbol{B}^j}{d\boldsymbol{r}_i} = F_i^{DFT}$$

$$\vdots$$

6

$$\boldsymbol{W}_{\boldsymbol{SNAP}} = -\sum_{k=1} \beta_k \sum_{i=1}^{N} \boldsymbol{r}_i \otimes \sum_{j=1}^{N} \frac{d\boldsymbol{B}^j}{d\boldsymbol{r}_i} = W^{DFT}$$

LINEAR REGRESSION

Set of linear equations:

FITTING SNAP POTENTIAL

$$\begin{pmatrix} N & \dots & \sum_{j=1}^{N} B^j \\ \vdots & & \vdots \\ 0 & \dots & -\sum_{j=1}^{N} \frac{dB^j}{dr_i^\alpha} \\ \vdots & & \vdots \\ 0 & \dots & -\sum_{i=1}^{N} r_i^\beta \sum_{j=1}^{N} \frac{dB^j}{dr_i^\alpha} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} E^{DFT} \\ \vdots \\ F_{i,\alpha}^{DFT} \\ \vdots \\ W_{\alpha,\beta}^{DFT} \end{pmatrix}$$

$$\Updownarrow$$

$$\boldsymbol{A} \cdot \boldsymbol{\beta} = \boldsymbol{y}$$

The matrix equation is solved for β using QR factorization (no inversion).

LINEAR REGRESSION

FITTING SNAP POTENTIAL

➡ **‘Pure’ SNAP** $E_{snap} = E_{DFT}$

Linear regress the β_k coefficients to fit reference values (i.e. snap does all the job) :

➡ **‘reference’ + SNAP** $E_{snap} = E_{DFT} - E_{REF}$

Use a reference potential (for example 2 body) to account for the main part of the energy and forces, then linear regress the β_k coefficients to fit the difference between DFT and reference potential values. The reference potential stabilizes the dynamic.

$$\boldsymbol{y} = \begin{pmatrix} E^{DFT} - E^{ref} \\ \vdots \\ F_{i,\alpha}^{DFT} - F_{i,\alpha}^{ref} \\ \vdots \\ W_{\alpha,\beta}^{DFT} - W_{\alpha,\beta}^{ref} \end{pmatrix}$$

LINEAR REGRESSION

FITTING A PURE SNAP POTENTIAL FOR Ge

- Buid the **A** matrix = bispectrum coefficients
- Build the **y** vector = DFT – reference potential values
- Solve the $\boldsymbol{A}.\boldsymbol{\beta} = \boldsymbol{y}$ equation for β. The coefficients β are the SNAP coefficients.

→ This works nicely (not an optimization problem), but results are not optimum due to differences in the nature of the reference points (Esolid vs Eliquid vs Ssolid vs Fliquid…)
Use weights as hyper parameters of the model

In practice the database is splitted into groups (crystal, liquid, metastable) with associated weights (for E,F,W). The weighted matrix equation is solved in the same way, and the weightvector is optimized.

$$\boldsymbol{\omega} \cdot \boldsymbol{A} \cdot \boldsymbol{\beta} = \boldsymbol{\omega} \cdot \boldsymbol{y}$$

Use of a differential evolution algorithm to optimize the weights.

LINEAR REGRESSION

DIFFERENTIAL EVOLUTION

Goal : minimize a function f relatively to its parameters {x}. No need of gradient evaluation

- Start with a population of sets of parameters $\{x\}_i$

- Each set of the population is mutated as:

