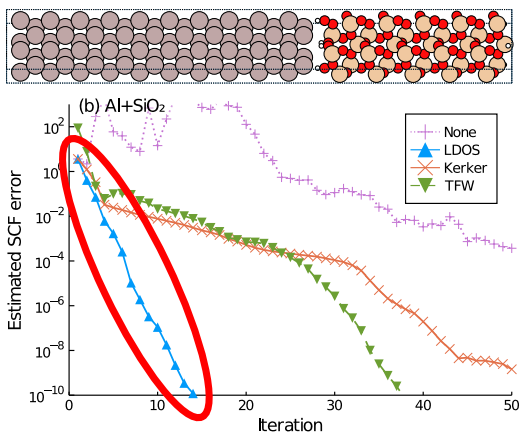
Robust error-controlled materials simulations

Michael F. Herbst

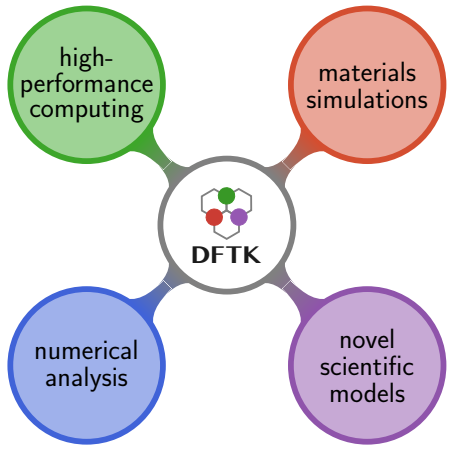
Mathematics for Materials Modelling (`matmat.org`), EPFL

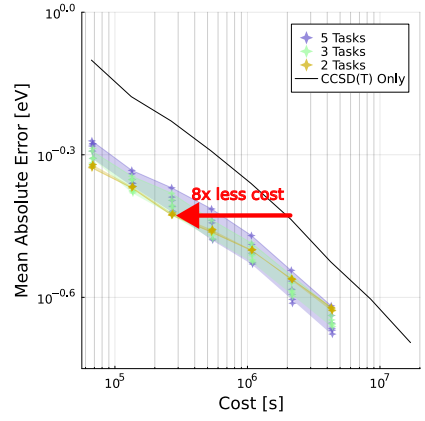
13 March 2024

`https://michael-herbst.com/slides/inaugural2024`



Novel materials simulation algorithms





Overcoming model deviations by multi-task learning

Energy consumption of materials discovery

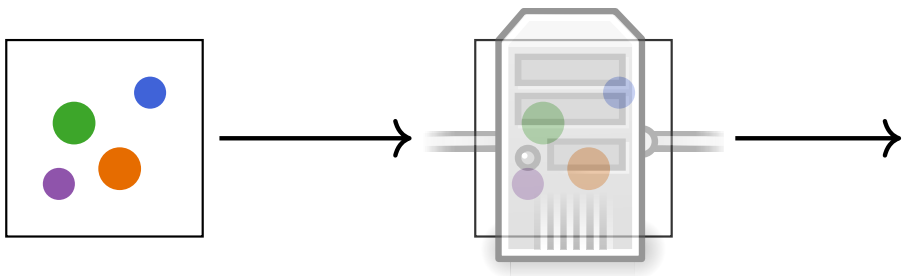
- Current solutions limited by properties of available materials
  - $\Rightarrow$ Innovation driven by discovering new materials
- Experimental research extremely energy intensive
  - 1 fume hood $\simeq$ 2-3 average households[1]

$\Rightarrow$ Complement experiment by computational materials discovery

[1] D. Wesolowski *et. al.* Int. J. Sustain. High. Edu. **11**, 217 (2010).

High-throughput materials screening

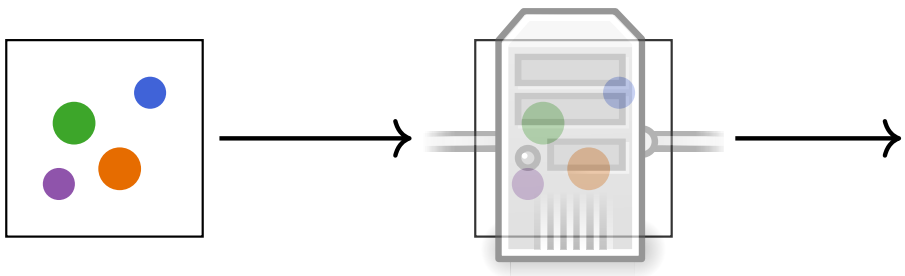
$$\min_{\Psi} \langle \Psi, H\Psi \rangle$$



- Energy consumption ?

High-throughput materials screening

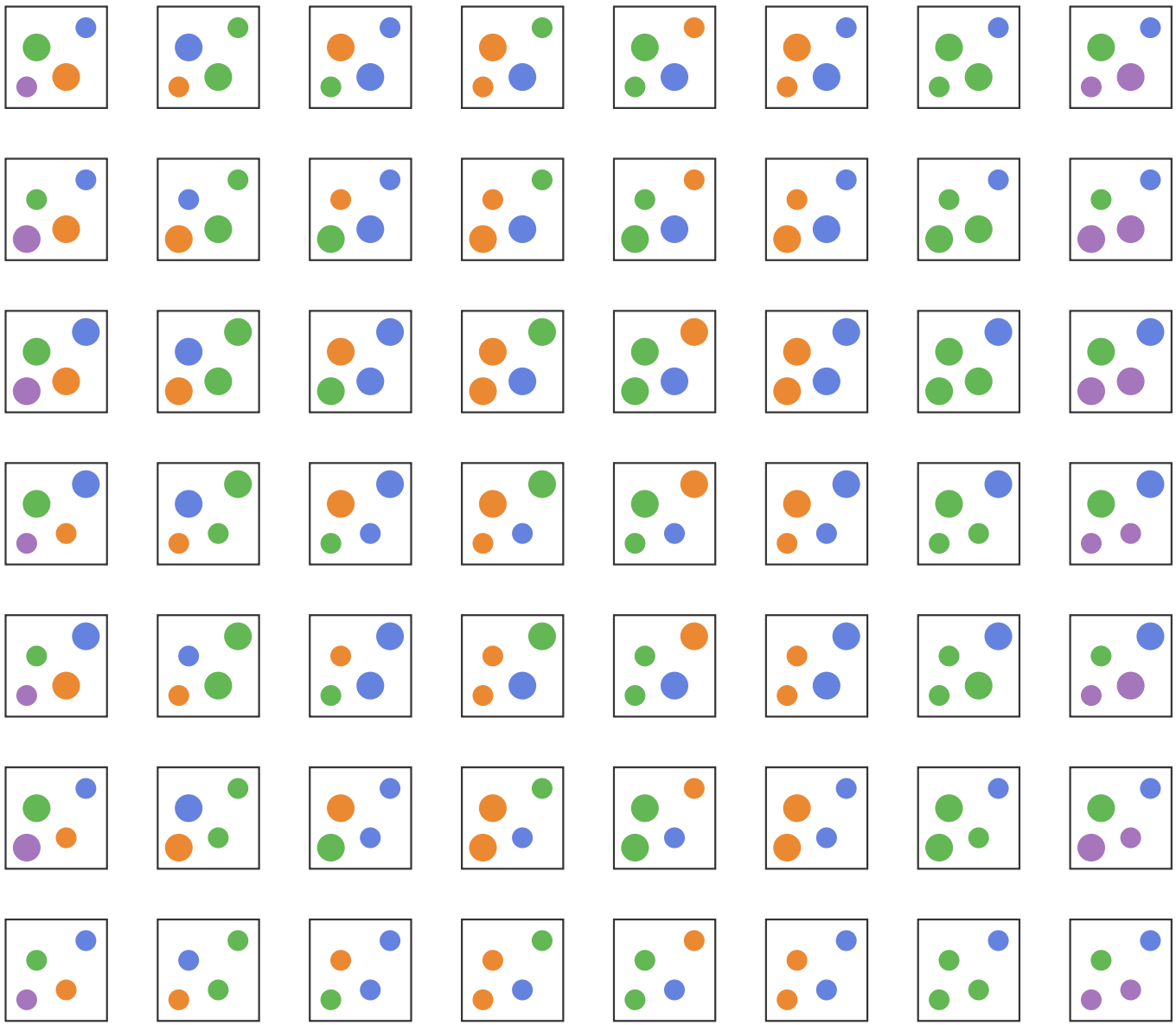
$$\min_{\Psi} \langle \Psi, H\Psi \rangle$$



