CP2K Developers Meeting

March 8th, 2024 14:00-16:00 CET
(see also https://www.cp2k.org/dev:meetings)

CP2K Developers Meeting

1. Current Development Efforts (all)
2. On-the-fly potentials (Martin Brehm)
3. GW and BSE for excitation energies (Max Graml)
4. Updates (Jan Wilhelm)
5. …
6. CP2K-HFX FPGA Update (PC2)
7. CP2K@CASUS (Frederick Stein)
8. Current Issues when running CP2K (all)
9. Feature Deprecation (all)
10. CP2K Release (all)
11. Open CP2K-Related Positions (all)
12. CP2K-related Events (all)

On-the-fly MLFF (Martin Brehm)

What are you currently working on or planning to work on?

- Jürg Hutter: there is overlap with a planned project
- Thomas Kühne: i-pi protocol might be an alternative integration option
- Ole: pytorch interface

Machine-Learning Force Fields

- Have become very popular recently *(e.g., Behler‘s NNPs)*
- Towards the vision *„ab initio accuracy at force field speed“*

However:

- Training process is not straight-forward *(coverage of full configuration space?)* → Requires an expert
- Often, configurations outside of the high-confidence space are encountered during production run…

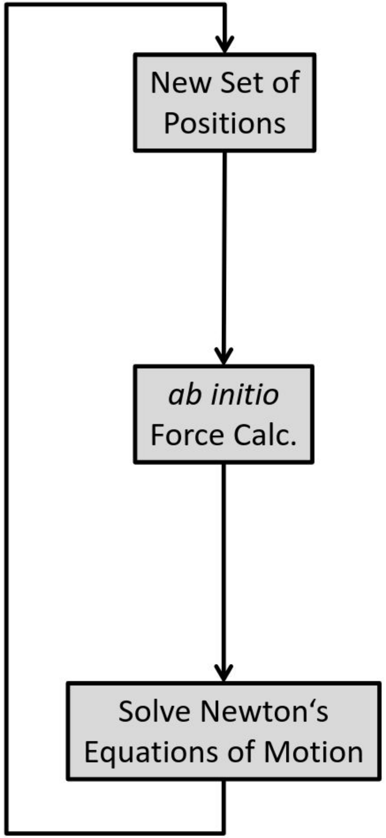
→ Stop production run, re-training, resume production run

Conclusion: If …

- … you are a „standard AIMD user“ without ML background,
- … you want to run a single long trajectory,
- … there is no trained MLFF available for your system,