$$\{x\}_{i,mutant} = \{x\}_{best} + (\{x\}_j - \{x\}_k)$$

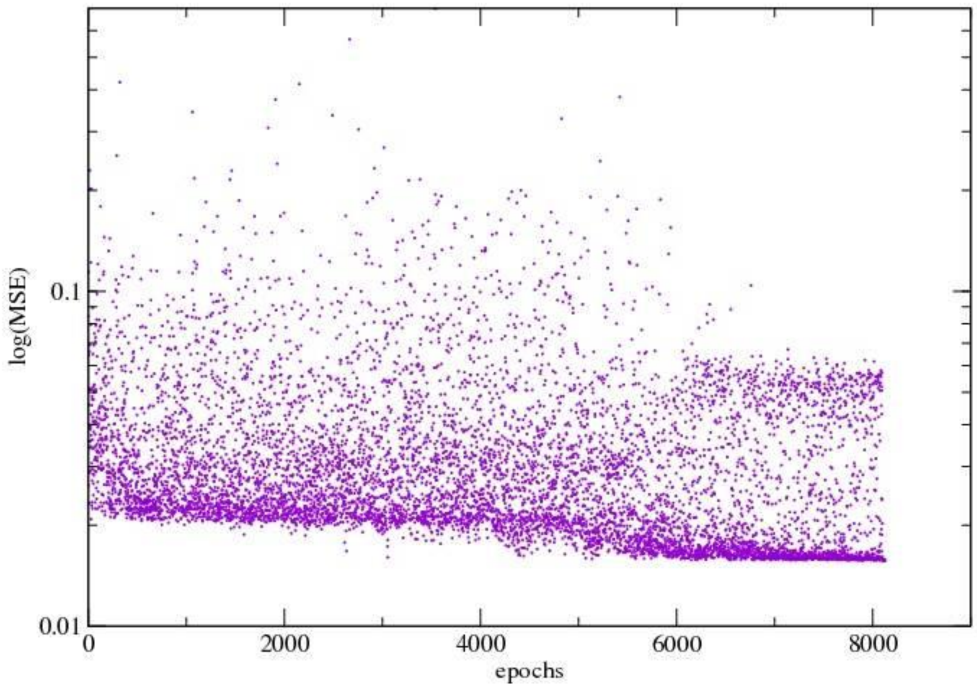
- Each parameter of the mutant replace the original one with a recombination probability (given as an input)