- Energy consumption ?
  - 8h of 36-core processor
    $\simeq$ 4h of average household
    $\simeq$ 1 CHF

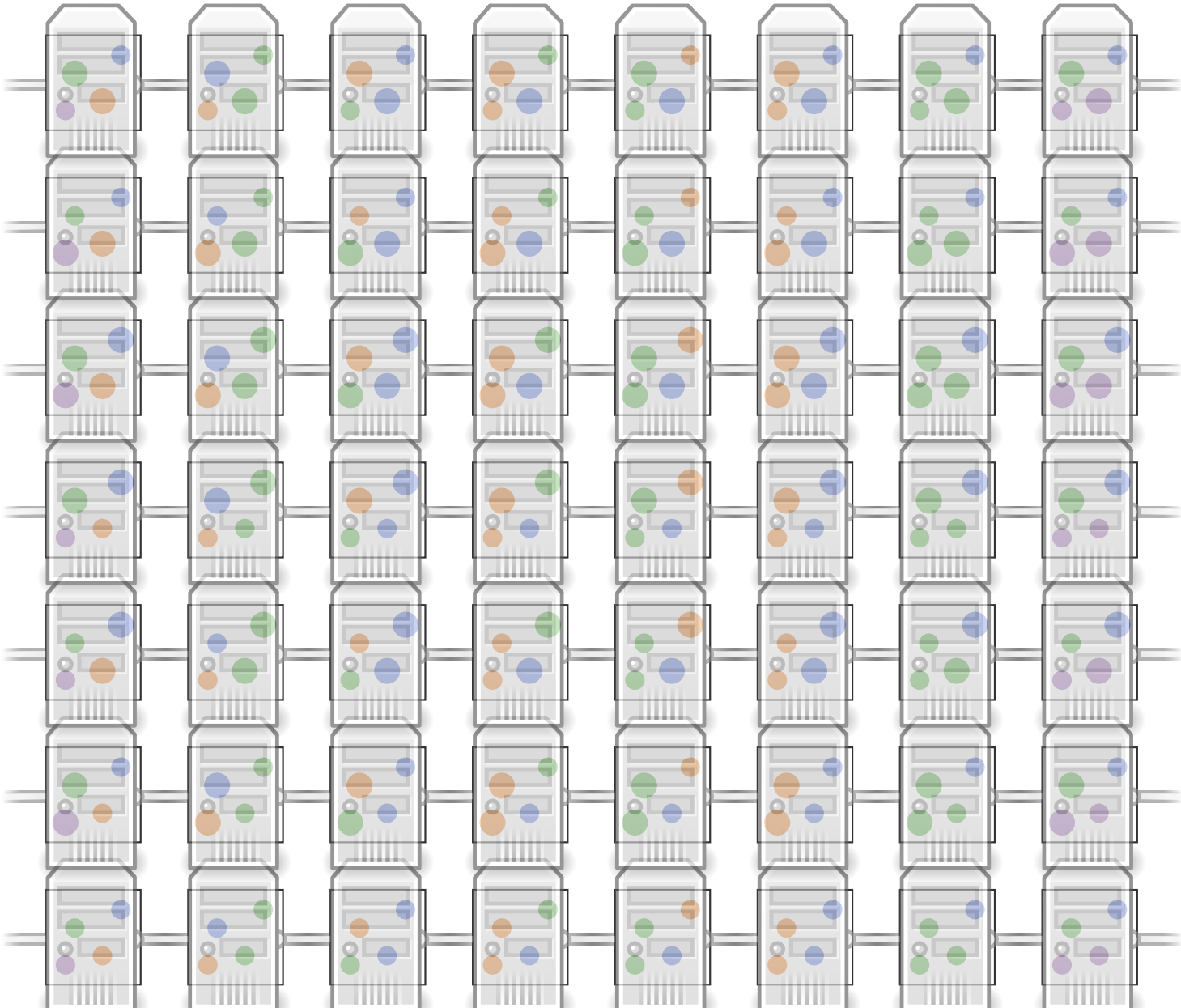
High-throughput materials screening

- We can **fully automate** this !

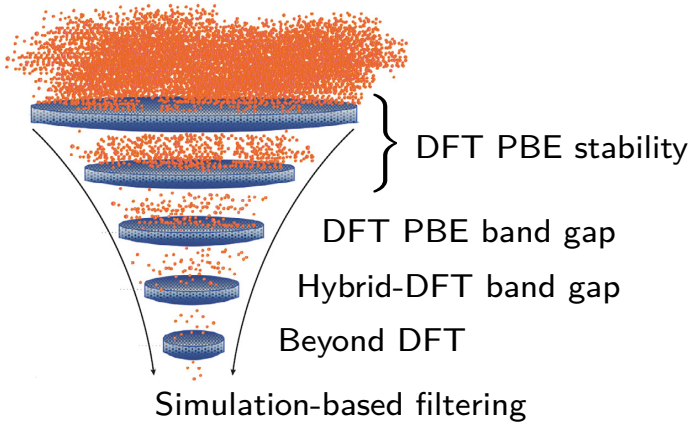


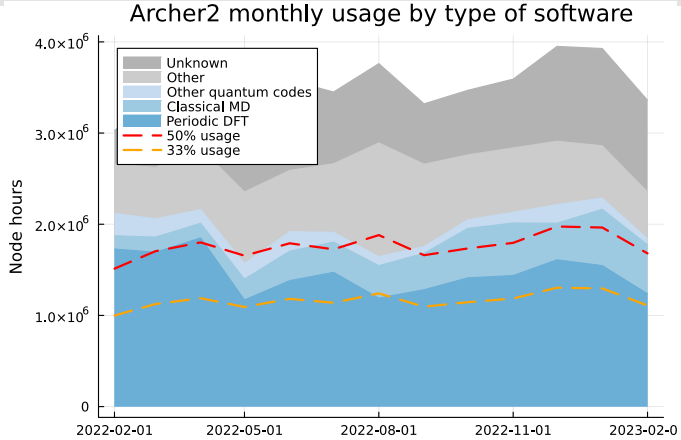
High-throughput materials screening

- We can **fully automate** this !



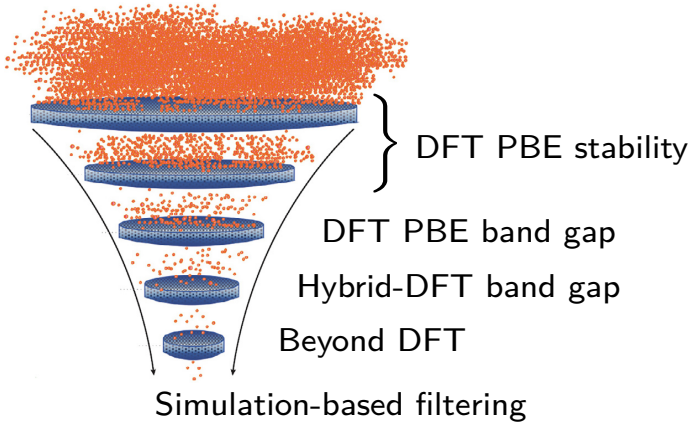
Computational materials discovery

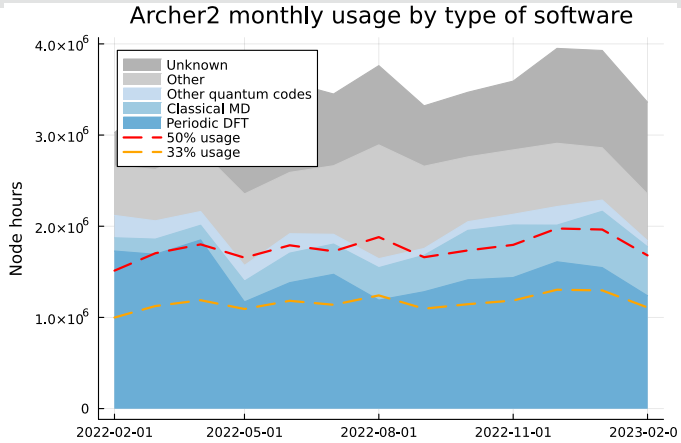




- Goal: Only promising candidates made in the lab
- Systematic simulations on $\simeq 10^4 - 10^6$ compounds
  - Noteworthy share of world's supercomputing resources