…, today‘s MLFFs are **not an option** for you :-/

Standard AIMD

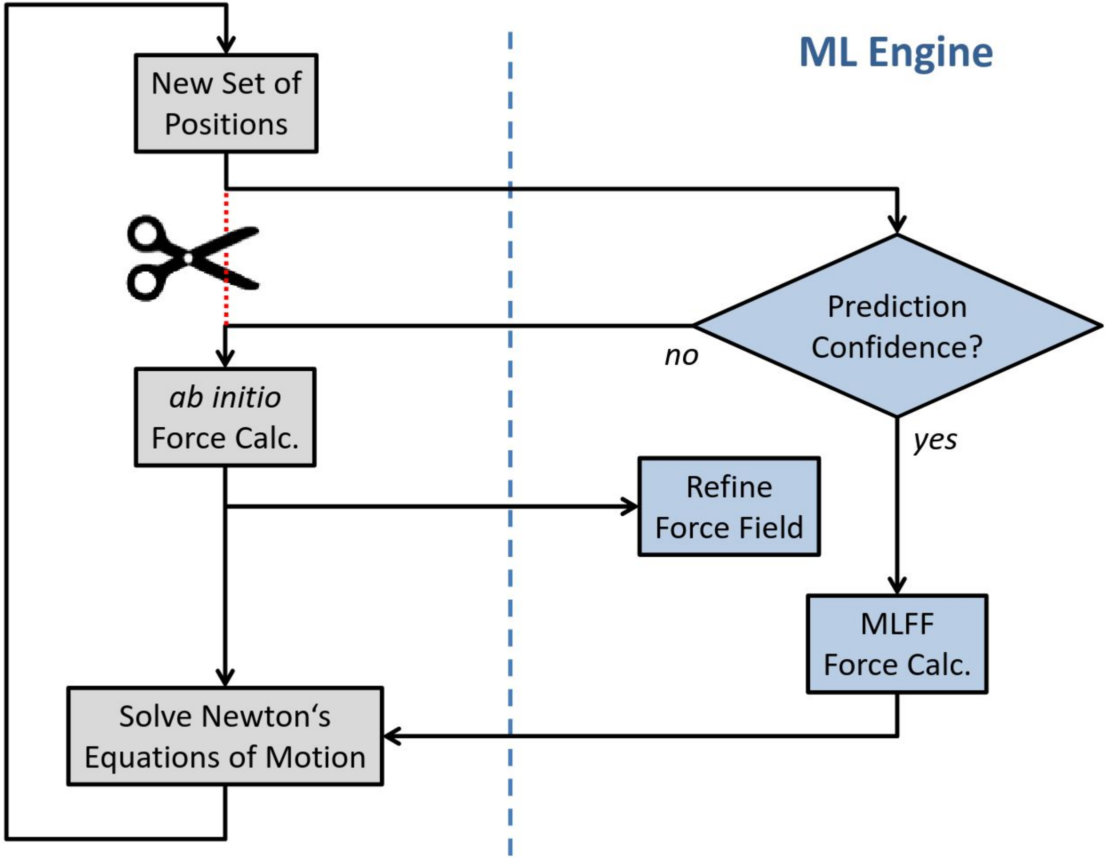


New Set of Positions

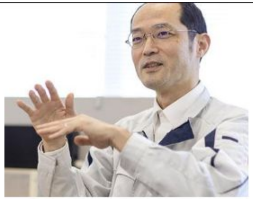
ab initio Force Calc.

Solve Newton‘s Equations of Motion

On-the-Fly Machine-Learning Force Field



Implementation in VASP



2019: Ryosuke Jinnouchi, while Postdoc in Wien, implements such an approach in VASP:

„On-the-fly machine learning force field generation: Application to melting points",
Phys. Rev. B **2019**, *100*, 014105, DOI 10.1103/PhysRevB.100.014105

- It is **not** based on neural networks – uses Bayesian Inference ML
- It is a **true black box** *(can be applied to any AIMD, no parameters need to be tuned, no expert needed)*
- After a few hundred AIMD steps, already 90% of steps via MLFF.
  → can easily save a factor of > 10 in computer time
- This is **not** an empty marketing promise *(a few colleagues of mine already use it for production)*

I find that's a really big thing.

I personally don't like VASP so much (commercial),
*see it as a part of my mission to convince scientists to use **free software**...*

Introducing Prokyon

I am planning to develop a C++ library **„Prokyon“**
for on-the-fly MLFF applications *(will be either GPL or L-GPL license)*.

- Will contain several models; the first one will be the Bayesian Inference ML as implemented in VASP
- Can be interfaced to any AIMD code in the future, but the **prime target will be CP2k**

