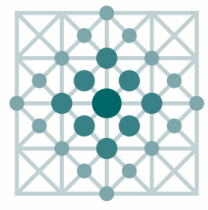
# High-throughput electronic-structure simulations: Where reliability really matters

Michael F. Herbst

Applied and Computational Mathematics, RWTH Aachen University
https://michael-herbst.com

05th May 2021



Applied and Computational Mathematics

RWTH AACHEN UNIVERSITY

→ https://michael-herbst.com/talks/2021.05.05_high_throughput_reliability.pdf

# Contents

# Societal challenges of 21st century

- Renewable energy
- Green chemistry and catalysts
- Drug design
- Transportation
- Data storage and communication

$\Rightarrow$ Need for novel materials

$\Rightarrow$ High-throughput computational screening

$\Rightarrow$ Common approach: Density-functional theory (DFT)

# Typical scale

- One DFT calculation: $\mathcal{O}(\text{hours})$ to $\mathcal{O}(\text{days})$
- E.g. Open Catalyst Project[1]
  - $1.3$ million DFT calculations
  - $> 250$ million DFT energy evaluations
  - Workflow success rate: $\simeq 50\%$[2]

$\Rightarrow$ Need high degree of automation

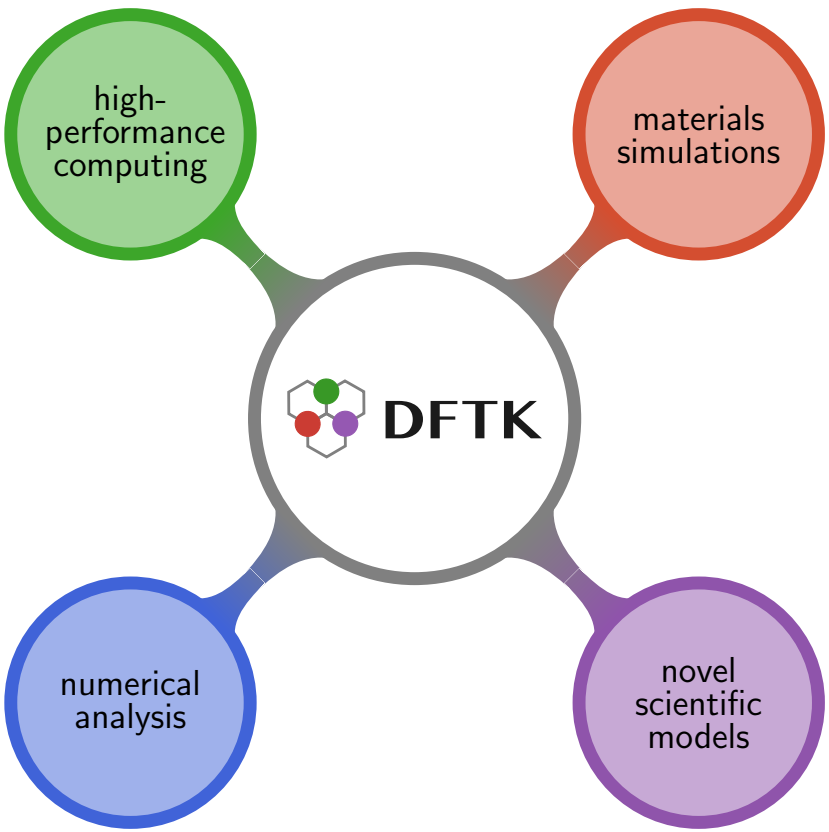
$\Rightarrow$ Reliability needs to be improved!

- Multidisciplinary research problem

[1] L. Chanussot *et. al.* The Open Catalyst 2020 (OC20) Dataset, 2020, arXiv 2010.09990.

[2] Z. Ulissi, private communication in ARPAE differentiate group seminar, Dec 2020.