Computational materials discovery





- Goal: Only promising candidates made in the lab
- Systematic simulations on $\simeq 10^4 - 10^6$ compounds
  - Noteworthy share of world's supercomputing resources

- Energy consumption of LUMI (one of the most efficient):
  - 60 million kWh / year $\simeq$ 1.5 EPFLs $\simeq$ 14000 households

Challenges of high-throughput regime

- Complexity of multiscale materials modelling
  - Many parameters to choose (algorithms, tolerances, models)
  - Automated workflows & data management software (see above)
- Despite elaborate heuristics: Thousands of failed calculations
  - ⇒ Wasted resources
  - ⇒ Increased human attention (limits througput)
- ~~MatMat~~ **research directions**:
  - Robustness of algorithms
  - Uncertainties of models
  - Error propagation & error balancing
  - Multifidelity: The ideal mix of fast & accurate methods

Opportunities of cross-disciplinary research

- Role of mathematical research:
  - Unify notation & expose structure (Where do models differ ?)
  - Formalise & analyse problems (How can things go wrong ?)
- Historic importance: Formulation of quantum mechanics
  $\Rightarrow$ Collaboration mathematics & physics
- Materials modelling: Source for research problems
  - Large-scale eigenvalue problems
    (L. Lin, Y. Saad, C. Yang, ...)
  - Acceleration, fixed-point methods
    (T. Kelly, A. Miedlar, Y. Saad, R. Schneider, H. vd. Vorst, H. Walker, ...)
  - Non-linear PDEs
    (Z. Bai, E. Cancès, G. Friesecke, M. Lewin, I. Sigal, ...)
  - Many more examples ...
    (Approximation theory, optimisation, uncertainty quantification, ...)

$\Rightarrow$ Gap remains: Mathematical understanding $\Leftrightarrow$ Simulation practice

Difficulties of cross-disciplinary research

(A computational science point of view . . .)

- Community conventions . . .
  - Language barriers, publication culture, speed of research, . . .
- . . . that are cemented in software:
  - Priorities differ $\Rightarrow$ What is considered "a good code" differs

Mathematical software

- **Goal:** Numerical experiments
- **Scope:** Reduced models
- High-level **language**: Matlab, python, . . .
- **Lifetime:** 1 paper
- **Size:** $<$ 1k lines
- Does not care about performance

Application software

- **Goal:** Modelling physics
- **Scope:** All relevant systems
- Mix of **languages:** C, FORTRAN, python, . . .
- **Lifetime:** 100 manyears
- **Size:** 100k – 1M lines
- Obliged to write performant code

Difficulties of cross-disciplinary research

(A computational science point of view . . .)

Mathematical software

- **Goal:** Numerical experiments
- **Scope:** Reduced models
- High-level **language**:
  Matlab, python, . . .
- **Lifetime:** 1 paper
- **Size:** < 1k lines
- Does not care about performance

Application software

- **Goal:** Modelling physics
- **Scope:** All relevant systems
- Mix of **languages:**
  C, FORTRAN, python, . . .
- **Lifetime:** 100 manyears
- **Size:** 100k – 1M lines
- Obliged to write performant code

- Working with these codes requires different skillsets
  - $\Rightarrow$ Orthogonal developer & user communities
- Obstacle for knowledge transfer:
  - Mathematical methods never tried in practical setting
    (and may well not work well in the real world)
  - Some issues cannot be studied with mathematical codes
    (and mathematicians may never get to know of them)
- **Hypothesis:** People compose if software composes

Software to enable cross-disciplinary research

Quantum spin systems (MA+PY)

- `ReducedBasis.jl`: Reduced basis methods for eigenproblems
- Computation of phase diagrams[1,2]

Implicit solvation models (MA+CH)

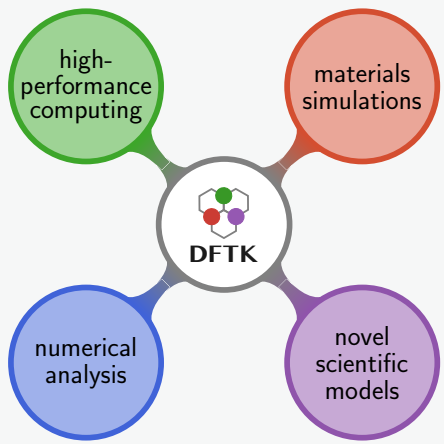
- `ddx`: Linear scaling solvation models up to protein-sized systems[3]

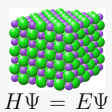
Computational spectroscopy (MA+CH)

- `adcc`: Algebraic Diagrammatic Construction methods[4]
- Error of core-valence separation[5]
  (Technique for simulating X-ray spectra)

First-principle materials modelling

- Density-Functional ToolKit[6]:
  Reduced settings *and* applications





[1] MFH, S. Wessel, M. Rizzi, B. Stamm. Phys. Rev. E **105**, 045303 (2022).

[2] P. Brehmer, MFH, S. Wessel, *et. al.* Phys. Rev. E **108**, 025306 (2023)

[3] M. Nottoli, MFH, A. Mikhalev, *et. al. ddx Polarizable Continuum Solvation* DOI 10.26434/chemrxiv-2024-787rx

[4] MFH, M. Scheurer, T. Fransson, *et. al.* WIREs Comput. Molec. Sci **10**, e1462 (2020).

[5] MFH, T. Fransson. J. Chem. Phys. **153**, 054114 (2020).

[6] MFH, A. Levitt, E. Cancès. JuliaCon Proc. **3**, 69 (2021).

Software to enable cross-disciplinary research

Quantum spin systems (MA+PY)

- `ReducedBasis.jl`: Reduced basis methods for eigenproblems
- Computation of phase diagrams[1,2]

Implicit solvation models (MA+CH)

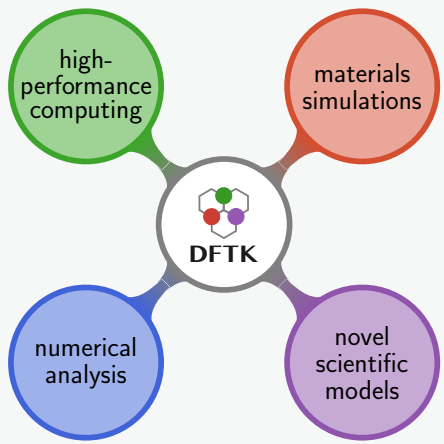
- `ddx`: Linear scaling solvation models up to protein-sized systems[3]

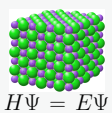
Computational spectroscopy (MA+CH)

- `adcc`: Algebraic Diagrammatic Construction methods[4]
- Error of core-valence separation[5] (Technique for simulating X-ray spectra)

First-principle materials modelling

- Density-Functional ToolKit[6]: Reduced settings *and* applications





[1] MFH, S. Wessel, M. Rizzi, B. Stamm. Phys. Rev. E **105**, 045303 (2022).

[2] P. Brehmer, MFH, S. Wessel, *et. al.* Phys. Rev. E **108**, 025306 (2023)

[3] M. Nottoli, MFH, A. Mikhalev, *et. al. ddx Polarizable Continuum Solvation* DOI 10.26434/chemrxiv-2024-787rx

[4] MFH, M. Scheurer, T. Fransson, *et. al.* WIREs Comput. Molec. Sci **10**, e1462 (2020).

[5] MFH, T. Fransson. J. Chem. Phys. **153**, 054114 (2020).

[6] MFH, A. Levitt, E. Cancès. JuliaCon Proc. **3**, 69 (2021).