Two planned modes of operation:

a) CP2k drives the AIMD. Prokyon is invoked via `MULTIPLE_FORCE_ENV` as a second `FORCE_ENV`

b) Prokyon drives the AIMD. CP2k is invoked via `libcp2k` *(as Gromacs QM/MM does)*

- Interfacing Fortran ↔ C ↔ C++ works well *(experience from `libvori`)*
- Technical details are currently being discussed *(parallelization)*
- Expect an Alpha version *(hopefully)* in summer…

GW and BSE for excitation energies (Max Graml)

What are you currently working on or planning to work on?

Bethe Salpeter equation for computing electronic excitations

Bethe Salpeter equation

$$\begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B} & \mathrm{A} \end{pmatrix}\begin{pmatrix}\mathbf{X}^{(n)}\\ \mathbf{Y}^{(n)}\end{pmatrix} = \Omega^{(n)}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\mathbf{X}^{(n)}\\ \mathbf{Y}^{(n)}\end{pmatrix}$$

$$\mathrm{A}_{\imath a,jb} = \left(\epsilon_a^{\mathrm{G_0W_0}} - \epsilon_\imath^{\mathrm{G_0W_0}}\right)\delta_{\imath j}\delta_{ab} + \alpha^{S/T} v_{\imath a,jb} - \mathrm{W}_{\imath j,ab}(\omega = 0)$$

$$\mathrm{B}_{\imath ajb} = \alpha^{S/T} v_{\imath a,bj} - \mathrm{W}_{\imath b,aj}(\omega = 0)$$

Hermitian equation

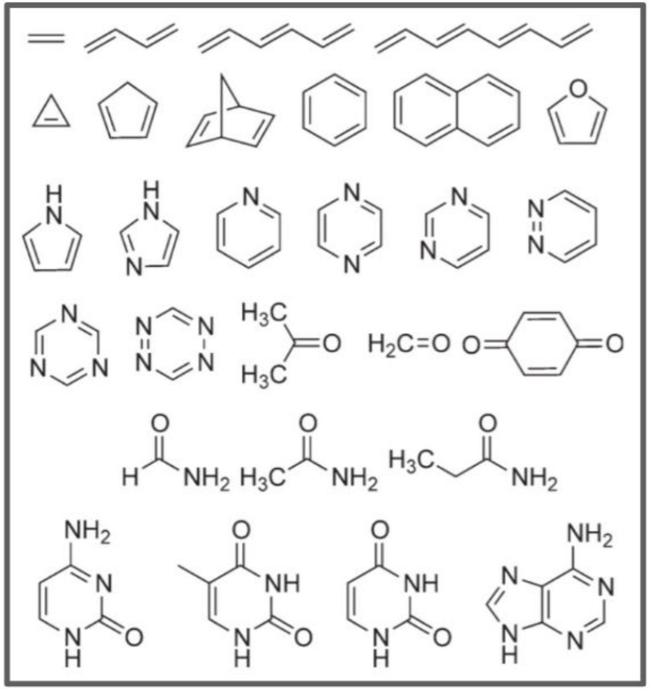
$$\mathrm{C}\,\mathbf{Z}^{(n)} = {\Omega^{(n)}}^2\,\mathbf{Z}^{(n)}$$
$$\mathrm{C} = (\mathrm{A}-\mathrm{B})^{0.5}\,(\mathrm{A}+\mathrm{B})(\mathrm{A}-\mathrm{B})^{0.5}$$

