Machine learning of equivariant functions inspired by atomistic modelling and three-dimensional image processing

Bastiaan J. Braams, Centrum Wiskunde & Informatica (CWI), Amsterdam, Netherlands

Presentation in PPPL Theory Seminar, 2019-11-27 (previously at GSCAEP, Beijing, 2019-09-18

Potential energy surfaces for atomistic dynamics

N atoms; classical nuclei, positions $x(i)$, $1 \leq i \leq N$.

Interaction potential $V(X)$ ($X \in \mathbf{R}^{3N}$). Force $F = -\nabla V$ and then "F=ma" dynamics.

Electron dynamics is gone. This is the Born-Oppenheimer (adiabatic) approximation.

Also of interest, the dipole moment surface (DMS) $\mathbf{d}(X)$, quadrupole moment surface, polarizability and other properties.

Fundamentally $V(X)$, $\mathbf{d}(X)$, etc., are obtained by expensive electronic structure calculations. We are concerned here with fitted surfaces that are cheap to evaluate.

Local representation: $V = \sum_i V_0(X^{(i)})$ where $X^{(i)}$ are collective nuclear coordinates for an environment of the i-th atom.

Other uses of potential energy surfaces

Born-Oppenheimer approximation allows quantum nuclei; it is not limited to semiclassical molecular dynamics.

Molecular spectroscopy: Eigenvalue problem $H\Psi = E\Psi$ for the nuclear wavefunction. Tractable for small molecules.

Diffusion Monte Carlo for sampling the ground state nuclear wavefunction: Random walk with birth and death processes.

Quantum statistics: $< A >_\beta = \frac{1}{Z(\beta)} \mathrm{tr}(Ae^{-\beta H})$. Thermal averages calculated using Path Integral Monte Carlo.

Ring Polymer Molecular Dynamics and variants; PIMC plus time evolution. Model for nuclear quantum effects.

Quantum scattering: $i\hbar \frac{\partial}{\partial t} \Psi(X,t) = H\Psi(X,t)$. Application to reaction dynamics is pretty much limited to 4-atom systems.

Three-dimensional image processing

Three-dimensional (3D) objects can be represented by a voxel image; the direct analog of a two-dimensional pixel image.

In many cases it suffices to characterize only the surface of the 3D object. Then a voxel image is very inefficient.

Consider a 3D object represented by its surface, and the surface represented by a cloud of points.

Each point may carry some properties: material, colour, ...

Want to learn properties of the 3D object.

Similarities to atomistic modelling: input data based on points in R^3; desired invariance under translation and rotation.

Different from atomistic modelling: no significant pairwise interaction; no invariance under inversion.

Invariants, covariants, equivariants

Vector spaces U and V, group G with representations on U and V. For $g \in G$ we write $g_U.u$ or $g_V.v$ for the action of g on a typical element $u \in U$ or $v \in V$.

Function $f : U \to V$ is covariant (equivariant) if for all $g \in G$ and $u \in U$, $f(g_U.u) = g_V.f(u)$.

(Invariants are the special case $g_V.v \equiv v$.)

Example, the dipole moment $\mathbf{d}(X)$, $X \in R^{3N}$. Invariant under $\mathrm{Sym}(N)$, covariant under $O(3)$.

Represent it by effective charge model: $\mathbf{d}(X) = \sum_i w_i(X)\mathbf{r}_i$; then the weight vector $w \in R^N$ is covariant under $\mathrm{Sym}(N)$ and invariant under $O(3)$.

Fitting or learning in the presence of symmetries

Example: Want to fit or learn $f : \mathbf{R}^N \to \mathbf{R}$, $f(x) = z$, using data $f(X_\alpha) = z_\alpha$. Typical point $x = (x_1, ..., x_N)$. Say that the underlying true function is totally symmetric in the $(x_i)_i$. Options:

(a) Ignore the symmetry, use any plausible model. (Maybe replicate the data using symmetry.) Obtain symmetry via accuracy.