- The fitness (i.e. cost function) is evaluated for the new candidate. The new candidate replaces the old one if better.

LINEAR REGRESSION

Differential evolution

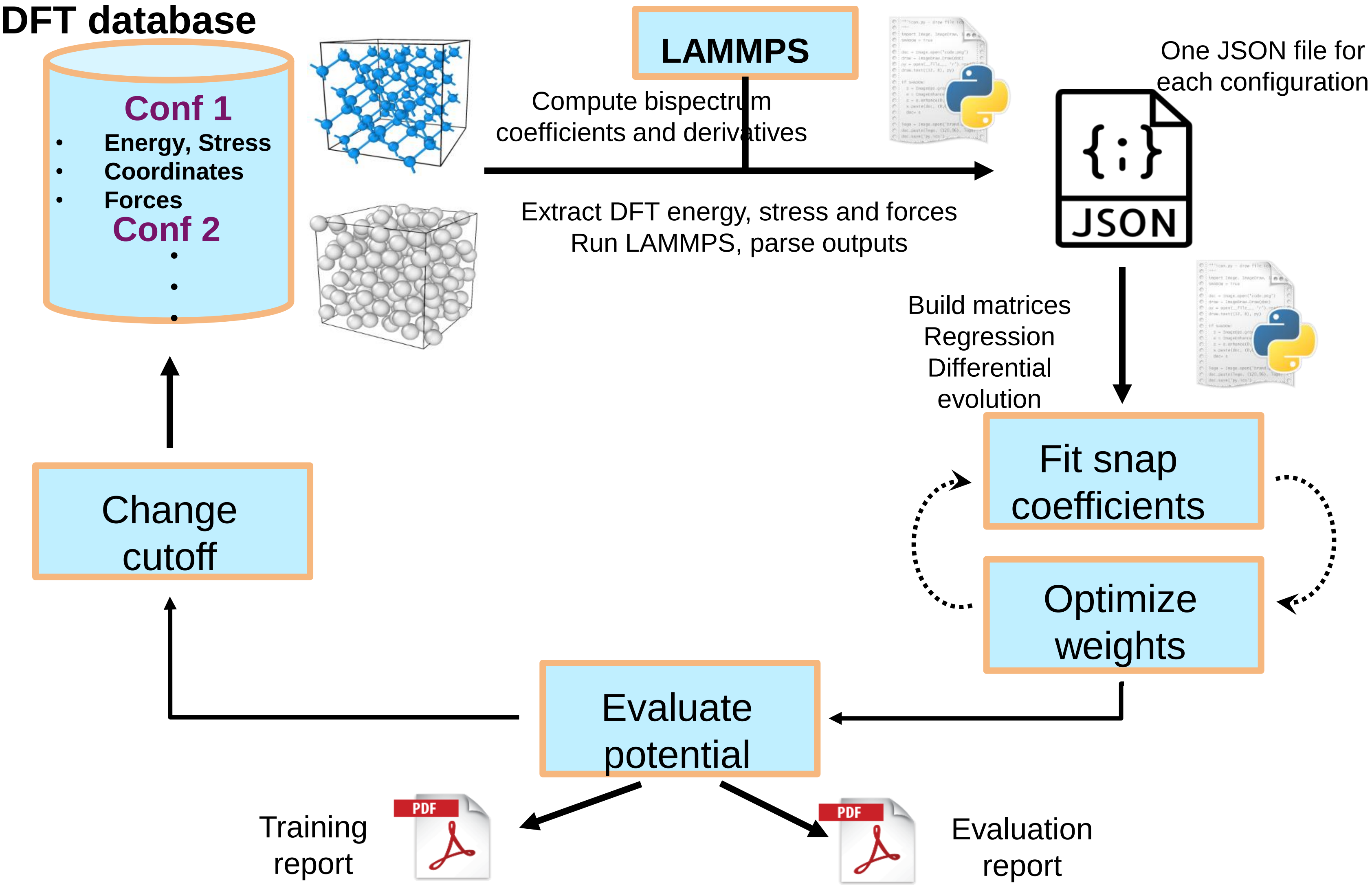
Optimization of 6 weights for a snap potential on Iron