# Density-functional toolkit (DFTK)




- `https://dftk.org`
- 2 years of development
- Pure julia code

- Supports mathematical developments *and* scale-up to regime relevant to applications
- Low entrance barrier: Only 6k lines of code!
- Scales to $> 800$ electrons

- `https://docs.dftk.org`

# Why julia?

*Walks like Python, talks like Lisp, runs like FORTRAN*

- Rich ecosystem (Optimisation, PDEs, stochastic processes, GPUs, Machine-Learning, statistics, linear algebra . . . )
- High-level, compiled and hackable
- No two-language problem: Everything stays within julia
- Multiple dispatch:
  - Generic fallbacks, fast code for special cases

  $\Rightarrow$ First get it to work then get it to work *fast*

  $\Rightarrow$ Write code once, re-use for many data structures / back ends
- `https://michael-herbst.com/learn-julia`

# Current research with DFTK

⇒ Vision: Improve high-throughput workflows:

- Use physics: Reliable black-box SCF algorithms[1]
- Use maths: Error estimates and automatic error balancing[2]
- Better algorithms: Numerical analysis of SCF methods[3]

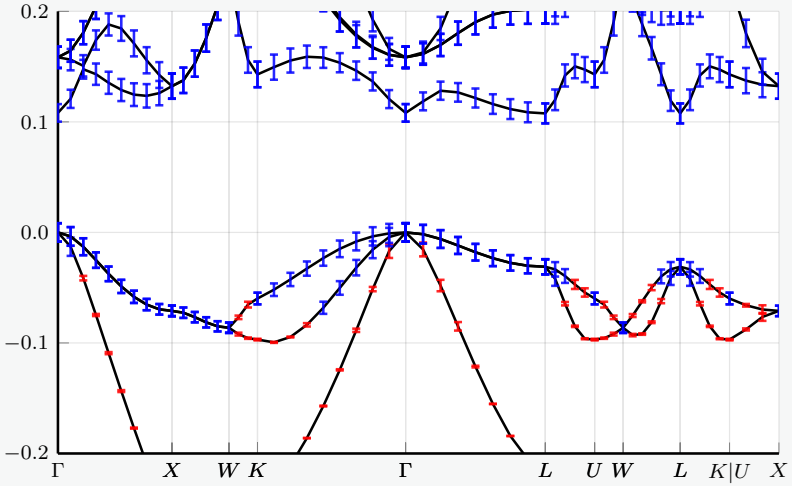
⇒ Reliable automatic selection of parameters in DFT workflows

- International and interdisciplinary user base:
  - RWTH Aachen, ENPC Paris, MIT, CMU, ...
- Involved in multidisciplinary research projects:
  - ACED-DIFFERENTIATE, EMC2 erc synergy, CESMIX

[1] M. F. Herbst, A. Levitt. J. Phys. Condens. Matter **33**, 085503 (2021).
[2] M. F. Herbst, A. Levitt and E. Cancès. Faraday Discuss. **224**, 227 (2020).
[3] E. Cancès, G. Kemlin, A. Levitt. arXiv 2004.09088 (2020).

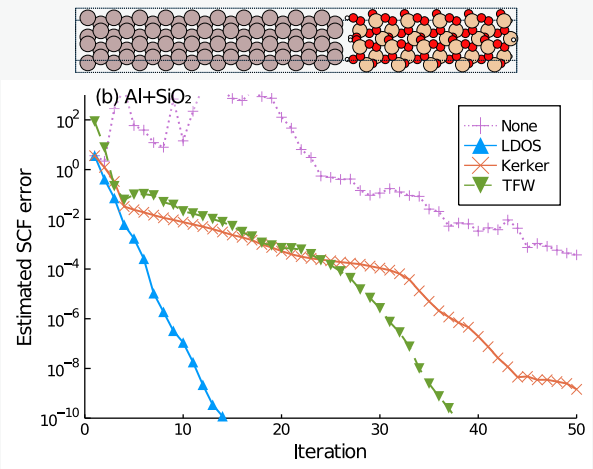
# Highlighted DFTK projects

## Error estimates for Kohn-Sham[1]




- A posteriori estimates for non-self-consistent Kohn-Sham
- Estimation of basis error, diagonalisation error, arithmetic error
- Time to publication: 10 weeks