(b) Use explicit invariants of a good functional form. Example: $y_k = p_k(x)$ (where the p_k are elementary symmetric polynomials for $1 \leq k \leq N$), then $f(x) = g(y(x))$ with some plausible model for g. Efficient; technically difficult for more complicated symmetries. (Braams+Bowman at Emory University.)

(c) Use explicit invariants of an easily generalizable form. Example: $y = \mathrm{Sort}(x)$, then $f(x) = g(y(x))$. Introduces nonsmoothness, often discontinuities. Obtain smoothness via accuracy.

New approaches from Big Data and Machine Learning

Machine Learning has brought specific methods, e.g. deep convolutional neural networks (Vision, Go).

Machine Learning is also bringing a change of attitude...

Nothing wrong with optimizing over very many variables (Stochastic Gradient Descent).

Nothing wrong with lots of local minima, even inequivalent ones. Don't ask for a guaranteed global optimum.

(NN with 20 layers and 256 nodes per layer and a ReLU nonlinearity has multiplicity of about 10^{10139}, being $(256!)^{20}$; with tanh nonlinearity about 10^{11680}, being $(2^{256} \times 256!)^{20}$.)

Review of problem statement; focus of remaining talk

Atomistic modelling: Fit or learn an energy, force field or other properties as a function of atomic positions and atom types.

Three-dimensional image processing: Fit or learn properties of an object as a function of coordinates of a cloud of points (and associated properties) on the surface of the object.

The atomistic problem has a long history of empirical, physically motivated force fields (CHARMM, AMBER, GROMOS, Tersoff potentials, EAM approach, ReaxFF).

We are concerned with recent "big data" approaches of two kinds.

* Richly parameterized linear models; big data linear regression.

* Deep neural networks.

Big data approaches based finally on linear regression

(More details in supplementary slides.)

Gaussian Approximation Potential (GAP), Smooth Overlap of Atomic Potentials (SOAP) kernel. Key reference: Bartók, Kondor, Csányi (2013) Phys Rev B 87.

Spherical Wavelet Expansion approach. Key references: Eickenberg, Exarchakis, Him, Mallat, Thiry (2018) J Chem Phys 148; Brumwell, Sinz, Kim, Qi, Hirn (2018) arXiv:1812.02320.

Spectral Neighbor Analysis Potential (SNAP). Key references: Thompson, Swiler, Trott, Foiles, Tucker (2015) J Comput Phys 285; Wood, Thompson (2018) J Chem Phys 148.

Atomic cluster expansion. Key reference: Drautz (2019) Phys Rev B 99.

Behler-Parrinello High-Dimensional Neural Network Potentials (HDNNP)

Key references: Behler, Parrinello (2007) Phys Rev Lett 98; also Behler (2017) Angew Chem Int Ed 56.

Ansatz $V = \sum_i V^{(a_i)}(X_i)$ where a_i denotes the atom type of the i-th atom and X_i are the collective nuclear coordinates for a local environment of the i-th atom.

Input layer for the neural network are symmetry functions of the local environment; invariant wrt rotation and permutation. Radial (two-body) and angular (three-body) functions.

Finally a feed-forward neural network with some activation function ϕ; inter-layer mapping: $v^{(l)} = \phi(W^l_{l-1}.v^{(l-1)} + d^l)$.

Weights W and displacements d to be learned.

SchNet, Deep Tensor network from TU Berlin

Key reference: Schütt, Sauceda, Kindermans, Tkatchenko, Müller (2018) J Chem Phys 148. Also Nature (2017).

Say N atoms. Each NN layer contains atomic feature vectors x_i for each atom ($1 \leq i \leq N$); positions are global parameters.

Transitions between layers, $l-1 \to l$ (before the nonlinearity):

Dense atom-wise, $x_i^l = W_{l-1}^l . x_i^{l-1} + d^l$;

Convolution feature-wise: $x_i^l = (X^{l-1} \star W_{l-1}^{\prime l})_i$. Convolutions depend on relative distances.

Smooth shifted SoftPlus instead of ReLU.

Weights to be fitted as functions of relative positions.

DeepMD approach led by E and Car, Princeton Univ.

Key reference: Zhang, Han, Wang, Car, E (2018) PRL 120.