Questions: Efficient optimization method in more dimensions ?

LINEAR REGRESSION

WORKFLOW



DFT database

Conf 1

- Energy, Stress
- Coordinates
- Forces

Conf 2

•
•
•

LAMMPS

Compute bispectrum coefficients and derivatives

Extract DFT energy, stress and forces
Run LAMMPS, parse outputs

One JSON file for each configuration

Build matrices
Regression
Differential evolution

Fit snap coefficients

Optimize weights

Evaluate potential

Change cutoff

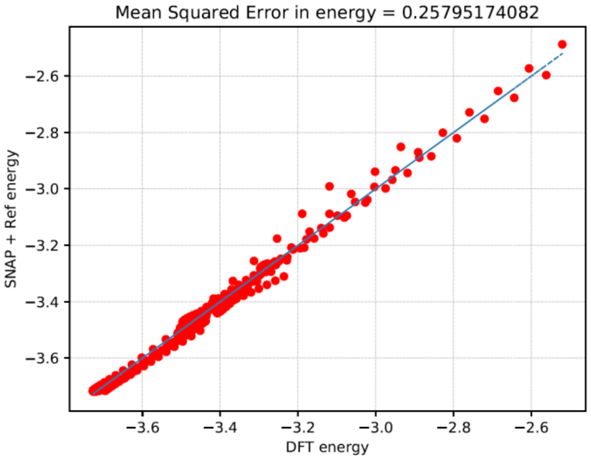
Training report

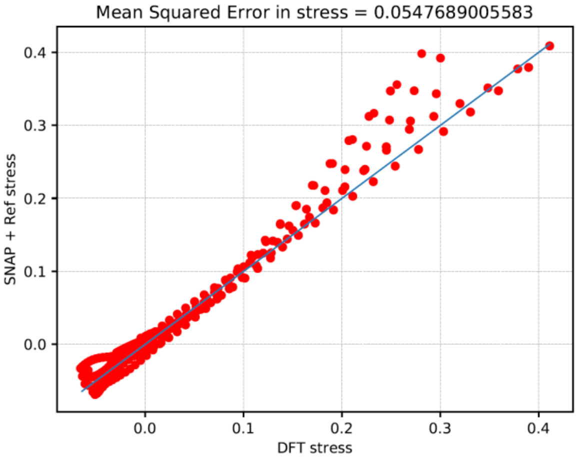
Evaluation report

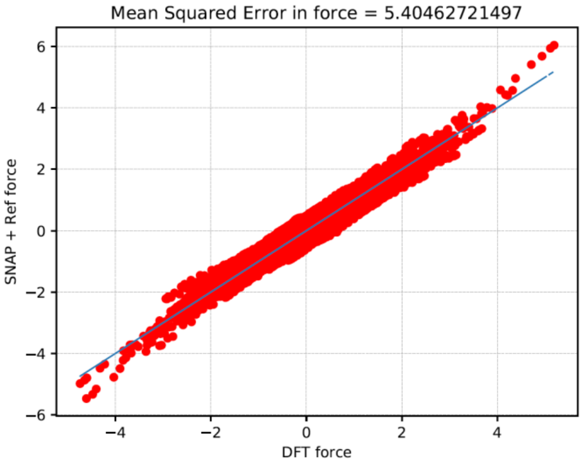
DE LA RECHERCHE À L'INDUSTRIE

LINEAR REGRESSION

SNAP Ge (best potential so far)


Mean Squared Error in energy = 0.25795174082

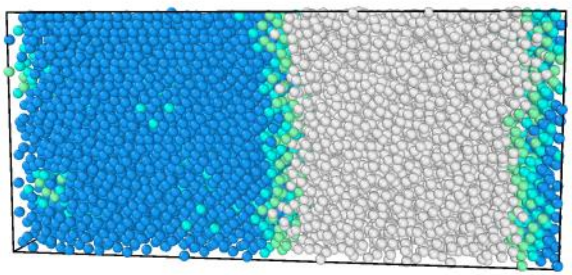

Mean Squared Error in stress = 0.0547689005583


Mean Squared Error in force = 5.40462721497

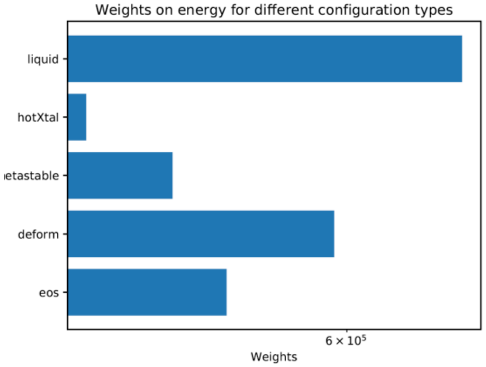
Method	Vacancy formation energy (eV)
DFT	2.33
Tersoff	3.72
SNAP	1.64

Method	Surface energy (111) (eV/Å^2)
DFT	0.07
Tersoff	0.205
SNAP	0.076

Coexistence melting point calculation
SNAP melting point ~ 700 K
Tersoff melting point ~ 2560 K