Contents

1. Robust and efficient first-principle simulations

2. Determining errors and exploiting error control

EPFL MATMAT

Density-functional theory (insulators)

- Goal: Understand electronic structures (Many-body quantum system)
- DFT approximation: Effective single-particle model

$$\begin{cases} \forall i \in 1 \dots N : \left(-\frac{1}{2}\Delta + V\left(\rho_\Phi\right) \right) \psi_i = \varepsilon_i \psi_i, \\ V(\rho) = V_{\text{ext}} + V_{\text{Hxc}}(\rho), \\ \rho_\Phi = \sum_{i=1}^{N} |\psi_i|^2 , \end{cases}$$

- Self-consistent field (SCF) fixed-point problem

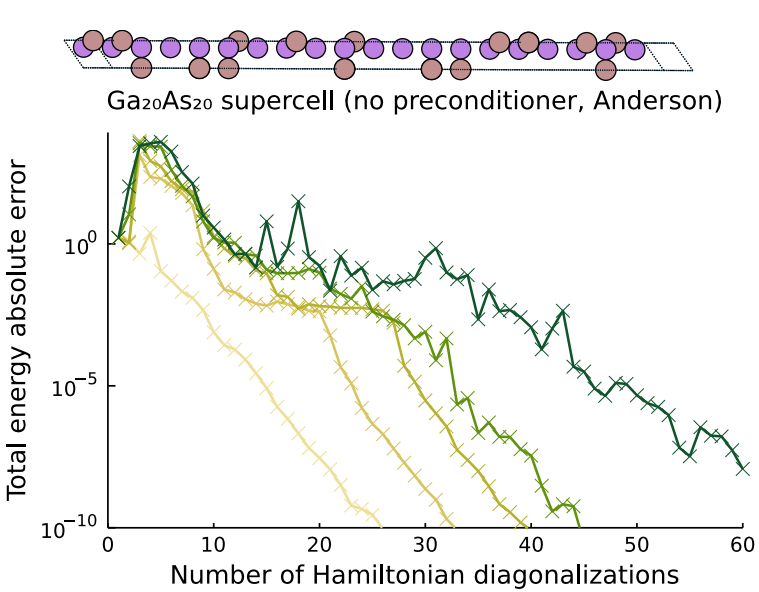
$$\rho\big(V(\rho)\big) = \rho$$

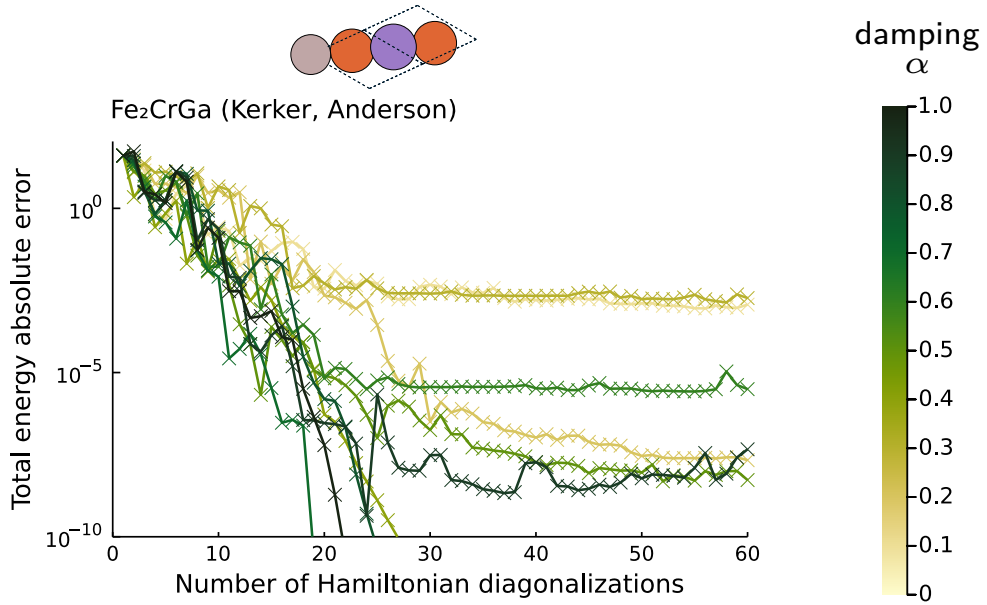
- Density mixing (preconditioner P, damping α)

$$\rho_{n+1} = \rho_n + \alpha\, P^{-1} \left[\rho\big(V(\rho_n)\big) - \rho_n \right]$$

- Best P & α highly system dependent (metal, insulator, . . .)
  - Usually chosen by trial and error (Impact on energy consumption . . .)

Illustration: Guessing a suitable damping α can be hard





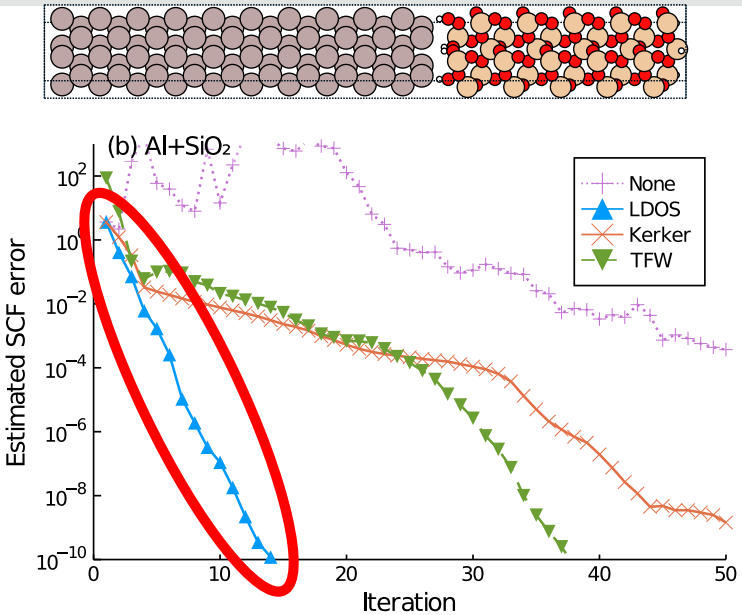
- Inefficient standard damping $(0.6 - 0.8)$
- Surprisingly small damping for smooth convergence

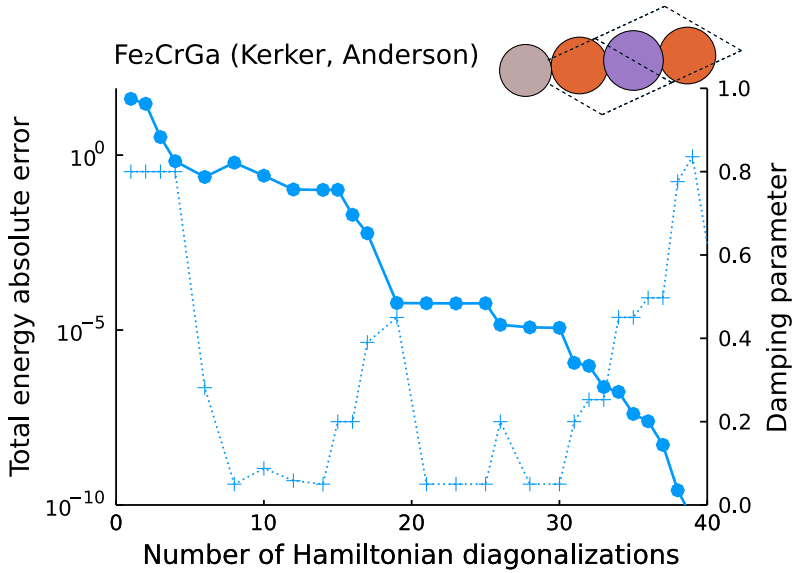
- Heusler alloy: Materials class with unusual magnetic properties

$\Rightarrow$ Numerically challenging behaviour

- SCF irregular: α versus convergence
- Usual heuristics breaks: Larger damping is better

Self-adapting black-box algorithms





- Preconditioning inhomogeneous systems (surfaces, clusters, . . .)
- LDOS preconditioner[1]: Parameter-free and self-adapting
- ca. 50% less iterations

- Damping α adapted *in each step* (using tailored quadratic model)
- Avoids trial and error (but may have a small overhead)
- Safeguard with theoretical guarantees[2]

⇒ Maths / physics collaboration:
Exchange of ideas between simplified & practical settings crucial

[1] MFH, A. Levitt. J. Phys. Condens. Matter **33**, 085503 (2021).