## SCF for inhomogeneous systems[2]



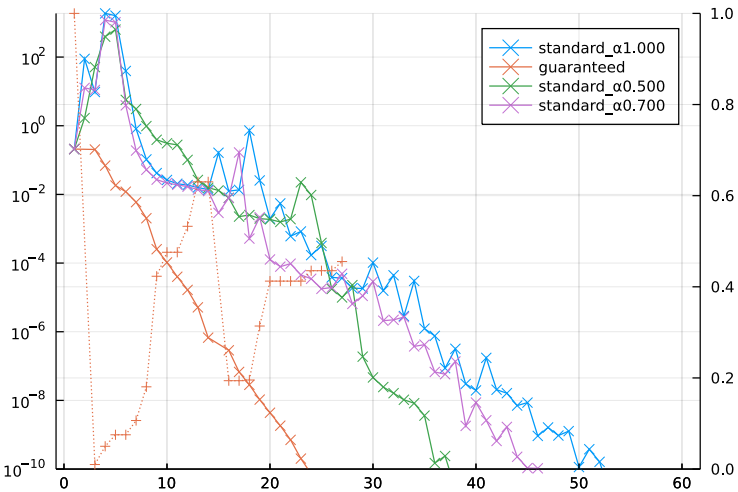

- SCFs for large inhomogeneous materials hard to converge
- LDOS: Black-box and parameter-free mixing scheme
- Same implementation for initial exploration and $> 800$ electrons

[1] M. F. Herbst, A. Levitt and E. Cancès. Faraday Discuss. **224**, 227 (2020).
[2] M. F. Herbst, A. Levitt. J. Phys. Condens. Matter **33**, 085503 (2021).

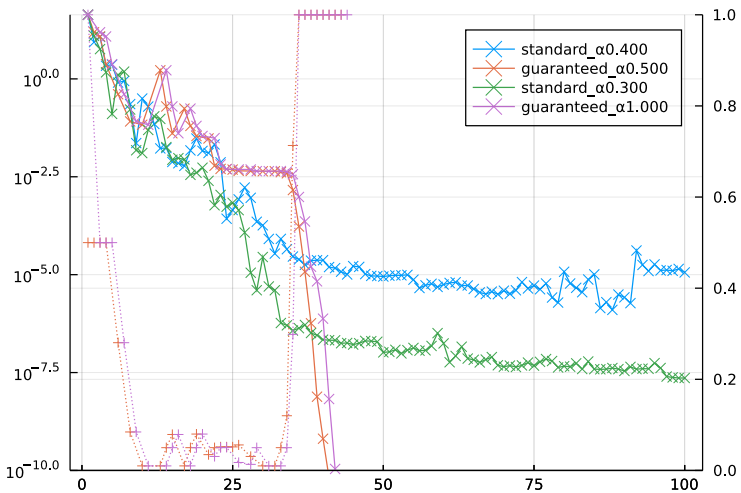
# Highlighted DFTK projects (2)

- Adaptive damping based on quadratic model
- Let the maths choose the damping
- Replace trial-and-error by something more rigorous
- Work in progress . . .