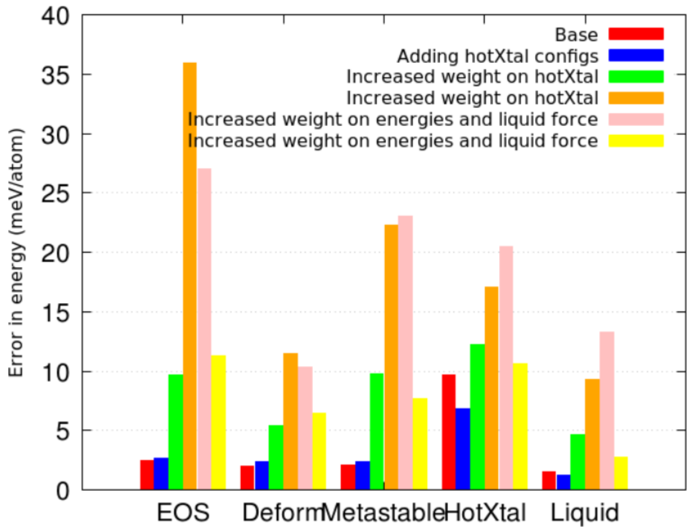
LINEAR REGRESSION

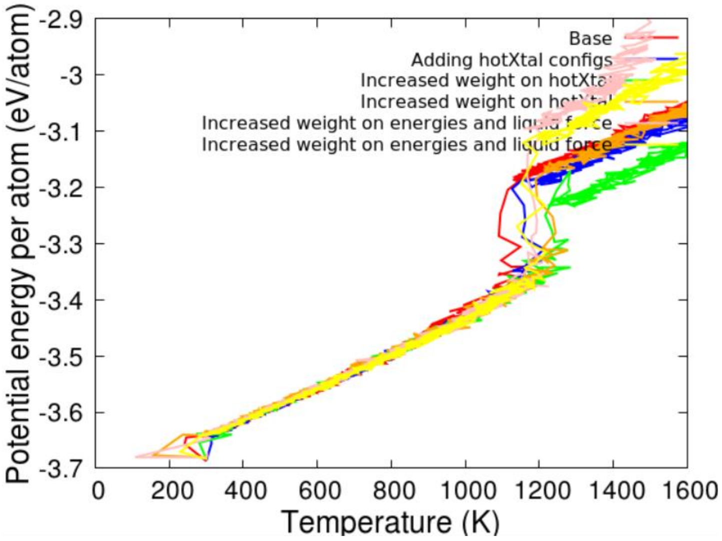


“adjusting/optimizing” the weights

↕

Infinity of different potentials





CONCLUSION

- Accuracy versus transferability ?
- Infinity of potential (database, weights): rationalization of the choice (use Pareto ?).
- 2 step process: learning and computing

Can we mix the 2 (on the fly or active learning) ? How can we integrate a “memory” in regression method without redoing all the learning job (pre-conditioned learning)

- Control the error of the prediction for dynamic
- Can we integrate the invariance into the regression method – so we could use simpler descriptors

QUESTIONS

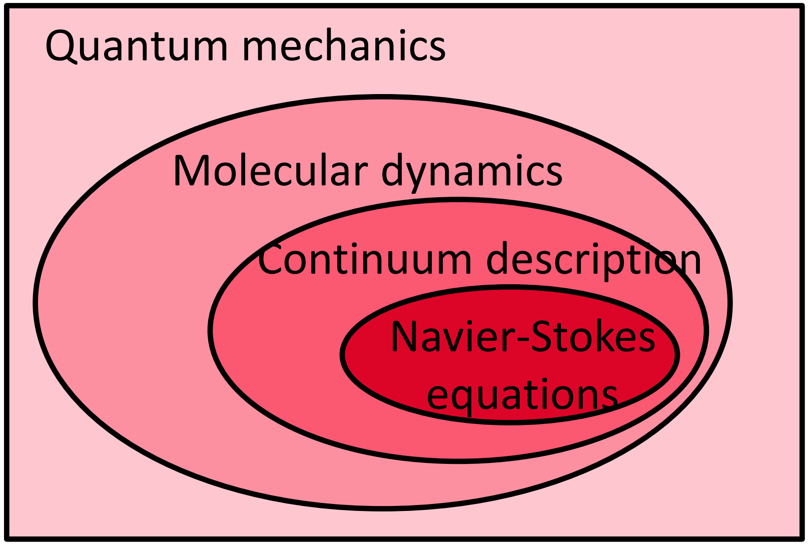
Thank you !

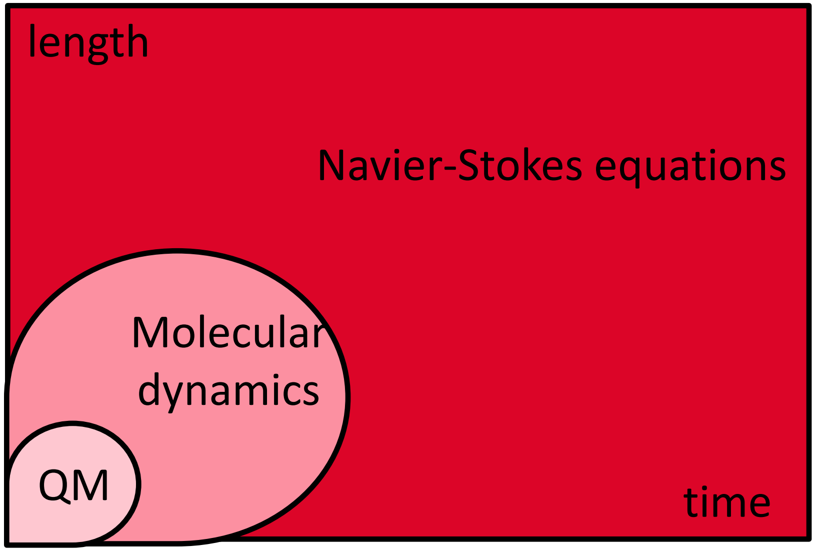
?