[2] MFH, A. Levitt. J. Comput. Phys. **459**, 111127 (2022).

What about properties ?

- DFT properties: Response of system to external changes:
  - Connection Theory $\Leftrightarrow$ Experiment
  - Modelling: Potential $V(\theta, \rho)$ depends on parameters θ
    (e.g. atomic positions, el. field)

- SCF procedure yields fixed-point density ρ_{SCF}

$$0 = \rho\Big(V(\theta, \rho_{\mathrm{SCF}})\Big) - \rho_{\mathrm{SCF}}$$

$\Rightarrow$ Defines implicit function $\rho_{\mathrm{SCF}}(\theta)$

- Properties are derivatives:
  - **Forces** (energy wrt. position), **dipole moment** (energy wrt. el. field), **elasticity** (energy cross-response to lattice deformation), phonons, electronic **spectra**, ...

$\Rightarrow$ Density-functional perturbation theory (implicit differ$^{\mathrm{n}}$.)

$$\frac{\partial \rho_{\mathrm{SCF}}}{\partial \theta} = \left[1 - \chi_0 K\right]^{-1} \chi_0 \frac{\partial V}{\partial \theta}$$

- Need *many* applications of χ_0 (Independent particle susceptibility)

Sternheimer equations

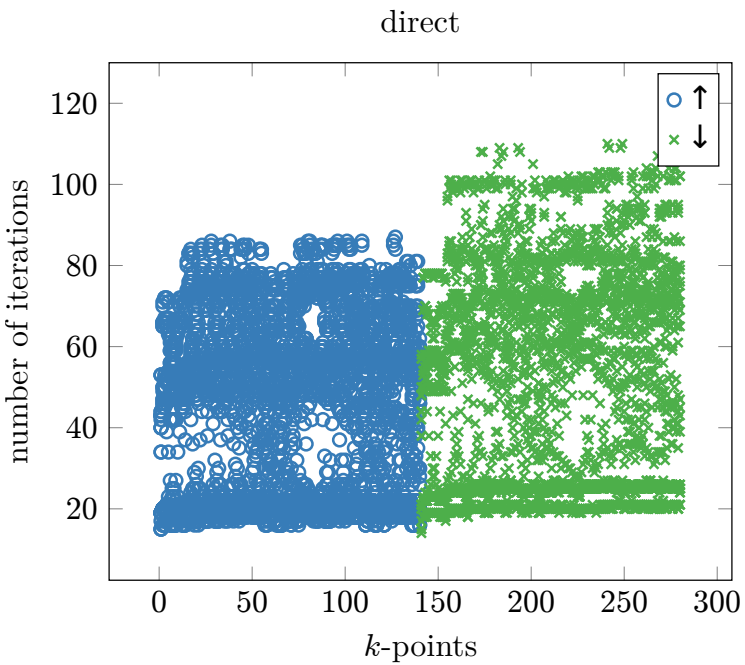
- Product $\chi_0 \delta V$ requires solving Sternheimer equations

$$\left(\tilde{H} - \varepsilon_i\right) \delta\psi_i = -P\,\delta V \psi_i \qquad \forall i = 1, \ldots, N$$

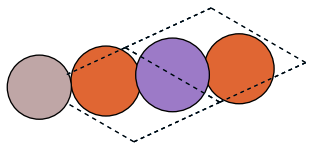
$H = -\frac{1}{2}\Delta + V$, $\tilde{H} = PHP$ and P some projector

(ε_i, ψ_i) eigenpairs of H

$\Rightarrow$ Badly conditioned for metallic systems (ε_i near eigenvalue of $\tilde{H}$)



(Number of iterations for various i)



Fe_2MnAl Heusler alloy

Schur complement approach to response[1]

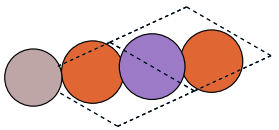
- Numerics of eigensolver:
  We have N_{ex} "extra" bands
- Use these to partition $\tilde{H}$:

$$\tilde{H} = \begin{pmatrix} E_{\text{ex}} & \mathbf{C} \\ \mathbf{C}^\dagger & \mathbf{R} \end{pmatrix}$$

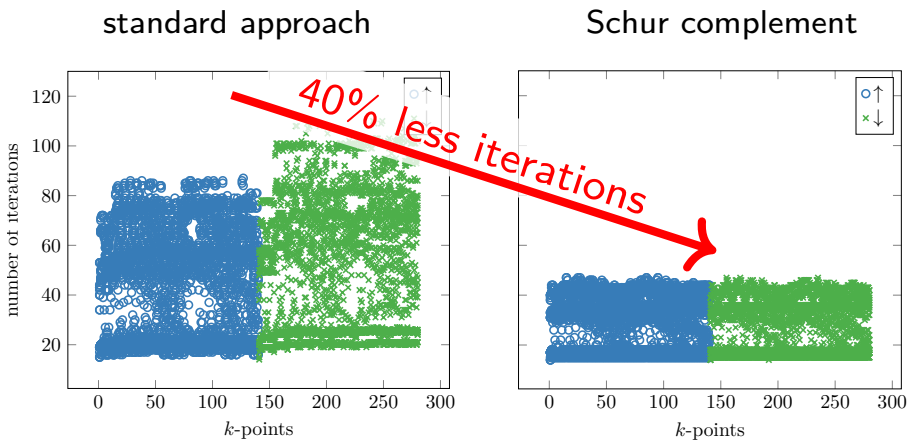
$E_{\text{ex}} = \text{diag}(\varepsilon_{N+1}, \ldots, \varepsilon_{N+N_{\text{ex}}})$
& $\mathbf{C}$, $\mathbf{R}$ projections of $\tilde{H}$

$\Rightarrow$ Use Schur complement:
Better-conditioned systems

$$(\mathbf{R} - \varepsilon_i)x = b$$



Fe_2MnAl Heusler alloy

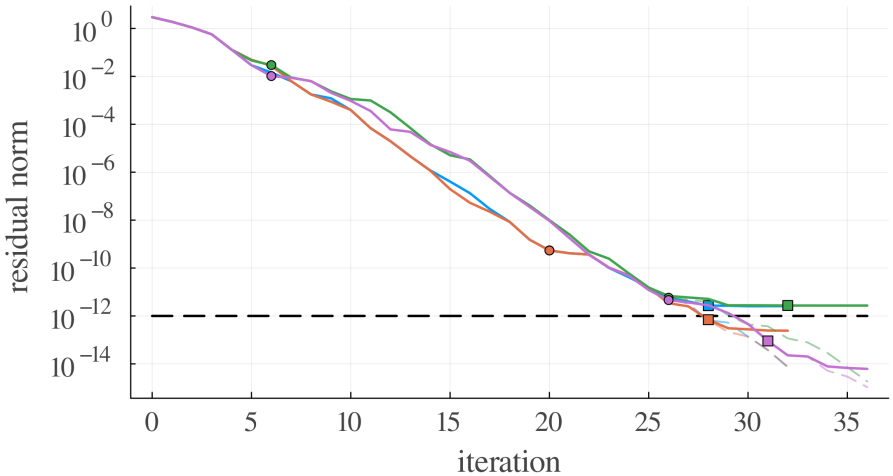


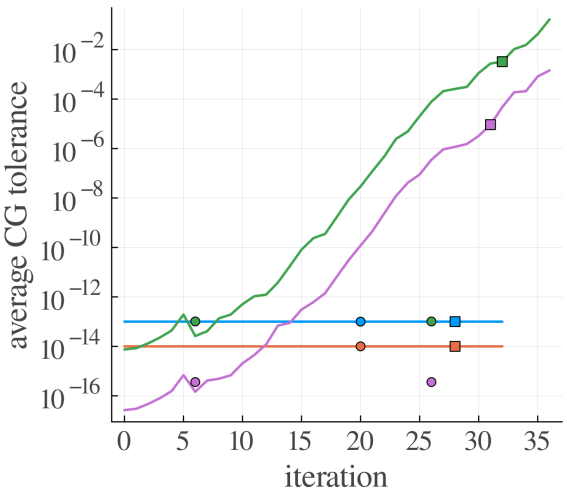
- Schur-based approach tames CG
- ca. 40% less iterations
- Development guided using a "real material"

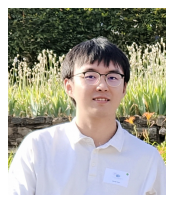
[1] E. Cancès, MFH, G. Kemlin, *et. al.* Lett. Math. Phys. **113**, 21 (2023).