GaAs without proper mixing




$Fe_2AlMn$ (challenging spin system)



# Ongoing DFTK projects

- Feshbach-Schur methods for Schrödinger operators (Stamm)
- Towards full *a posteriori* error analysis (Stamm, Dusson, Cancès, Levitt)
- Improved initial guesses
  - Reduced order modelling (Stamm)
  - Machine learning (Viswanathan)

# Outlook

- Combination of adaptive damping an LDOS mixing
- Localised states / semiconductors (Blügel)
- RPA and post-DFT (Friesecke, Thicke)
- Mixed-precision methods on GPU (Edelman)
- Multi-fidelity methods (Marzouk)

# ADC in one slide

- Algebraic-diagrammatic construction (ADC) approach to electronic excitations
- Post-HF: Builds on Møller-Plesset PT ground states
- Intermediate states

$$|\Psi_n\rangle = \sum_I X_{I,n} \left|\tilde{\Psi}_I\right\rangle$$

- Hermitian eigenvalue problem

$$\mathbf{M}\mathbf{X} = \mathbf{\Omega}\mathbf{X}, \qquad \mathbf{X}^\dagger\mathbf{X} = \mathbf{I},$$

  with $\mathbf{M}$ ADC matrix and $\mathbf{\Omega}$ excitation energies.
- $\mathbf{M}$ sparse, so iterative methods employed (Jacobi-Davidson)

# Some questions related to ADC

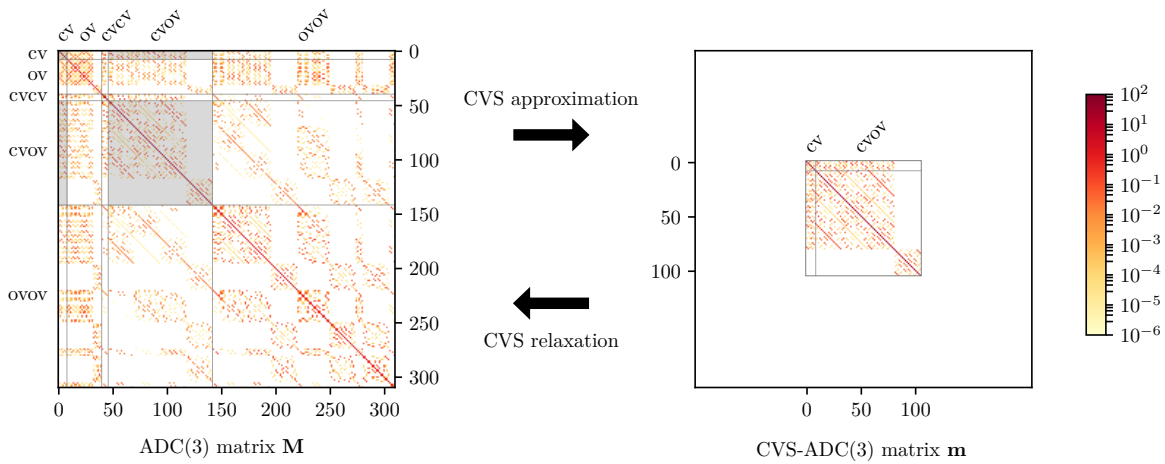
- ADC exists for multiple variants:
  - Core-valence separation (CVS)
  - Spin-flip
  - Frozen-core (FC) / frozen-virtual (FV)
- More specific numerics?
  - LOBPCG, Schur complement, preconditioning?
- Errors of CVS, FC, FV? Can these be undone?
- Interpolating from ADC($n$) to ADC($n+1$)

$\Rightarrow$ Difficult to address in previous frameworks

- Recently developed: *adcc* Python package (`https://adc-connect.org`)