Recall $V = \sum_i V^{(a_i)}(X_i)$ where a_i is the atom type and X_i are the collective nuclear coordinates for a local environment of the i-th atom.

Local environment is rotated in a problem-dependent manner. Let j enumerate neighbouring atoms within a cut-off distance.

Environment descriptor $D(X_i) = \mathrm{Sort}\{D_{ij}\}$ sorted by chemical species and by distance R_{ij}.

$D_{ij} = (1/R_{ij}, x_{ij}/R_{ij}^2, y_{ij}/R_{ij}^2, z_{ij}/R_{ij}^2)$

Obtain $V^{(a)}$ as output of a deep neural network with inputs the ordered $(D_{ij})_j$.

DeepPot-SE, Deep Potential Smooth Edition, from Princeton University and IAPCM

Key reference: Zhang, Han, Wang, Saidi, Car, E (2018) NIPS.

Follow-on to the DeepMD work, but now with due respect for continuity, energy conservation and vector covariance.

Environment descriptors without rotation or sort:
$D_{ij} = s_{ij}(r_{ij}) \times (1, x_{ij}, y_{ij}, z_{ij})$; $s_{ij} \to 0$ for large r_{ij}.

Two-stage NN: an encoding network and a fitting network.

Encoding network maps local environment to a feature space preserving point group and permutation symmetry.

Fitting network is fully connected feed-forward neural network with skip connections.

Perspective from Machine Learning .. Intro

Basic feed-forward neural network: $v^l = \phi(W^l_{l-1}.v^{(l-1)} + d^l)$.

$v \in R^b$; ϕ e.g. pointwise ReLU or a smooth variant; weights W and displacements d to be learned.

Structured network: $v \in X^b$, $X \sim \mathbf{R}^k$ with some structure. W has a corresponding structure.

Example (Convolutional Neural Network, CNN): $X \sim \mathbf{R}^{m \times n}$, pixelized greyscale image. Elements of W are convolutions with compact kernel.

Parameterized network: Data depend on parameters p, weights to be learned as a function of p. (Linear regression formulation.)

Point Cloud Convolutional Networks

Permutation equivariant neural networks ("Deep Sets"): Zaheer, Kottur, Ravanbhakhsh *et al.*, NIPS 2017.

Structured and parameterized: $X \sim \mathbf{R}^N$, $v \in \mathbf{R}^{N \times b}$; interpret as feature vector of size b with components in $\mathbf{R}^N$ associated with N points in $\mathbf{R}^3$. Parameters p: positions $\mathbf{r}_i$ $(1 \leq i \leq N)$. Weights W represent local convolutions; depend on local distances.

Point cloud convolutional networks are covariant (equivariant) under $\mathrm{Sym}(N)$.

See also: [SpiderCNN convolutional filters on point sets: Xu, Fan, Xu, Zeng, Qiao, Proc ECCV 2018], [PointConv deep convolutions on point sets: Wu, Qi, Fuxin, Arxiv 2018], other work on permutation equivariant NN.

Point Clouds with Additional Structure

Basic point cloud network has $v \in \mathbf{R}^{N \times b}$; feature vector of size b with components in $\mathbf{R}^N$. Group is $\mathrm{Sym}(N)$; feature vectors are otherwise unstructured. Imagine additional structure in the feature vector.

Additional group G of transformations on v, especially permutation group or $\mathrm{SO}(3)$; parameters and feature vector have definite transformation properties under G.

Group-Equivariant CNN: Cohen and Welling, ICML 2016.

Spherical Convolutional Neural Networks: Cohen, Geiger, Köhler, Welling, ICLR 2018.

Tensor Field Networks: Thomas, Schmidt, Kearnes *et al.*, Arxiv, 2018.

Gauge Equivariant CNN: Cohen, Weiler, Kicanaoglu, Welling, Arxiv, 2019.

Conclusions

There are promising new approaches to force fields with inspiration from big data and machine learning. (No assessment here of relative merits.)

There are valuable related developments from machine learning community inspired at least in part by application to atomistic force fields.

BJB wish list for developments and future work ...