WIP: Inexact Krylov methods

- DFPT + Sternheimer: Nested linear problems
- Inexact Krylov methods:[1] Framework to tolerate *less tight* solutions of Sternheimer
- First results indicate 25%–50% less Hamiltonian applications (the expensive step)





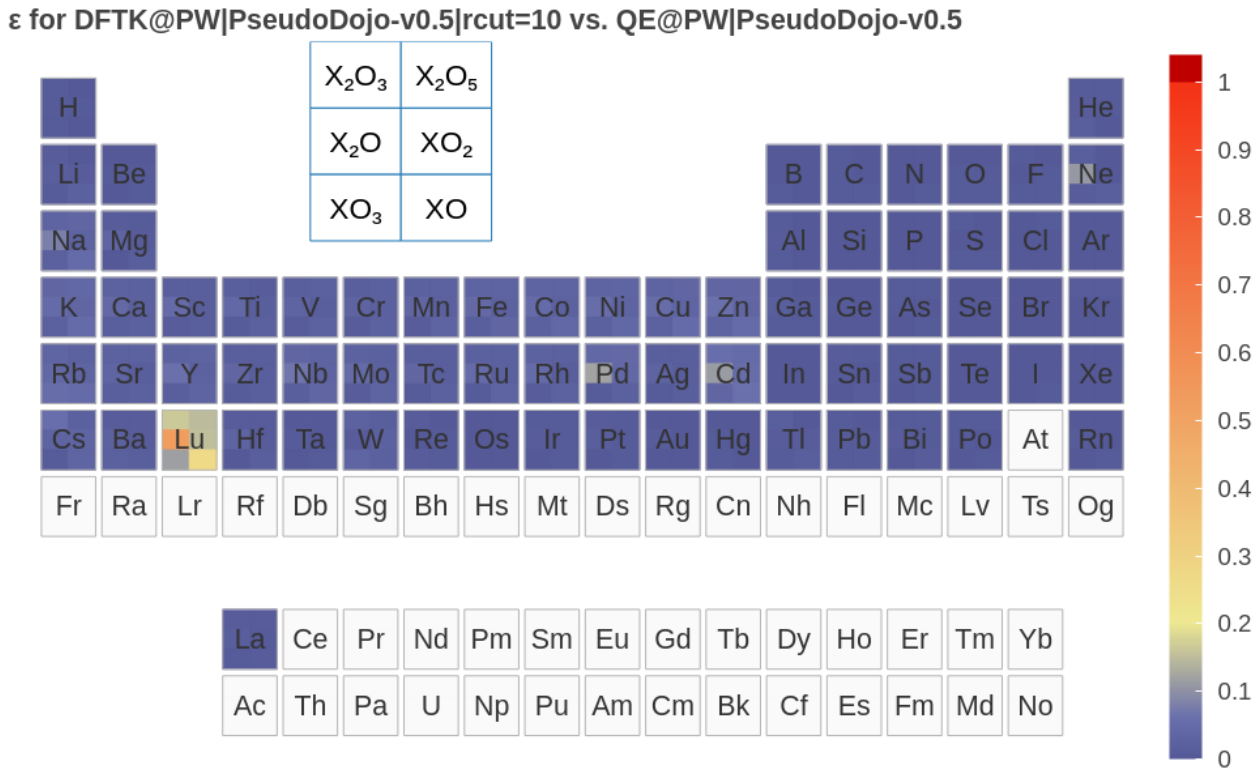


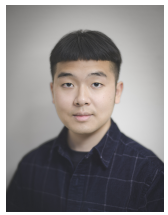
Bonan Sun

[1] V. Simonicini, D. Szyld. SIAM J. Sci. Comput., **25**, 454 (2003).

WIP: Close the gap between maths and high-throughput

- DFTK plugin for AiiDA workflow manager
- **Goal:** Simplify automated testing of novel algorithms
- Verification study Quantum-Espresso vs. DFTK





Yihan Wu

Contents

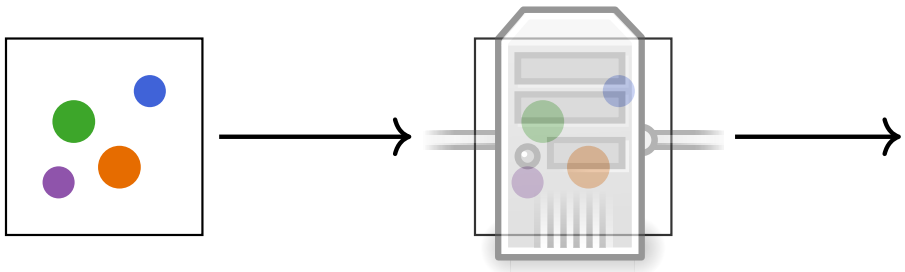
1. Robust and efficient first-principle simulations

2. Determining errors and exploiting error control

EPFL MATMAT

Recall our goal

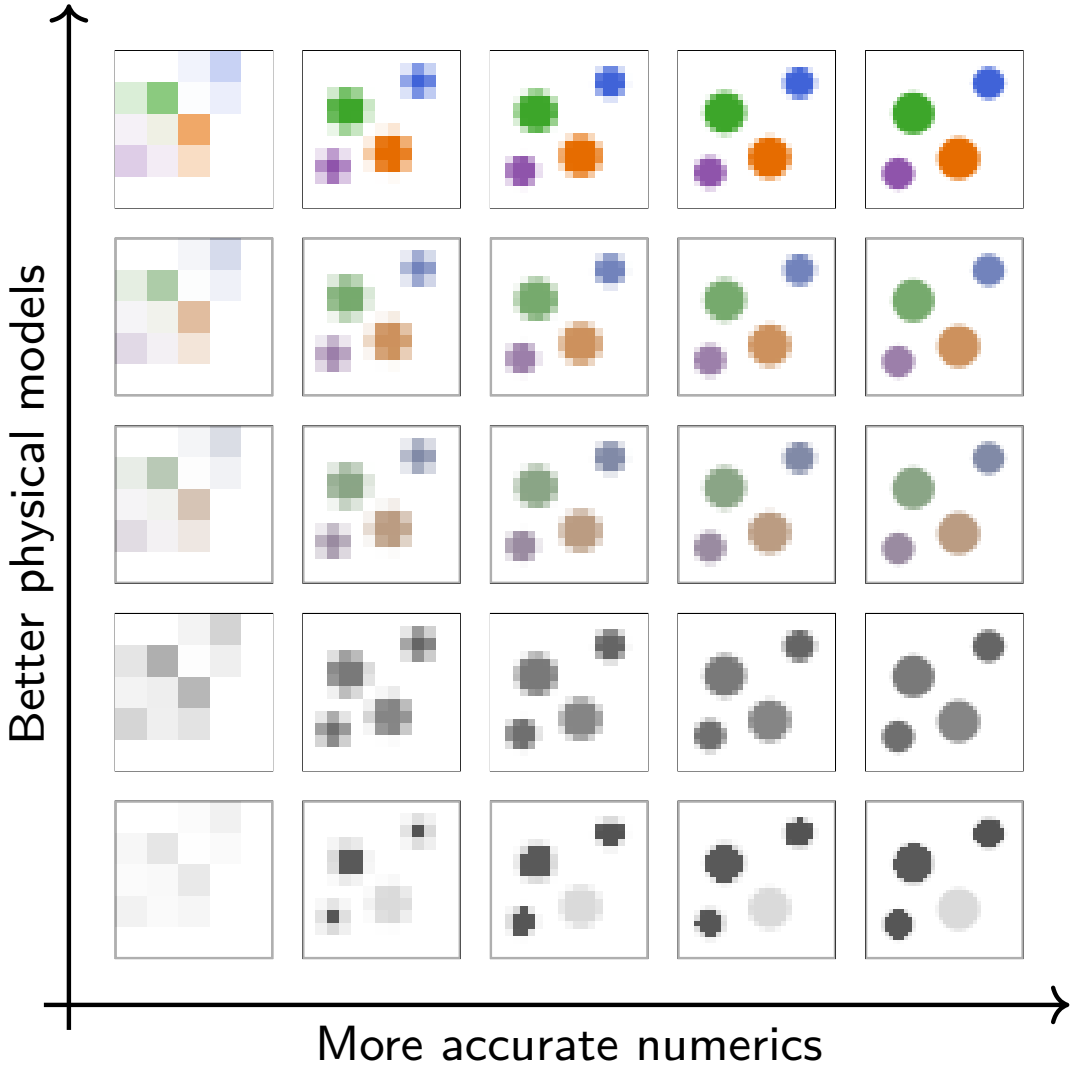
$$\min_{\Psi} \langle \Psi, H\Psi \rangle$$



- Need to repeat on 10^6 unseen systems . . .
  - Which model/numerics ?
  - How accurate is our answer ?