Tamm Dancoff approximation

$$\mathrm{A}\,\mathbf{X}_{\mathrm{TDA}}^{(n)} = \Omega_{\mathrm{TDA}}^{(n)}\,\mathbf{X}_{\mathrm{TDA}}^{(n)}$$

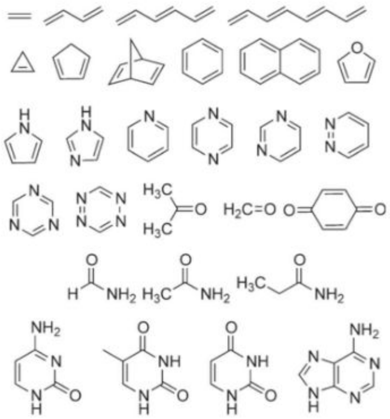
Canonical computational cost: $O(N^6)$

Benchmark on Thiel's set



Thiel's set: J. Chem. Phys. 128, 134110 (2008)

Bethe Salpeter equation for computing electronic excitations



Bethe Salpeter equation

$$\begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B} & \mathrm{A} \end{pmatrix}\begin{pmatrix}\mathbf{X}^{(n)}\\ \mathbf{Y}^{(n)}\end{pmatrix} = \Omega^{(n)}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\mathbf{X}^{(n)}\\ \mathbf{Y}^{(n)}\end{pmatrix}$$

$$\mathrm{A}_{ia,jb} = \left(\epsilon_a^{\mathrm{G_0W_0}} - \epsilon_i^{\mathrm{G_0W_0}}\right)\delta_{ij}\delta_{ab} + \alpha^{S/T} v_{ia,jb} - \mathrm{W}_{ij,ab}(\omega = 0)$$

$$\mathrm{B}_{iajb} = \alpha^{S/T} v_{ia,bj} - \mathrm{W}_{ib,aj}(\omega = 0)$$

Hermitian solution

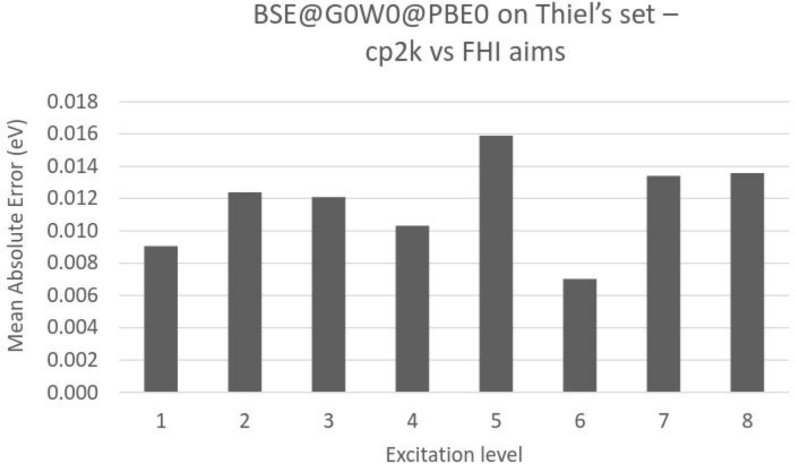
$$\mathrm{C}\,\mathbf{Z}^{(n)} = {\Omega^{(n)}}^2\,\mathbf{Z}^{(n)}$$
$$\mathrm{C} = (\mathrm{A}-\mathrm{B})^{0.5}\,(\mathrm{A}+\mathrm{B})(\mathrm{A}-\mathrm{B})^{0.5}$$

Tamm Dancoff approximation

$$\mathrm{A}\,\mathbf{X}_{\mathrm{TDA}}^{(n)} = \Omega_{\mathrm{TDA}}^{(n)}\,\mathbf{X}_{\mathrm{TDA}}^{(n)}$$