Simultaneous learning of energy, dipole, quadrupole moment through local charges with long-range interactions.

Fit or learn bands in solids invariant under $\mathrm{SL}(3, \mathbf{Z})$.

Learn local effective Hamiltonians for excited states.

More studies genuinely focussed on applications.

Other uses of potential energy surfaces

Supplementary slides.

Molecular spectroscopy

Eigenvalue problem $H\Psi = E\Psi$:

$$-\sum_i \frac{\hbar^2}{2m_i}\Delta_i\Psi(X) + V(X)\Psi(X) = E\Psi(X)$$

Configuration interaction approach (Hartree products):

$$\Psi(X) = \sum_\alpha c_\alpha \Psi_\alpha(X)$$

$$\Psi_\alpha(X) = \Pi_i \psi^{(i)}_{\alpha(i)}(x(i))$$

This provides the ro-vibrational spectrum. Tractable for small molecules: e.g. $2(H_2O)$, CH_3OH; up to 9 atoms in our work.

Diffusion Monte Carlo

Ground state wavefunction: $H\Psi = E_0\Psi$.

$$-\sum_i \frac{\hbar^2}{2m_i}\Delta_i\Psi(X) + V(X)\Psi(X) = E_0\Psi(X)$$

Steady state for reaction-diffusion equation:

$$\frac{\partial\Psi}{\partial t} - \sum_i \frac{\hbar^2}{2m_i}\Delta_i\Psi(X) + V(X)\Psi(X) = E_0\Psi(X)$$

Can be solved in many dimensions using random walk with birth and death processes.

Result is ground state energy E_0; plus sample from the ground state wavefunction. (Sample $|\Psi^2|$ via descendant weighting.)

Quantum statistics

Partition function $Z(\beta) = \mathrm{tr}(e^{-\beta H})$. Thermal averages:

$$< A >_\beta = \frac{1}{Z(\beta)} \mathrm{tr}(A e^{-\beta H})$$

Use $e^{-\beta H} = (e^{-(\beta/n)H})^n$; $H = T + V$; $n \to \infty$. Let $\beta_n = \beta/n$, $\omega_n = 1/\beta_n \hbar$;

$$V_n(\mathbf{X}) = \sum_i (V(X_i) + \frac{1}{2} m \omega_n^2 (X_{i+1} - X_i)^2)$$

$$Z_n(\beta_n) = \int e^{-\beta_n V_n(\mathbf{X})} dx$$

Path Integral Monte Carlo.

Ring Polymer Molecular Dynamics

Due to David Manolopoulos (Oxford). PIMC plus time evolution. Classical hamiltonian:

$$H_n(x,p) = V_n(x) + \sum_i \frac{p_i^2}{2m}$$

$$\frac{dx}{dt} = \frac{\partial H_n}{\partial p}\ , \quad \frac{dp}{dt} = -\frac{\partial H_n}{\partial x}$$

Seen as a model for calculating the quantum Kubo correlation function.

$$\tilde{c}_{A,B}(t) = \frac{1}{\beta Z(\beta)} \int_0^\beta \mathrm{tr}(e^{-(\beta-\lambda)H} A(0) e^{-\lambda H} B(t)) d\lambda$$

Quantum scattering

Time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi(X,t) = H\Psi(X,t)$$

Wavepacket propagation in an unbounded domain. Application to reaction dynamics is pretty much limited to 4-atom systems.

Big data approaches based finally on linear regression

Supplementary slides.

GAP-SOAP approach of G. Csányi, B. Bartók et al.

Key reference: Bartók, Kondor, Csányi (2013) Phys Rev B 87.

Gaussian Approximation Potential (GAP), also referred to as Kernel Ridge Regression, high-dimensional version of Radial Basis Functions. Authors use both language of Machine Learning and language of function fitting, regression analysis.
$f(X) = \sum_\alpha w_\alpha K(X, X_\alpha)$.

Smooth Overlap of Atomic Potentials (SOAP) kernel $K(X, X')$.

$X \mapsto \rho$, $S(\rho, \rho') = \int \rho(r)\rho'(r)dr$.

