High-throughput materials discovery and spectroscopy by beyond-DFT workflows and data-analysis frameworks

Claudia Draxl
3 projects ...

- FONDA
- FAIRmat
- NOMAD CoE
My research field

Computational materials science
Focus on theoretical spectroscopy

... and why do I care about workflows?
3 projects ... NOMAD CoE
Many materials / properties require methods better than semi-local DFT
Basically everything relevant for e.g. energy research

Complex systems require larger simulation cells
1 atom out of thousand(s) may determine the color

... and/or better methodology
Exascale codes

Bring DFT, Green-function based methods, and coupled-cluster theory to exascale

Develop libraries for code families

Covering planewaves, LAPW’s, numerical atom-centered orbitals

All-electron & pseudopotentials

Demonstrator codes
Our instrument

Our instrument exciting is a full-potential all-electron density-functional-theory package implementing the families of linearized augmented planewave methods. It can be applied to all kinds of materials, irrespective of the atomic species involved, and also allows for exploring the physics of core electrons. A particular focus are excited states within many-body perturbation theory.


exciting is a full-potential all-electron density-functional-theory package implementing the families of linearized augmented planewave methods. It can be applied to all kinds of materials, irrespective of the atomic species involved, and also allows for exploring the physics of core electrons. A particular focus are excited states within many-body perturbation theory.
Example: core-level spectra in MAPbI$_3$

Pb M$_5$ edge  

I L$_3$ edge

Exascale workflows

Serve as a glue
Controling computational parameters
Way more parameters in advanced methodology
Combining tasks / ‘recipes’
To be executed either in parallel or consecutively
Handling jobs
To guarantee error-free or error-corrected execution
Collecting results
Are
Method-specific
Largely code-specific
Example: high-throughput absorption spectra

Bethe-Salpeter equation

\[ \sum_{\nu' c' k' } H_{\nu c k, \nu' c' k' } A_{\nu' c' k'}^{\lambda} = E_{\lambda} A_{\nu c k}^{\lambda} \]

```xml
<input>
<title>BSE</title>

<structure>
  <define_some_structure_here>
  </structure>

<groundstate />

<xscript xtype="BSE" />
</input>
```
Bethe-Salpeter equation

Exciton wavefunction

Optical excitations

Valence

Conduction

Core excitations

Core

Conduction

Spectra

\[ \text{Im} \varepsilon_M(\omega) = \frac{8\pi^2}{\Omega} \sum_{\lambda} \sum_{vck} A^\lambda_{vck} \frac{\langle \mathbf{v} \mathbf{k} | \mathbf{p} | \mathbf{c} \mathbf{k} \rangle^2}{\epsilon_{ck} - \epsilon_{vk}} \delta(\omega - E^\lambda) \]
High-throughput absorption spectra ...

Bethe-Salpeter equation

\[ \sum_{v' c' k'} H_{vck, v' c' k'}^{e-h} A_{v' c' k'}^\lambda = E_{\lambda} A_{vck}^\lambda \]

\[ H_{vck, v' c' k'}^{\text{diag}} = (\varepsilon_{ck} - \varepsilon_{vk}) \delta_{vv'} \delta_{cc'} \delta_{kk'} \]

\[ H_{cvk, c' v' k'}^{\text{dir}} = \int d^3r d^3r' \frac{\psi_{vk}(r) \psi_{ck}^*(r') \varepsilon^{-1}(r, r') \psi_{v' k'}^*(r) \psi_{c' k'}(r')}{|r - r'|} \]

\[ H_{vck, v' c' k'}^{x} = \int d^3r d^3r' \psi_{vk}(r) \psi_{ck}^*(r) \vec{v}(r, r') \psi_{v' k'}^*(r') \psi_{c' k'}(r') \]
Bethe-Salpeter equation

\[ \sum_{\nu'c'k'} \hat{H}_{vck,\nu'c'k'} A_{\nu'c'k'}^\lambda A_{vck}^\lambda = E^\lambda A_{vck}^\lambda \]
Workflow for converging (core) spectra

- Structure
- Ground-state
- Dry run checking parameters
- Converge various parameters
- Converged result

Species, atom, edge
Range of input values, convergence criteria [1]
Database, analysis

excitingtools

Parser for over 50 exciting output files
+ tools for post-processing

Utilities for constructing exciting input files

Easy integration with workflow managers

https://pypi.org/project/excitingtools/

https://github.com/exciting/excitingtools/
Code pieces

**jobflow**
Dynamic workflows, HPC execution

**excitingworkflow**
Define workflow, calculation, convergence

**excitingtools**
Allows automatic interaction with exciting

**exciting**
Computation of absorption spectra
XANES benchmark project

Probing 3 different codes

Multi-code Benchmark on Simulated Ti K-edge X-ray Absorption Spectra of Ti-O Compounds

Fanchen Meng,1 Benedikt Maurer,2 Fabian Peschel,2 Sencer Selcuk,1 Mark Hybertsen,1
Xiaohui Qu,1 Christian Vorwerk,3,* Claudia Draxl,2, † John Vinson,4, ‡ and Deyu Lu1, §

1 Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, United States
2 Institut für Physik and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin Germany
3 Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States
4 Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

(Dated: March 15, 2023)

https://arxiv.org/abs/2303.17089
XANES benchmark workflow

Workflow combining 3 different codes

- Ions
  - PseudoDojo
  - SSSP
  - Species files

- Structure
  - k-grid
  - k-path

- Materials Project
  - Project

- DFT parameters

- DFT ground state

- Band structure

- X-ray absorption spectra

- Macroscopic dielectric function

- Spectral convergence
  - $|G+q|_{\text{max}}$ convergence
  - k-grid convergence