CONTEXT

Venn diagram illustrating the hierarchy of frameworks for describing a fluid, in terms of their **theoretical validity in parameter space**

Venn diagram illustrating the hierarchy of frameworks for describing a fluid, in terms of **their space-time domain of applicability**



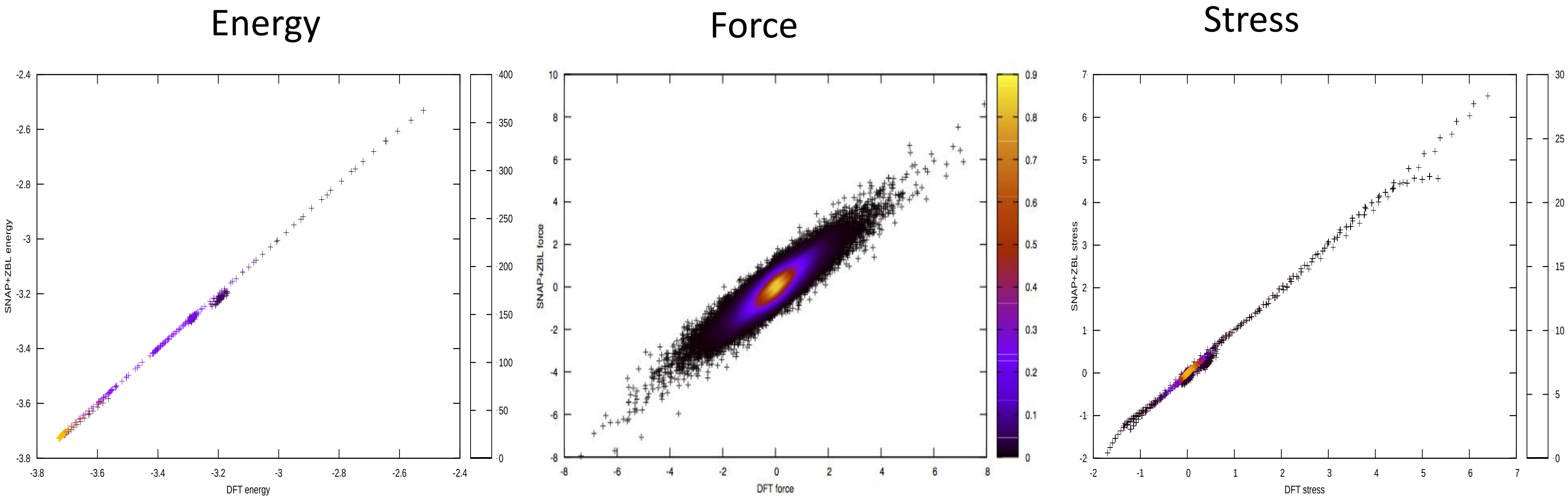


K. Kadau, J.L. Barber, T.C. Germann, B.L. Holian and B.J. Alder, Phil. Trans. R. Soc. A (2010) **368**, 1547.
B.L. Holian, C.W. Patterson, M. Mareschal, and E. Salomons, Phys. Rev. E. **47**, r24 (1993).
B.L. Holian, M. Mareschal, and R. Ravelo, Phys. Rev. E. **83**, 026703 (2011).