$K(X, X') = \int |S(\rho, R.\rho')|^n \, dR$, $R \in O(3)$.

Integrals evaluated via spherical harmonic expansion.

Spherical Wavelet Expansion approach led by S. Mallat

Key references: Eickenberg, Exarchakis, Him, Mallat, Thiry (2018) J Chem Phys 148; Brumwell, Sinz, Kim, Qi, Hirn (2018) arXiv:1812.02320.

General with respect to chemical elements.

Global transform, no explicit reference to local environments.

Smoothed densities $\rho(r) = \sum_i n_i g(r - r_i)$, smooth kernel g; separate densities for core and valence electrons.

Solid harmonic wavelet basis functions (Brumwell et al.): $\psi^m_{\gamma,l,j}$; convolutions with densities ρ.

Now symmetrize with respect to rotations and translations ...

Finally multilinear regression.

Spectral Neighbor Analysis Potential (SNAP)

Key references: Thompson, Swiler, Trott, Foiles, Tucker (2015) J Comput Phys 285; Wood, Thompson (2018) J Chem Phys 148.

Local environment of atom i, expansion into hyperspherical harmonics, expansion coefficients $u^j_{m,m'}$.

Bispectrum components are defined that are real-valued and invariant under 3D rotation: triple products $B_{j_1,j_2,j}$ of the $u^j_{m,m'}$. (It involves a hyperspherical version of Clebsch-Gordon coefficients.)

Finally local energy E_i is expressed as a linear (Thompson et al, 2015) or quadratic (Wood and Thompson, 2018) function of the bispectrum components $B_{j_1,j_2,j}$. (Linear regression either way.)

Atomic cluster expansion by R. Drautz, Bochum

Key reference: Drautz (2019) Phys Rev B 99.

Local environment of atom i, descriptors $A_{iv} = \sum_j \phi_v(\vec{r}_{ji})$, where the ϕ_v are a family of basis functions: $v = (nlm)$ and then

$$\phi_v(\vec{r}) = \sqrt{4\pi} R_{nl}(\|\vec{r}\|) Y_l^m(\vec{r}/\|\vec{r}\|).$$

Cluster products involving Clebsch-Gordon coefficients

$$B^{(K)}_{i,\vec{n},\vec{l}} = \sum_{\vec{m}} \mathrm{CG}(\vec{l},\vec{m}) \times A_{i,n_1,l_1,m_1} \cdots A_{i,n_K,l_K,m_K}.$$

Finally $E_i = \sum_{K,\vec{n},\vec{l}} c^{(K)}_{\vec{n},\vec{l}} B^{(K)}_{i,\vec{n},\vec{l}}$

Drautz (2019) also describes a nonlinear version to overcome slow convergence of the cluster expansion.

Problems with potentials for materials and PMI

Supplementary slides.

Hydrogen retention in irradiated tungsten

IAEA Coordinated Research Project (CRP) on Plasma-wall interaction with irradiated tungsten and tungsten alloys in fusion devices (2013-2018). See https://www-amdis.iaea.org/CRP/.

Need to understand effect of radiation om microstructure and effect of microstructure on hydrogen retention and migration.

Must use surrogate irradiation; need modelling to interpret experimental data.

Most basic computations: primary radiation damage and hydrogen migration.

Relatively short timescale. (Long timescale: segregation, corrosion.) Molecular dynamics is the main tool.

Problems with potentials for tungsten

Talk by A. Sand (Helsinki, with Kai Nordlund) at IAEA, 2017-11-16: "Energetic cascades in tungsten: sensitivity to interatomic potentials and electronic effects."

