Uncertainties in classical effective potentials: sources and quantification strategies

Peter Brommer

Warwick Centre for Predictive Modelling
University of Warwick

WCPM Seminar
14 May 2015

Motivation

Quantum theory of electrons and nuclei

QM (QED) is a theory with enormous predictive power:

- Energy levels of hydrogen atom to a few ppm.

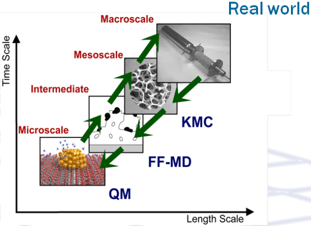
Solution to all our modelling needs? . . . No.

"Sequential" Multiscale Modelling

Use proper simulation tool for each scale.

- Parameterise from small to large.
- No direct coupling between models.

Propagate uncertainty from one scale to the next.



essenceofescience.se/nobel-2013/

Outline

1. Standard Sequential QM/MM Multiscale Modelling
    - Density Functional Theory
    - Molecular Dynamics
    - Effective Potentials
    - Force Matching
2. Uncertainty
    - Sources
    - Quantification
    - Examples

(Plane Wave) Density Functional Theory

A sequence of approximations

Solve Schrödinger (Dirac) equation of electrons and nuclei

- for stationary nuclei (Born-Oppenheimer approximation),
- mapping the many-electron problem to many one-body problems,
- which use approximative functionals to represent XC,
- while core electrons are treated by pseudopotentials;
- wave functions are represented using plane wave basis set,
- cut off at finite energy and sampled on a finite grid;
- the problem is then solved by iteration to self-consistency.

Depending on some of the choices, further corrections are necessary.

Uncertainty and errors?

Not all errors are controllable.

- Would merit a talk of its own.

Molecular Dynamics

What is Molecular Dynamics?

Equations of motion of a system of interacting particles are integrated numerically.

- Direct simulation of the basic laws of physics: Newton’s (or Hamilton’s) equations.

Needed

- Initial condition: structure model
- Equation of motion: model of the interactions

Big systems or long simulation times are feasible only with classical effective potentials.

Effective Potentials

Eliminate electronic degrees of freedom

Re-write equation of motion

$$\frac{\partial \boldsymbol{P}_I}{\partial t} = -\nabla_I(\epsilon_0(\boldsymbol{R}) + V_{NN}(\boldsymbol{R})) = \nabla_I(\tilde{V}_{\text{eff}}(\boldsymbol{R}))$$

Usually, V_{eff} much simpler. Expand in manybody contributions:

$$V_{\text{eff}}(\boldsymbol{R}) = \sum_i \phi_1(\boldsymbol{r}_i) + \frac{1}{2}\sum_{\substack{i,j \\ i\neq j}} \phi_2(\boldsymbol{r}_i, \boldsymbol{r}_j) + \frac{1}{6}\sum_{\substack{i,j,k \\ i\neq j\neq k\neq i}} \phi_3(\boldsymbol{r}_i, \boldsymbol{r}_j, \boldsymbol{r}_k) + \dots$$

Even simpler: Central (homogeneous and isotropic) pair potentials:

$$V(\boldsymbol{R}) = \frac{1}{2}\sum_{\substack{i,j \\ i\neq j}} \phi_{ij}(r_{ij}), \qquad \text{with } r_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$$

Interactions

Choice of interaction model depends on material to simulate

(different ways to truncate manybody expansion)

- Central pair potentials, EAM potentials for metals.
- Angular dependent potential (ADP), MEAM.
- Covalent potentials (Tersoff, Stillinger-Weber, . . .) for ceramics.
- Coulomb potential (Ewald method, Wolf summation).
- Dipolar interaction for oxides.
- Simulation of organic molecules: Force fields for polymer chains, water, amino acids,. . .

Potentials specified

- by values at sampling points (tabulated potentials)
- or by parameters (analytic potentials).