$\Rightarrow$ Question of error control

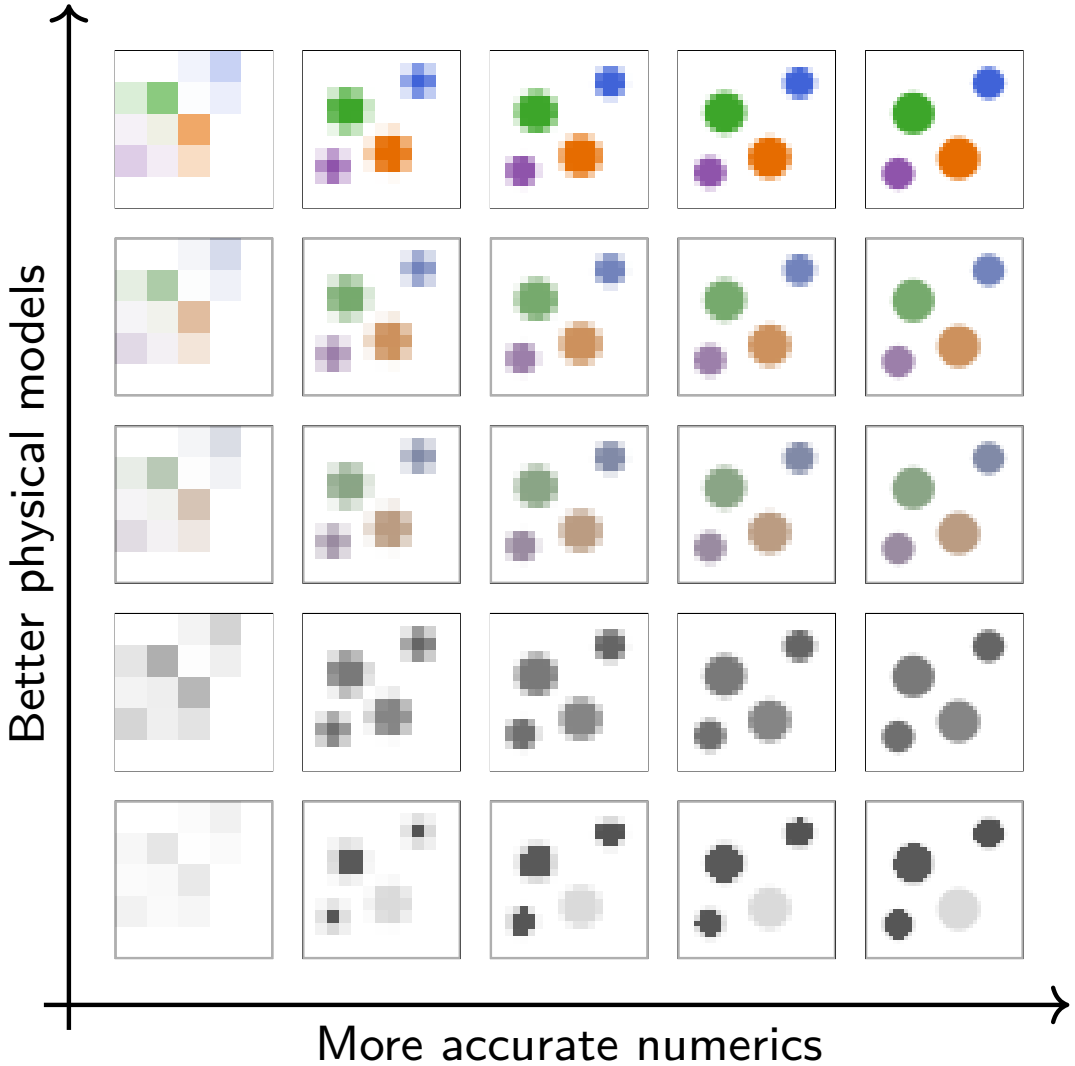
Error comes in different flavours



- Ideally want to balance errors

$\Rightarrow$ Need reliable error indicators !

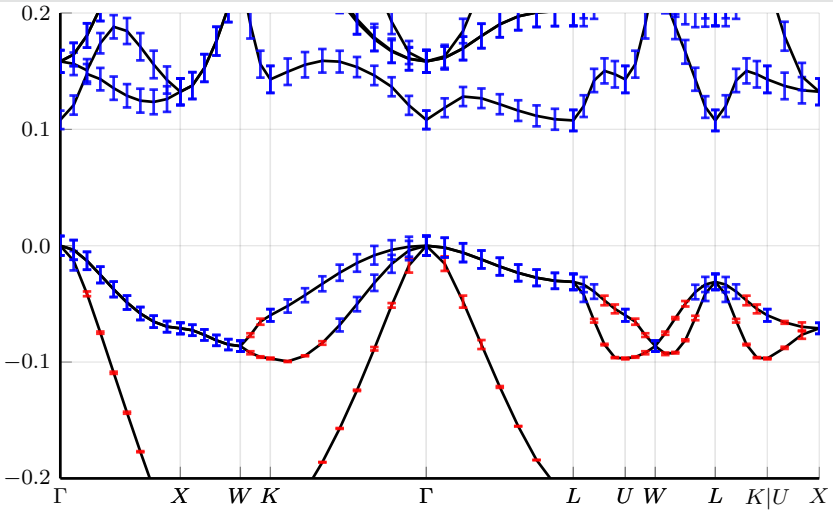
Error comes in different flavours



- Ideally want to balance errors

$\Rightarrow$ Need reliable error indicators !

Numerical error: Analytical techniques



Band structure with guaranteed errors[1]

- Momentum towards numerical error estimators for DFT
  - Focus on basis set error (some also tackle floating-point, SCF convergence)
- Results promising, but many challenges & caveats remain
  - Numerical experiments & problem simplifications crucial

  ⇒ **DFTK** is major research tool for this development[1-4]
- Techniques for DFT error less developed (and hard to tackle analytically)

---

[1] MFH, A. Levitt, E. Cancès. Faraday Discus. **223**, 227 (2020).

[2] E. Cancès, G. Dusson, G. Kemlin *et. al.* SIAM J. Sci. Comp., **44**, B1312 (2022).

[3] E. Cancès, G. Kemlin, A. Levitt. J. Matrix Anal. Appl., **42**, 243 (2021).