Canonical computational cost: $O(N^6)$

Typical excitation energy:
3-5 eV

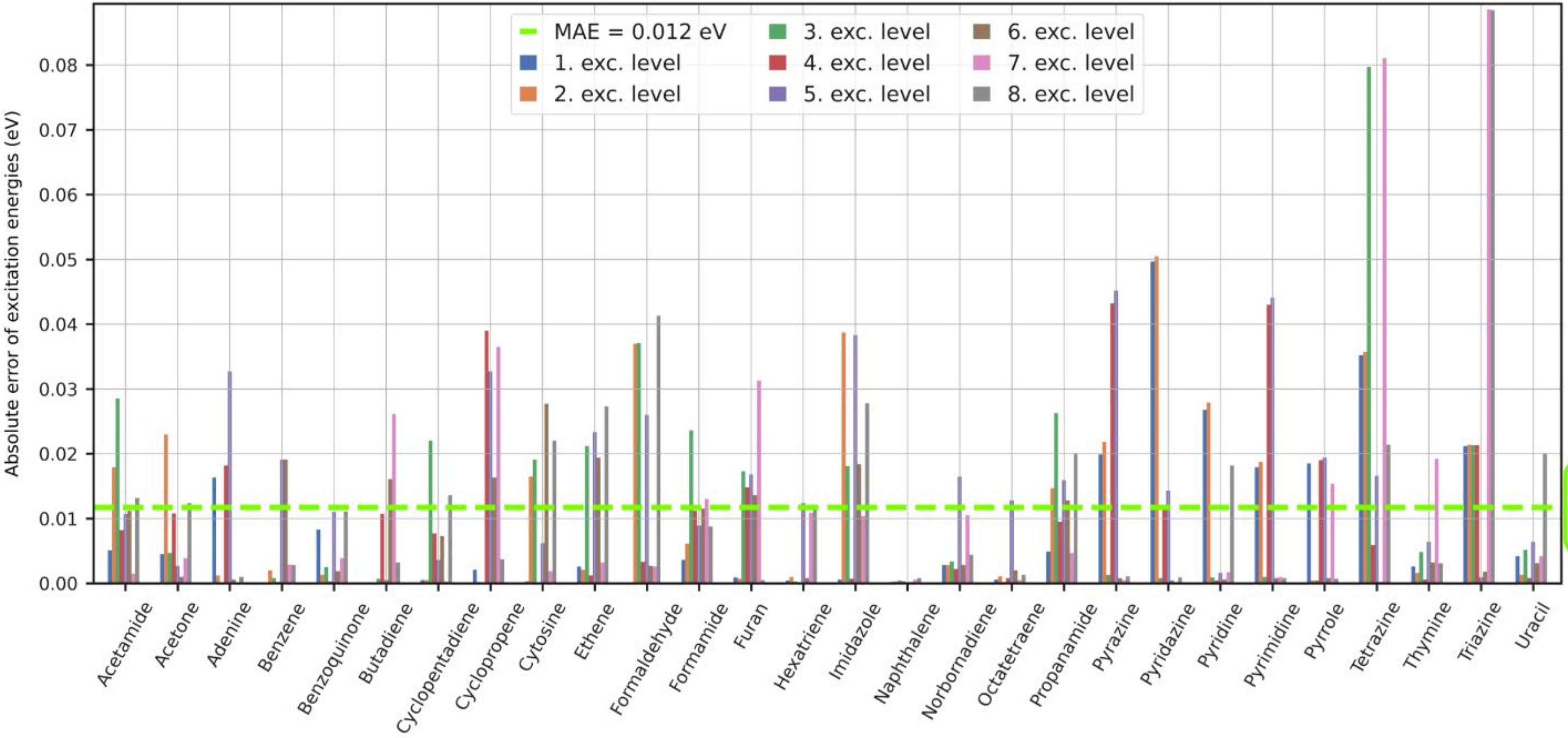

BSE@G0W0@PBE0 on Thiel's set – cp2k vs FHI aims

Thiel's set: J. Chem. Phys. 128, 134110 (2008)

Bethe Salpeter equation for computing electronic excitations

Typical excitation energy:
3-5 eV

BSE@G0W0@PBE0 on Thiel's set – cp2k vs FHI aims



MAE:
0.012 eV

Thiel's set: J. Chem. Phys. 128, 134110 (2008)

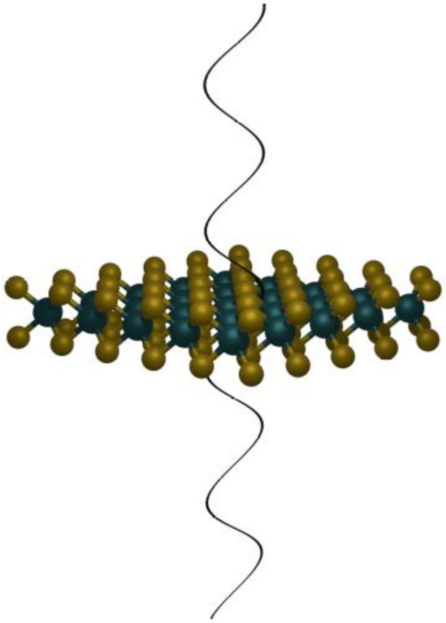
Updates (Jan Wilhelm)

What are you currently working on or planning to work on?