# Current research with adcc: CVS relaxation[2]

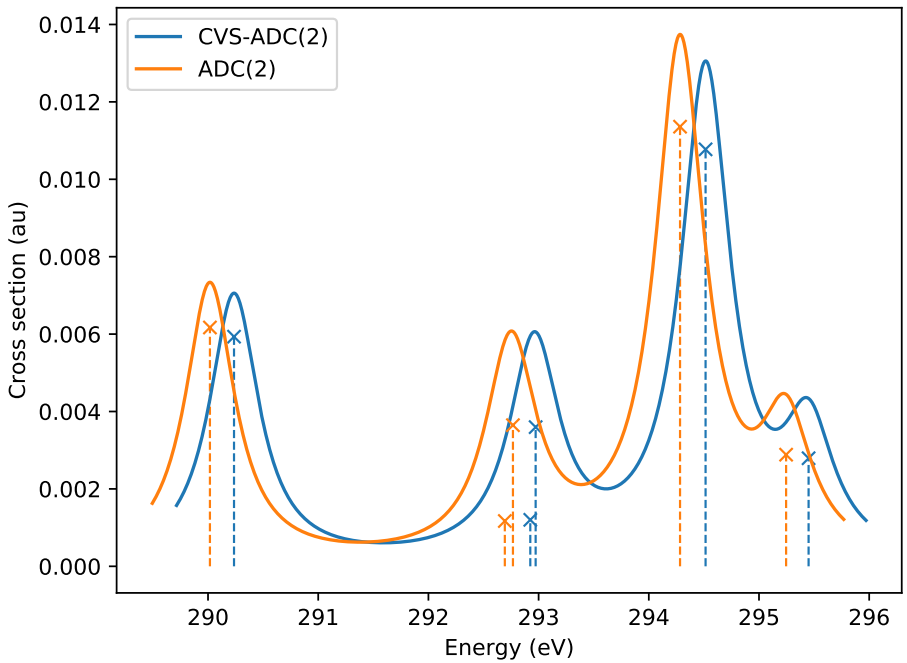
- CVS approximation used to target core excitations
- How large is the error?
- Used adcc[1] to develop numerical correction scheme




[1] M. F. Herbst, M. Scheurer, T. Fransson *et. al.* Wiley Interdiscip. Rev. Comput. Mol. Sci. **10**, e1462 (2020).

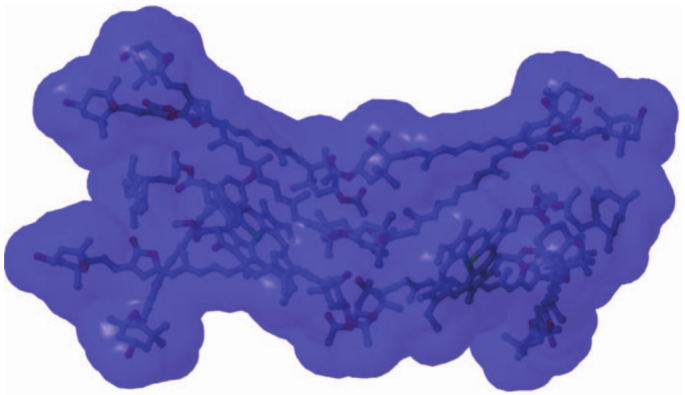
[2] M. F. Herbst and T. Fransson. J. Chem. Phys. **153** 054114 (2020).

# Typical 1s excitation (Fluoroethene)




- Error moderate for 1s core excitations
- WIP: 2s, 2p, . . . (Error definitely larger!)
- Joint with Thomas Fransson (KTH) and Sokolov group (Ohio)

# Continuum solvation models based on domain decomposition[1]



- Linear scaling, recently FMM acceleration
- Dissemination into modular package `ddX` (Stamm, Lipparini)
- WIP: Integration with Gaussian, Psi4, pyscf, adcc, ...

[1] F. Lipparini, G. Scalmani, L. Lagadère, *et. al.* J. Chem. Phys. **141** 184108 (2014).

# Summary

- DFTK
  - Mathematical developments for eigenvalue problems
  - Plane-wave Kohn-Sham DFT ground state & response theory
  - Supports LDA / GGA $> 800$ electrons
  - Broad development platform
  - $\Rightarrow$ High-throughput DFT methods
- Molecular systems
  - Error analysis / computational spectroscopy using adcc
  - linear-scaling continuum solvation with ddX

# Questions?


https://adc-connect.org

**DFTK** https://dftk.org


https://michael-herbst.com/learn-julia

mfherbst

https://michael-herbst.com/blog

herbst@acom.rwth-aachen.de