LINEAR REGRESSION

Ge ZBL+SNAP POTENTIAL

Correlations



Energy

Force

Stress

LINEAR REGRESSION

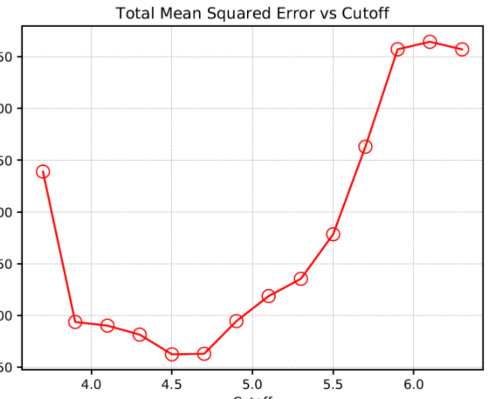
Mean absolute error for this category

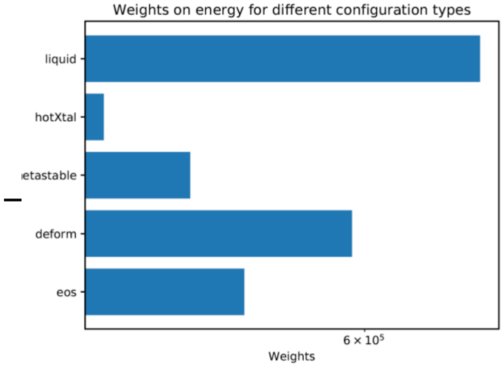
Bounds and weight in objective function

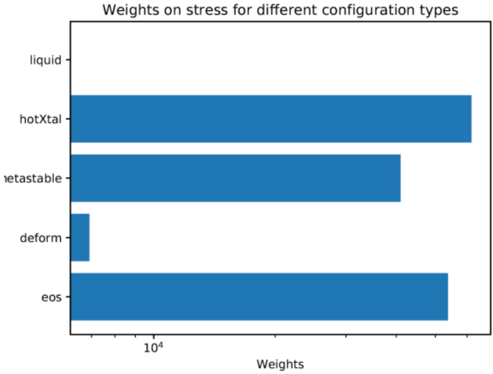
Weight after optimization

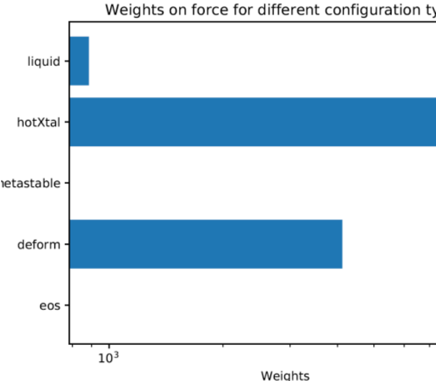
Error reported in Molybdenum SNAP

Configuration type	$N_{configs}$	N_{atoms}	**Energy** (8.9 meV/atom)	**Force** (0.30 eV/Å)	**Stress** (0.89 GPa)
Diamond EOS	41	8	(1e4, 1e6, 500) [573513] (9.63 meV/atom)	-	(1e2, 1e5, 1) [53797] (0.61 GPa)
Diamond Deform	246	8	(1e4, 1e6, 500) [597205] (5.45 meV/atom)	(1e2, 1e4, 1) [4125] (0.17 eV/Å)	(1e2, 1e5, 1) [6920] (0.16 GPa)
BCC, FCC, SC EOS	164	2/4/1/4	(1e4, 1e6, 500) [561968] (9.81 meV/atom)	-	(1e2, 1e5, 1) [41018] (2.5 GPa)
Diamond @ 800K	100	216	(1e4, 1e6, 500) [544007] (12.25 meV/atom)	(1e2, 1e4, 100) [9637] (0.12 eV/Å)	(1e2, 1e5, 100) [61534 (0.27 GPa)
Diamond @ 1000K	100	216	(1e4, 1e6, 500) [544007]	(1e2, 1e4, 100) [9637]	(1e2, 1e5, 100) [61534 (0.27 GPa)


Total Mean Squared Error vs Cutoff
Cutoff


Weights on energy for different configuration types
liquid
hotXtal
etastable
deform
eos
6 × 10^5
Weights


Weights on stress for different configuration types
liquid
hotXtal
etastable
deform
eos
10^4
Weights


Weights on force for different configuration ty
liquid
hotXtal
etastable
deform
eos
10^3
Weights