- GW on 2D materials is factor 10 000 to 100 000 faster compared to plane-wave GW

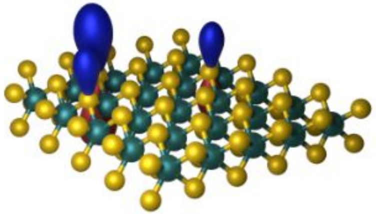
***GW* with plane-wave basis set**

$$\chi_{\mathbf{GG}'} = \langle e^{i\mathbf{Gr}} | \chi(\mathbf{r},\mathbf{r}') | e^{i\mathbf{G}'\mathbf{r}'} \rangle$$

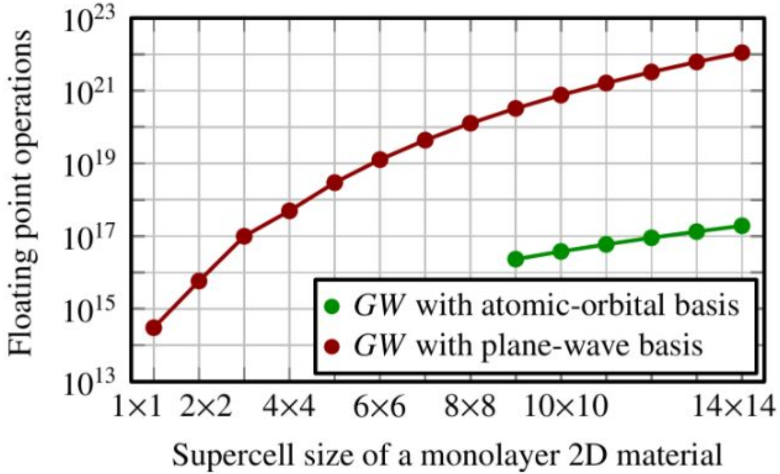


***GW* with atomic-orbital basis set**

$$\chi_{PQ} = \langle \varphi_P(\mathbf{r}) | \chi(\mathbf{r},\mathbf{r}') | \varphi_Q(\mathbf{r}') \rangle$$



https://doi.org/10.1021/acs.jctc.3c01230



- Memory efficient iteration of 3-center integrals, extreme example: 11x11 cell of MoS_2 (363 atoms, 2D material), TZVP-MOLOPT, **1 node** on Noctua (1024 GB RAM), 204 hours

$$\Sigma_{\lambda\sigma}(i\tau) = \sum_{\text{atom } A} \sum_{\text{atom } B} \sum_{\nu \text{ (at atom } A)} \sum_{Q \text{ (at atom } B)} \left[\sum_{\mu} (\lambda\mu|Q)\, G_{\mu\nu}(i\tau)\right] \left[\sum_{P} (\nu\sigma|P)\, W_{PQ}(i\tau)\right]$$

code in src/gw_methods ; input:

```
&PROPERTIES
  &BANDSTRUCTURE
    &GW
      NUM_TIME_FREQ_POINTS    10
      MEMORY_PER_PROC    15     ! Used 64 MPI processors -> memory per process: 1024 GB/64 = 16 GB
    &END
  &END
&END
```

Comparison computation time:

„old“: store 3c integrals: 7488 core hours

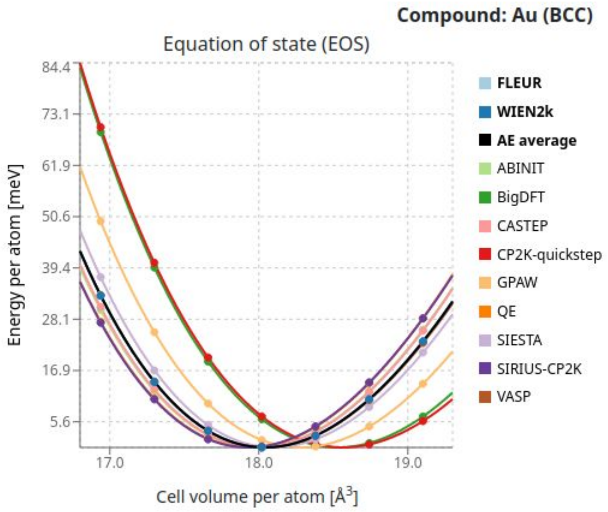
„new“: recalc 3c integrals: 26122 core hours
(recalculation of 3c: 9600 core hours)

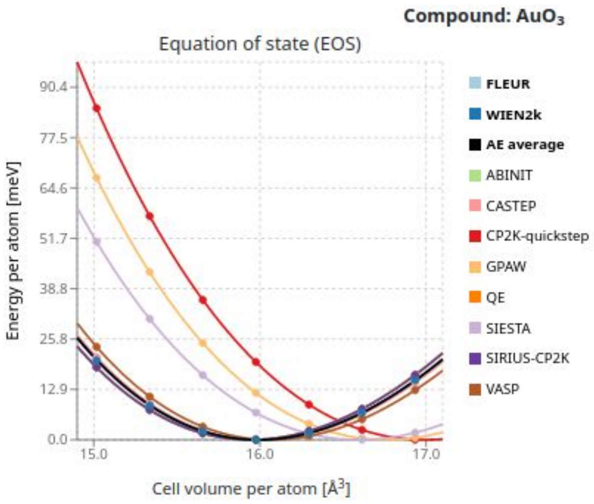
- Optimization of CP2K on Supermuc-NG Phase 2 (Intel-GPU Ponte Vecchio, Hans Pabst)

- Any plans for SCF with spinors and SOC from GTH pseudos?