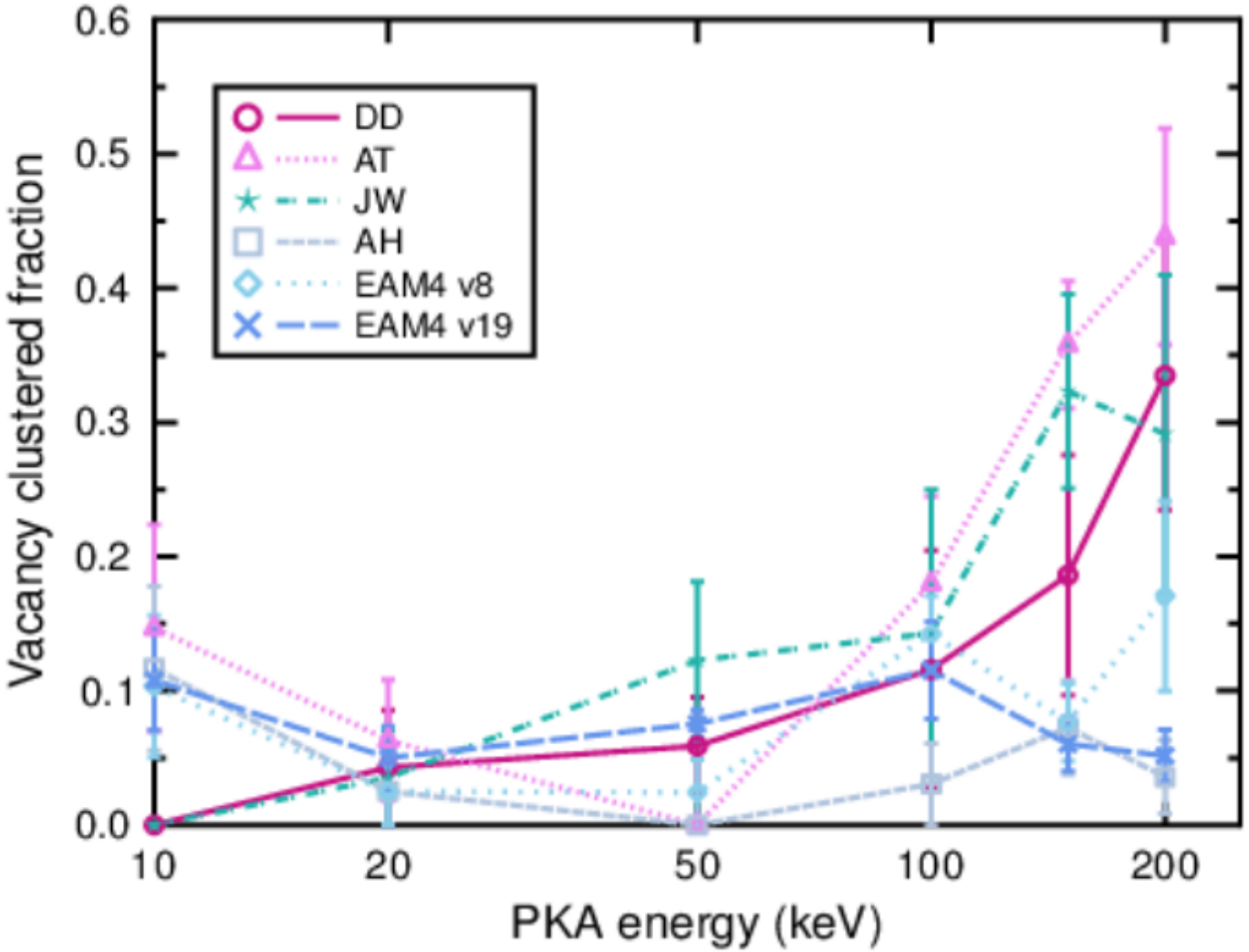
"*Potentials with largely similar point defect formation and migration energies disagree regarding clustered fraction of defects for high PKA energies. Some potentials predict only very small clusters, others show formation of clusters of* > 100 *point defects.*"

"*Why the different predictions, despite extensively fitted 'good' potentials??*"