Example: EAM potentials

EAM potentials

$$E = \sum_{ij} \Phi_{ij}(r_{ij}) + \sum_i F_i(\rho_i), \quad \text{where} \quad \rho_i = \sum_j \Psi_j(r_{ij})$$

Bond strength depends on environment –
better suited to describe vacancies and other defects.

Analytic vs. tabulated potentials

- Φ, F, ρ are the potential functions.
- Can be represented by analytic functions, e.g. $F = c\sqrt{\rho}$.
- Alternative: Tabulated at sampling points.

Potential Development

How to obtain effective potentials?

Potential serves to determine energies and forces
→ determines the physics of the system!

- Depending on the system (metal, oxide, etc.), a suitable potential type must be chosen.
- Within such a potential family, the potential parameters determine the physical properties a particular material.
- The parameters are chosen such that the desired material properties are correctly reproduced.
- The material properties to be reproduced are often computed ab-initio, instead of measured experimentally.

Reference quantities

Experimental Quantities of Interest

e.g. phase diagrams, elastic properties, diffusivities. . .

- often expensive to calculate ⇒ “simple” systems only.

Literature and first-principles quantities

e.g. bond lengths, heats of formation, bond angles.

- Sampling configurational space?

Example: ReaxFF.

First-principles forces, stresses, energies

- easily calculated (1 MD step).
- sample configurational space.

Force Matching!

Ercolessi & Adams, Europhys. Lett. **26**, 583 (1994)

Force Matching with *potfit*

Open source force matching code ***potfit***

- Flexible and modular.
- Supports pair, (M)EAM, ADP potentials (metals).
- Oxide potentials.
- Electron-temperature dependent potentials (laser ablation).
- Interfaces to DFT and MD codes.

Widely used code

- 40 downloads/month,
- 50 citations with potentials,
- from more than ten distinct groups around the globe.

Brommer, Gähler, Model. Simul. Mater. Sci. Eng. **15**, 295 (2007).

`http://potfit.sourceforge.net/`

Potential Generation

1. Select potential model, starting potential.
2. Select reference structures (100–200 atoms, MD simulation at various temperatures, strained structures).
3. Calculate forces, stresses, energies with ab-initio code.
4. Optimize starting potential with *potfit*.
5. Generate reference structures with new potential.
   $\Rightarrow$ more realistic configurations.
6. Test potential.

If results are not satisfying

- use more/different reference configurations,
- replace insufficient potential model.

and iterate procedure.

Uncertainties

Sources of uncertainty for force-matched potentials

Generic errors:

- “Imported” uncertainty: cannot beat DFT.
- Algorithmic uncertainty: global optimum?

Force Matching specific (structural & parameter) uncertainties:

- Bad reference data selection (parameter uncertainty).
- Wrong functional form (model bias).
- Overfitting (parameter uncertainty).
- Wrong potential model (model bias).

Properties of force-matched potentials:

- (Generally) good representability.
- Limited transferability.

Caveat emptor! (US$2M NSF CDI grant, cf. https://openkim.org/).

Interpolated potentials

Special case: tabulated or interpolated potential

Interpolated potentials can have many parameters (>100).

- No bias from particular functional form.
- Parameters have no meaning.

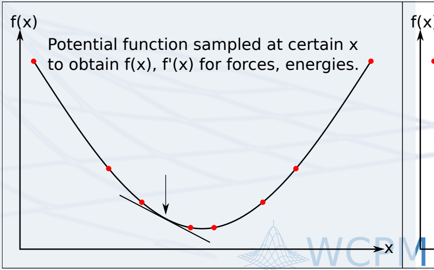
No formal uncertainty propagation in literature.

Confidence of sampling point values

Forces: evaluate potential functions and gradients.

- Both in training and use.

Training set and application: sample similarly.



Potential function sampled at certain x to obtain f(x), f'(x) for forces, energies.