  (KS-matrix will be complex with SOC and double in size, SCF with SOC will be important for calculation of forces with SOC)

Outliers in verification paper, e.g. for gold.

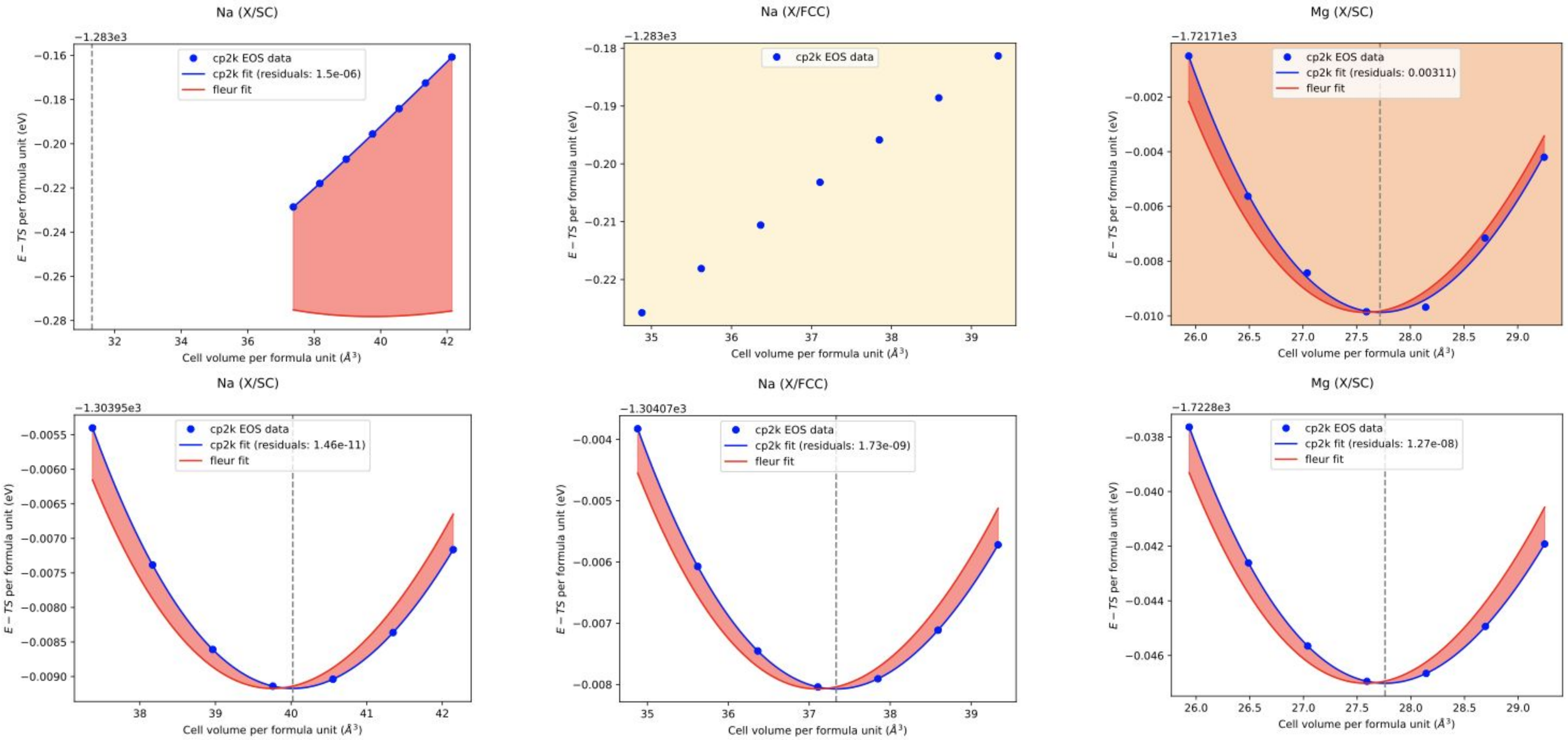




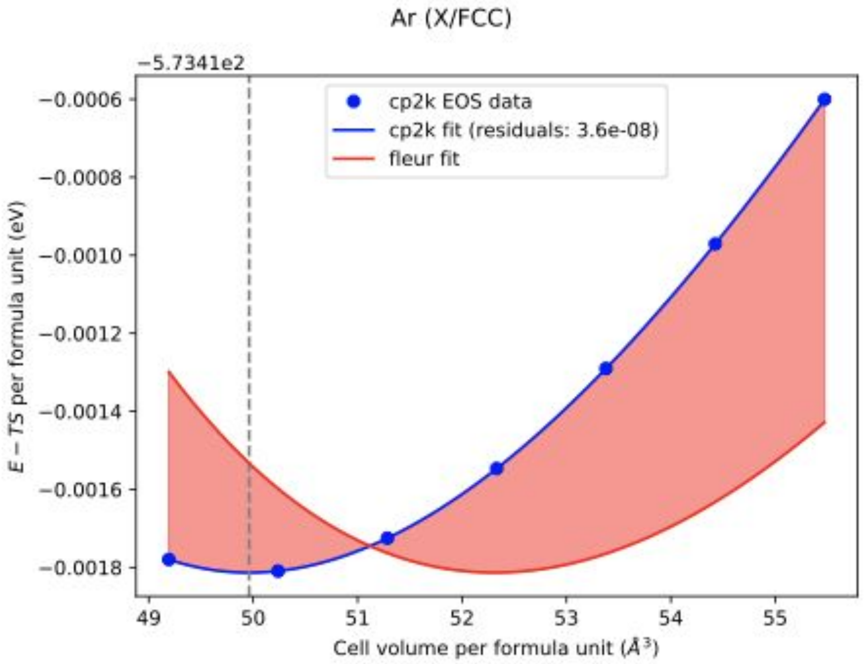
https://www.nature.com/articles/s42254-023-00655-3

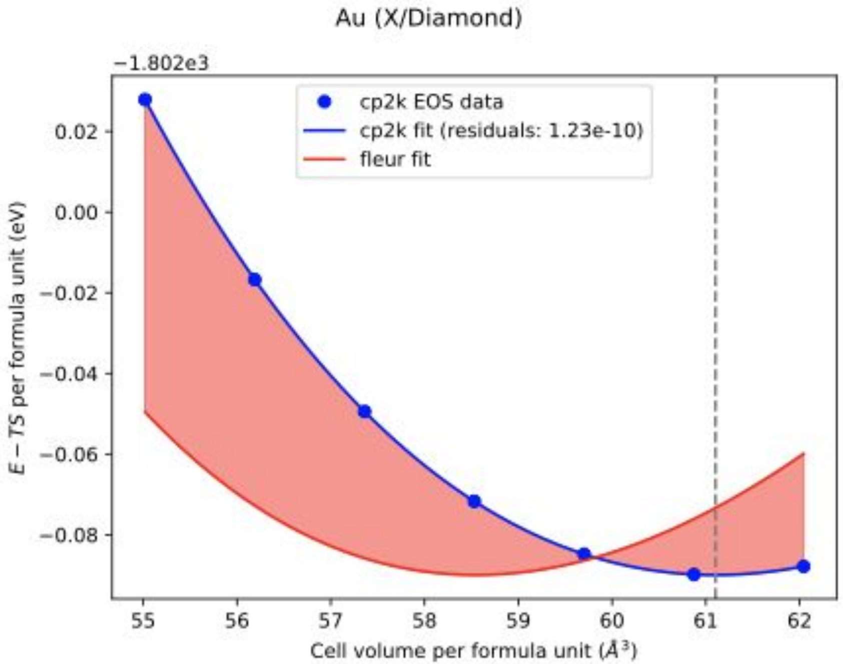
https://acwf-verification.materialscloud.org

Outliers in verification paper: UZH Protocol



Outliers in verification paper: UZH Protocol





Similarly: Ba, Kr, (Ne), Pd, Rh, Rn, Ru, (Xe)

CP2K HFX with accelerators (PC2)

Electron repulsion integral engines:

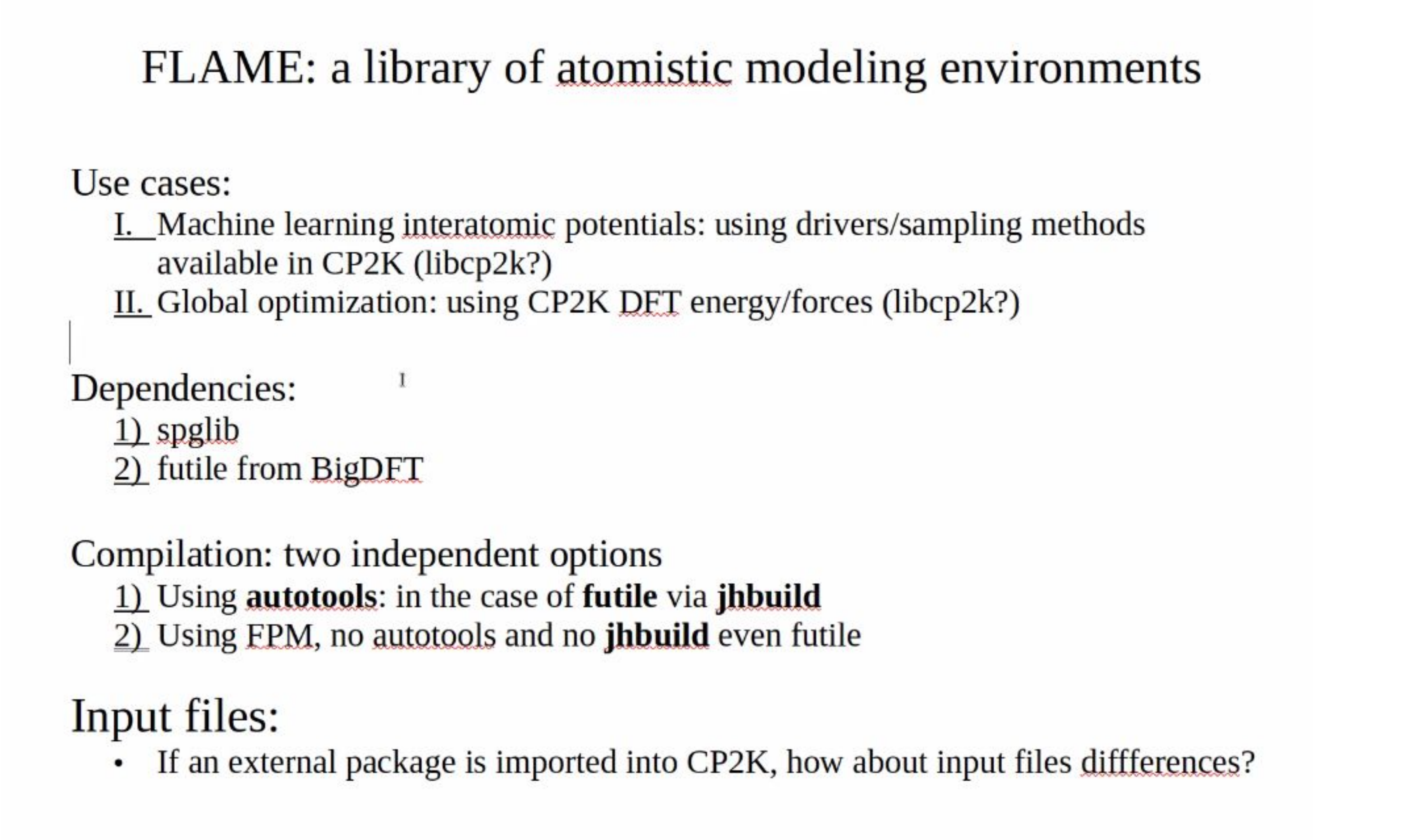
- Intel FPGA (Xin Wu, Tobias Kenter)
- Xilinx FPGA (Zhenman Fang) and Xilinx AI Cores (Johannes Menzel)
- Nvidia GPUs (Marcello Puligheddu)
- AMD GPUs (just started)

Ongoing: integration into CP2K (rewrite of HFX and load balancing)

Current Development Efforts

What are you currently working on or planning to work on?

- Importing a code for machine learning potential into CP2K (Alireza Ghasemi)
  - Possible solutions?
    - Interface as a library?
    - Experience with previous case imported into CP2K?
  - Compilation?
  - Input files?



FLAME: a library of atomistic modeling environments

Use cases:

I. Machine learning interatomic potentials: using drivers/sampling methods available in CP2K (libcp2k?)

II. Global optimization: using CP2K DFT energy/forces (libcp2k?)

Dependencies:

1) spglib
2) futile from BigDFT

Compilation: two independent options

1) Using **autotools**: in the case of **futile** via **jhbuild**
2) Using FPM, no autotools and no **jhbuild** even futile

Input files:

- If an external package is imported into CP2K, how about input files diffferences?

CP2K@CASUS

- Job offering as CP2K developer at CASUS
  - deadline was in March 6
  - still pending interview
- FFTW3+MPI:
  - prepare PR for some refactoring
  - in progress: switch to FFTW3 blocking scheme
- Performance Engineering (Andreas Knüpfer)
- Finite Temperature RPA

manual.cp2k.org

Input Reference

- Input descriptions support Latex-formulas, Markdown, and unicode incl. emoji.
- Special treatment for XC_SECTIONS.
- Highlighting of mentioned keywords ⭐.
- New precommit check for missing spaces in multi-line descriptions.

Methods Section

- Curated structure, see e.g. the section on X-Ray or HFX.
- Live preview for pull requests on GitHub.
- Most howtos have been moved over from Wiki to Github.
- Still missing:
  - Older topics, e.g. Metadynamics
  - Overview pages that, e.g. Optical Spectroscopy

Current Issues when Running CP2K

-

CP2K-Release

Open CP2K-Related Positions

- position in Jürg Hutters group
- position at CASUS: see slide CP2K@CASUS

CP2K-Related Events:

Plans:

- Paderborn+CASUS/HZDR:
  - 3rd and 4th of April: “Post-DFT/HF methods for the condensed phase with CP2K”: ADMM, RI, RPA,... (workshop with talks, virtual)
    - https://events.uni-paderborn.de/e/cp2kpostdft
  - ~Q3/24: Gromacs & CP2K on QM/MM (school with tutorial, 3-4 days in person)
- UK computer centre ARCHER2 is running an online workshop on use of CP2K on the 8th April.
  - If anyone would like to contribute a short presentation (20-30 mins) ideally linked to a tutorial or example page it would be really appreciated. RPA, LS, excited state contributions would be particularly great.
  - mwatkins@lincoln.ac.uk if interested.