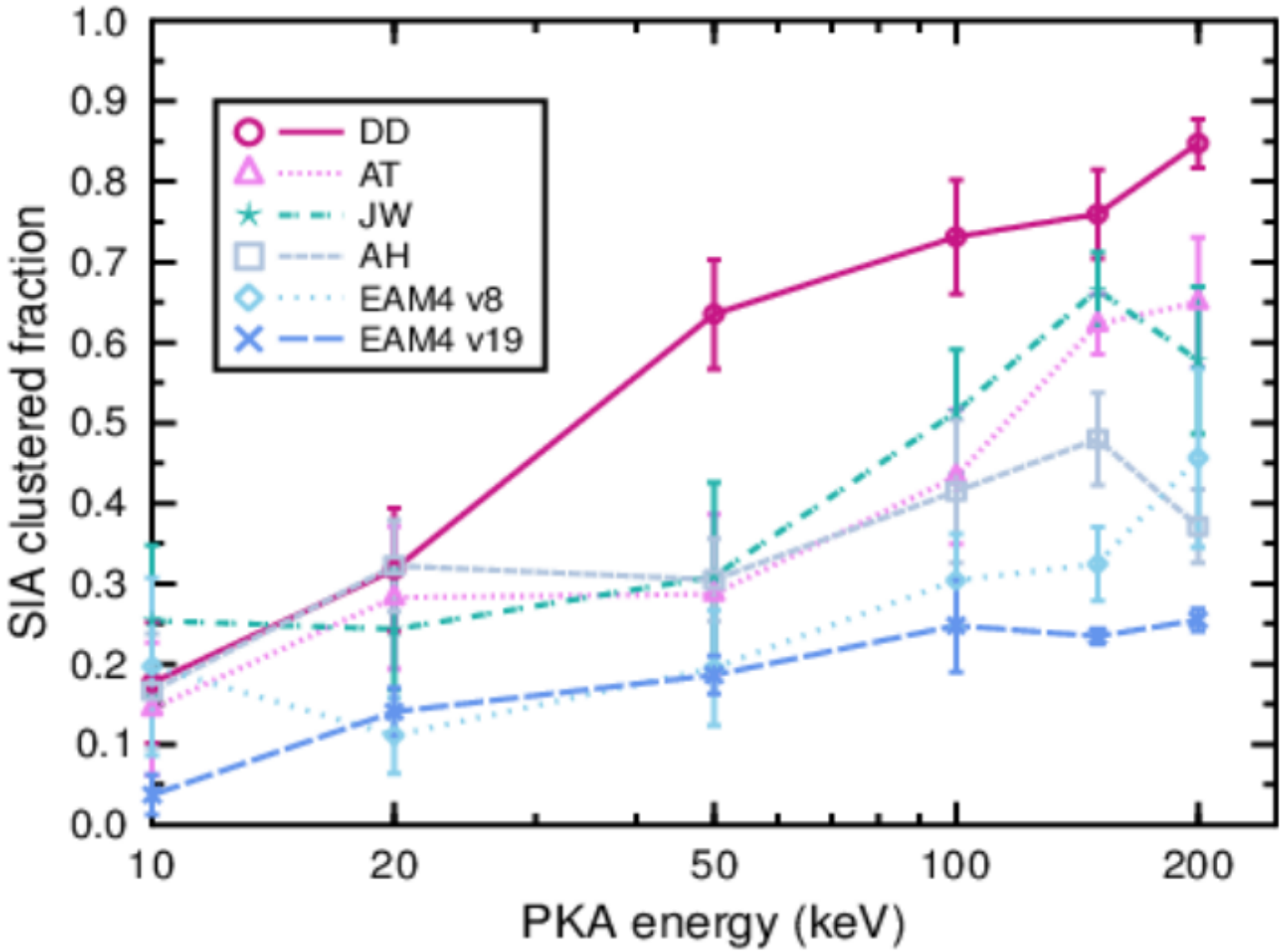
Discuss blending to short-range Ziegler-Biersack-Littmark (ZBL) potential.

Many-body effects beyond embedded atom (EAM) approach.