Potential derivation as an inverse problem

Forward problem: data (e.g. forces) δ from parameters u

$$\delta = G(u)$$

- u: Parameter vector
- δ: Result vector

Well-posed problem.

Inverse problem: parameters u from noisy data δ

$$\delta = G(u) + \eta$$

Ill-posed. Optimisation problem:

$$\min_u \frac{1}{2}||\delta - G(u)||^2$$

Potential derivation as a Bayesian inverse problem

Interpret u, η, δ as RV/random fields

Attack problem using “standard” techniques:

- MCMC methods
- MAP estimators
- Sparse deterministic approximations

⇒ Account naturally for noisy reference data.

Computationally involved (to say the least).

In the following:

- How has this been used so far (selection)?

The problem is two-fold:

1. Quantifying uncertainty in potential (inverse problem).
2. Propagate uncertainty to MD simulations (forward problem).

“Sloppy model” MD

Replace best-fit effective potential by ensemble

Conditional probability of parameter set u:

$$P(u|\delta, G) \propto \exp\left[-\frac{C(u)}{T}\right],$$

- cost function $C(u) = \frac{1}{2}||\delta - G(u)||^2$
- formal temperature T, with $T_0 = 2C_0/N_p$, $C_0 = \min_u C(u)$

For any observable O, calculate mean $\langle O\rangle|_{T,D,M}$, variance $\sigma_O^2|_{T,D,M}$.

Frederiksen *et al.*, PRL **93** 165501 (2004)

Sloppy model pros and cons

Advantages

- Integrates seamlessly into force matching/MD software stack:
  - Generate ensemble of potentials.
  - Run N independent MD simulations.
- Accounts for model & parameter errors.
- Computational cost: factor N.

Disadvantages

- No error propagation from DFT (DFT=truth).
- Sampling of parameter space can be tricky (anisotropic).

Open PhD position: Contact PB!

Inference of force field parameters

Bayesian formulation

$$\pi(u|\delta) = \frac{1}{C} p(\delta|u) q(u),$$

- $q(u)$: prior on the parameters.
- $\pi(u|\delta)$: posterior density.
- $p(\delta|u)$: likelihood, contains diffence data/predictions.

Direct Bayesian approach prohibitive,

if p expensive (e.g. full MD sim).

- Use surrogate models
  - non-intrusive spectral projection (NISP)
  - nondeterministic PC expansion

Rizzi *et al.*, Multiscale Modeling & Simulation **10**, 1460–1492 (2012)

Rizzi's method pros and cons

Demonstration on model system

TIP4P (water model, 4 parameters).

- Recover 3 parameter values (4th parameter assumed known)...
- ...from 30 "measurements" of density, self-diffusion, enthalpy.
- Model errors not tested (work underway).

Open questions

- How does it scale in parameters, observables?
- Use in force matching context?

UQ+P HPC framework Π4U

The framework

For model class *M*:

$$\pi(u|\delta, M) = \frac{1}{C} p(\delta|u, M) q(u|M),$$

- Transitional MCMC algorithm.
- Surrogate models to reduce computational cost.

Angelikopoulos *et al.*, J. Chem. Phys. **137**, 144103 (2012)
Hadjidoukas *et al.*, J. Comp. Phys. **284** 1–21 (2015)

Π4U pros and cons

Demonstration system

Argon, LJ potential (2 parameters).

- Analytic expressions available for many QoI.
- Functionality also works with full MD.
- Model error: additional parameter.

Computational cost

Time to solution: 3 days (48 nodes × 16 cores).

Conclusions

Force Matching

Extending atomistic simulations to new materials:

- Preserve DFT precision to larger systems, longer times.
- Foundation for other atomistic and meso-scale problems.

Essential part of multi-scale modelling stack.

Uncertainty

Work on UQ for effective potentials has just started.

- UQ in potential derivation: Bayesian inverse problem.
- UQ in MD simulations: Re-use old tools or shape new ones?

New modelling paradigm still needs work.