- Core-hole potential

- Supercell

- # of cond. bands

- OCEAN

- exciting

- XSPECTRA
Ti K edge

Ti-O structures

Intensity [arb. units]

Relative Energy (eV)
XANES code: Lightshow

Tool serving 3 different codes

Lightshow: a Python package for generating computational x-ray absorption spectroscopy input files

Matthew R. Carbone,1,* Fanchen Meng,2, † Christian Vorwerk,3 Benedikt Maurer,4 Fabian Peschel,4 Xiaohui Qu,2 Eli Stavitski,5 Claudia Draxl,4 John Vinson,6 and Deyu Lu2

1Computational Science Initiative, Brookhaven National Laboratory, Upton, New York 11973, United States
2Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, United States
3Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States
4Physics Department and IRIS Adlerhoff, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany
5National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, New York 11973, United States
6Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

(Dated: November 9, 2022)

https://arxiv.org/abs/2211.04452
Extreme-scale data

Artificial-Intelligence Toolkit

Brought to "real-time" performance
3 projects ...
FAIRmat is the NFDI consortium for solid state physics and the chemical physics of solids

https://fairmat-nfdi.eu

NOMAD is a web-based software for FAIR research data management in materials science

https://nomad-lab.eu

>140 mio calculations
FAIRmat is the NFDI consortium for solid state physics and the chemical physics of solids

https://fairmat-nfdi.eu

NOMAD is a web-based software for FAIR research data management in materials science

https://nomad-lab.eu

Solar cells

>140 mio calculations
A long way to the roof ...

Classical solar-cell research

Synthesis → Characterization → Device → Publication → Literature → Synthesis

- Literature: Monolithic perovskite/silicon tandem solar cell with >29% efficiency by enhanced hole extraction

Classical solar-cell research

The Perovskite Database Project

Pioneering project

100 authors

Manual handling of 42000 devices

16000 publications

More than 400 metadata categories

Synthesis information

Difficult to update

An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles

Heterogenous sources

Efficiency tables

Emerging inorganic solar cell efficiency tables (version 2)
Andriy Zakutayev1, Jonathan D Major2, Xiaojing Hao3, Aron Walsh4,5,6, Jiang Tang6,
Teodor K Todorov7, Lydia H Wong8,9 and Edgardo Saucedo10,9
Published 16 April 2021 • © 2021 The Author(s). Published by IOP Publishing Ltd

Journal of Physics: Energy, Volume 3, Number 3
Citation Andriy Zakutayev et al 2021 J. Phys. Energy 3 032003

Databases

Lab notebooks

SolarCellParameters

- Cell name
- data_file
- Certified values
- Certification institute
  - Light intensity
    - Unit: mW/cm²
  - Open circuit voltage
    - Unit: V
  - Short circuit current density
    - Unit: mA/cm²
- Fill factor
  - Unit: %
- Efficiency
  - Unit: %

Welcome Pepe Marquez
Explore solar cells in NOMAD

- Open Circuit Voltage \( V \)

Efficiency

Short Circuit Current Density (A/m²)

P. Marquez
NOMAD OASIS

Synthesis

Characterization

SLOW

Combinatorial synthesis
High throughput characterization
Autonomous labs
AI for an extra boost
Combinatorial explosion

Device stack

Synthesis

Characterization

SLOW

Device

contact

ETL

absorber

HTL

ETL: electron transport layer

HTL: hole transport layer
Computational high-throughput screening?

DFT level
Problematic

Many-body level
Not feasible
Problematic

Property 1
Property 2
Property 3
Property 4
Electronic Structure of (Organic-)Inorganic Metal Halide Perovskites: The Dilemma of Choosing the Right Functional

Cecilia Vona, Dmitrii Nabok, Claudia Draxl

First published: 17 December 2021 | https://doi.org/10.1002/adts.202100496
PbI$_2$ – a precursor

Band gaps from DFT and beyond ...
Electronic structure

Semi-local functionals

LDA, GGAs

\[
\begin{align*}
T + V_{\text{ext}}(r) + V_{\text{H}}(r) + V_{\text{xc}}(r) \psi_{i}^{\text{KS}}(r) &= \epsilon_{i}^{\text{KS}} \psi_{i}^{\text{KS}}(r) \\
\end{align*}
\]
Best performance?

$\gamma$-CsPbl$_3$
More to tune ...

**HSE:** mixing $\alpha$, screening $\omega$
3 projects ...
Science of workflows

FONDA
Foundation of Workflows for Large-Scale Scientific Data Analysis

Specific project
Deriving Trust Levels for Multi-Choice Data Analysis Workflows

Verification & validation
Data quality
Configuration-space sampling

>140 mio calculations
Similarity measures

Example: spectral fingerprints*

Tailored weighting of spectral features

Choose energy range, binning
Focus on particular region

Tanimoto’s similarity coefficient

\[ T_c(i,j) = \frac{f_i \cdot f_j}{|f_i|^2 + |f_j|^2 - f_i \cdot f_i} \]

\[ 0 \leq T_c \leq 1 \]

Spectral fingerprints

Data quality and error estimates

Example: bulk PbI₂

PBE – HSE
Tc=0.60

PBE – PBE+SOC
Tc=0.71

Cancellation of effects

Finding similar materials

Example

Materials similar to PbI$_2$

Based on 1,899,783 mio calculations
Wrap-up ...

FONDA

FAIRmat

NOMAD CoE

Customize your workflows ...

NOMAD
Thanks!!