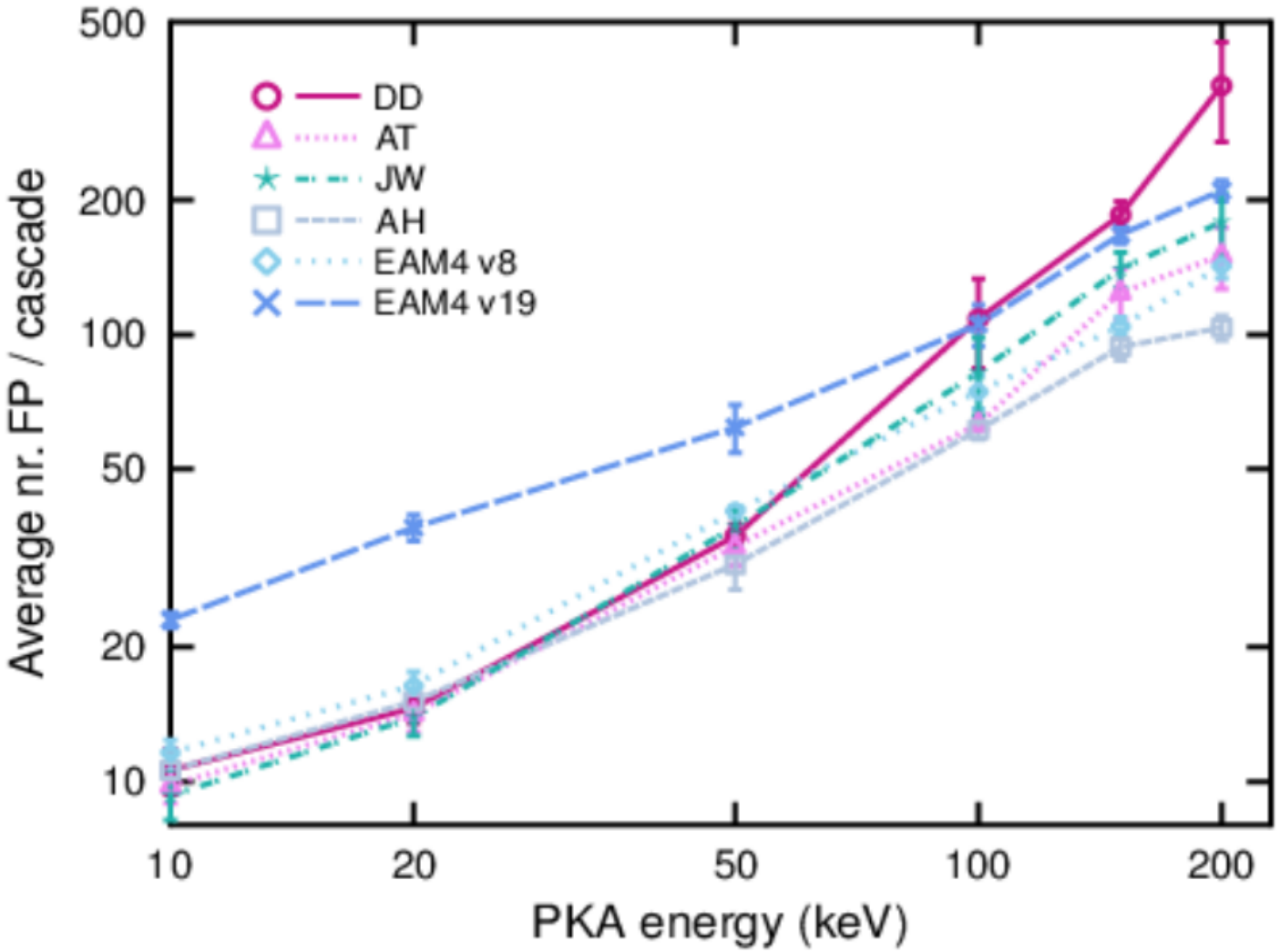
Vacancy Clustered Fraction



Self-Interstitial Atom Clustered Fraction



Average Number Frenkel Pairs per Cascade



Force fields for fusion materials and plasma-material interaction

The potential is (almost) everything; and that needs to be reflected in the effort.

Keep in mind the following target application: Primary radiation damage in W-H-He. PKA event, melt region, resolidification. More difficult than pure W (see above); not as difficult as steel.

Quantum effects on the nuclear motion: barely ever relevant.

Electronic excitation beyond simple stopping: can be important, could be taken into account. (Langevin approach, potential depends on electron temperature.)