[4] E. Cancès, G. Kemlin, A. Levitt. J. Sci. Comput., **98**, 25 (2024)

DFT error: Computing model sensitivities

- Consider model sensitivity of force $\mathcal{F}(\rho_{\text{SCF}}(\theta))$:

$$\frac{d\mathcal{F}}{d\theta} = \frac{\partial \mathcal{F}}{\partial \rho_{\text{SCF}}} \frac{\partial \rho_{\text{SCF}}}{\partial \theta} \tag{1}$$

- Computed by response theory (we've seen this before !):

$$\frac{\partial \rho_{\text{SCF}}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta}$$

- Parameters appear in innermost layer (model definition)
  - Each DFT model: Different derivatives $\frac{\partial V}{\partial \theta}$ (can be horrible)
  - Each quantity of interest: Different sensitivity expression (1)

  $\Rightarrow$ Combinatorial explosion

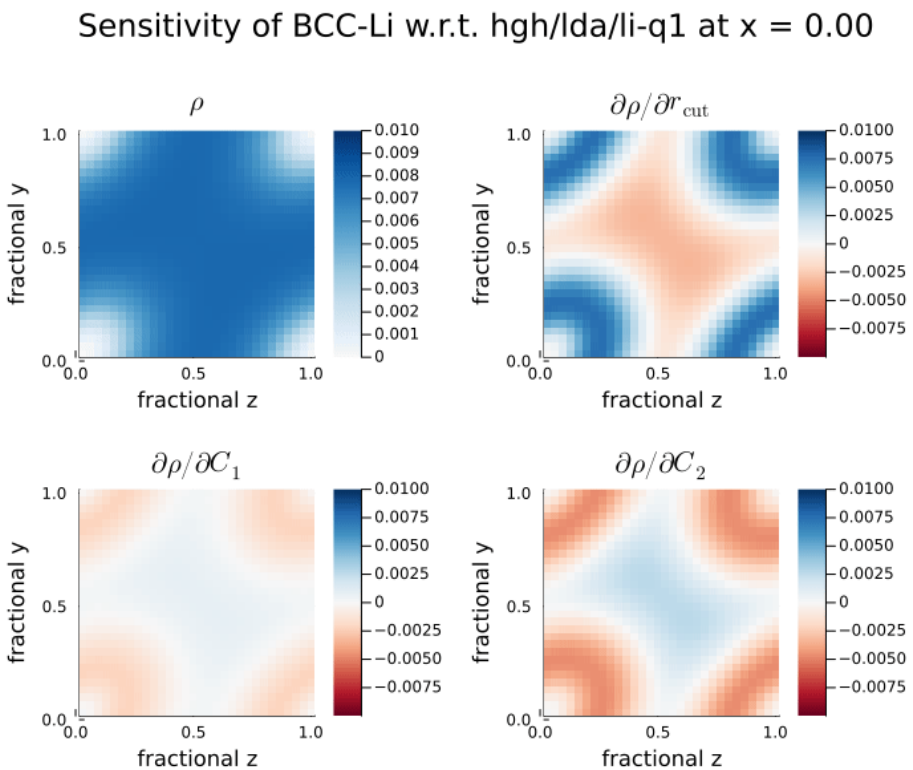
WIP: Sensitivity analysis in one line of code

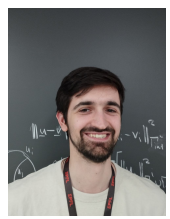
- **DFTK**: Algorithmic differentiation (AD)
  - Generic framework for derivatives: Request gradient, AD delivers

  ⇒ New properties/derivatives by non-DFT experts!

⇒ Setting for uncertainty quantification:

- Pseudopotential sensitivity of electronic density





Niklas Schmitz

WIP: Exploiting AD: Derivatives to guide materials design

- Materials design: Best design variables θ ...
  (strain, number of layers, alloy composition, ...)

$$\min_{\theta} \|\text{target} - \text{prediction}(\theta, R_*(\theta))\| \tag{1}$$

- ...subject to minimal energy (material should be stable)

$$R_*(\theta) = \arg\min_{R} \mathcal{E}(\theta, R) \tag{2}$$

- State of the art: Gradient-free methods for (1)

- Blocker for gradient methods: We need unusual gradients

$$\frac{\partial R_*}{\partial \theta} = -\left(\frac{\partial^2 \mathcal{E}}{\partial R^2}\right)^{-1} \frac{\partial^2 \mathcal{E}}{\partial R \partial \theta}$$

$\Rightarrow$ Combinatorial explosion to implement them all

$\Rightarrow$ Easily obtained with AD in DFTK! (at least some)

Cédric Travelletti

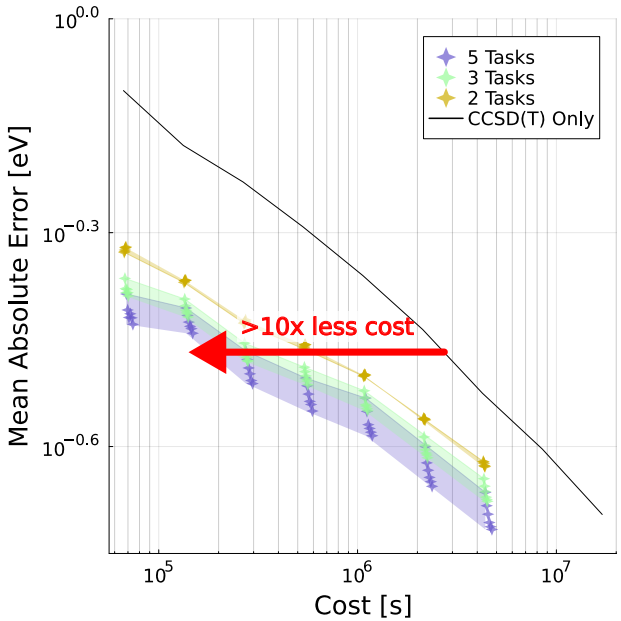
Exploiting heterogeneous modelling

	density-functional theory (DFT)			coupled cluster
model	PBE	PBE0	PBE0_DH	CCSD(T)
scaling	$O(N^3)$	$O(N^3)$	$O(N^3)$	$O(N^7)$
advantage	cheap	cheapish	cheapish	accurate
small organics	seconds to minutes			hours/days

- > 100 years of quantum chemistry: Zoo of models
- **Guiding idea:** Can we combine data from different functionals to balance accuracy / cost / deviating predictions?
  - $\Rightarrow$ Setting of Statistical learning (tricky analytically)
- Opportunities:
  - Reduce data generation cost for learning
  - Dataset of opportunity
- Challenge: Not always a clear which model is more accurate

Multitask learning: Modelling correlations & discrepancies[1]

- **Goal:** Surrogate for highest fidelity using mostly heterogeneous low-fidelity data
- Training: Limited CCSD(T) data plus mixture of different DFT models (tasks)
- Error: Model prediction versus true CCSD(T) result
- Cost: Computational time for data generation



	Core	Additional	Target
CCSD(T)	●●●●●●●●●●		●●●●●●●●●
PBE0_DH	●●●●	●●●●	
PBE0	●● ●●	●● ●●	
PBE	●● ●●	●● ●●	
BLYP	●● ●●	●● ●●	●●●●●●●●●

Katherine Fisher
(MIT grad student)

[1] K. Fisher, MFH, Y. Marzouk *Multitask meth. to predict molec. prop. from heterogeneous data* arXiv 2401.17898

Summary: Research in the MatMat group

- High-throughput materials discovery
  - Need for automation: Robustness and efficiency are key!
  - Understand errors ⇒ Precise accuracy tuning
- Robust and efficient simulation algorithms
  - Reduce user-chosen parameters, improve numerical schemes
  - Build on combining mathematical and physical insight
- Error control for first-principle modelling
  - Combination of statistical and analytical approaches
  - Sensitivity analysis & derivative-guided design
  - Multi-tasking surrogates: The best of accurate & cheap models
- **DFTK**: A DFT software for cross-disciplinary research
  - Reduced settings (error analysis) *and* high-throughput testing
  - Unlocks opportunities of algorithmic differentiation

  ⇒ Overcome barriers: People compose if software composes

Acknowledgements

MatMat group

- Niklas Schmitz (MatMat)
- Cédric Travelletti (MatMat)

Robust algorithms

- Eric Cancès (École des Ponts)
- Gaspard Kemlin (Université de Picardie)
- Antoine Levitt (Université Paris-Saclay)
- Benjamin Stamm (Stuttgart)
- Bonan Sun (EPFL)

Multi-task methods

- Katharine Fisher (MIT)
- Youssef Marzouk (MIT)

Aiida interface & verification

- Giovanni Pizzi (PSI)
- Junfeng Qiao (EPFL)
- Yihan Wu (EPFL)
- Austin Zadoks (EPFL)

Fonds national suisse

MARVEL
NATIONAL CENTRE OF COMPETENCE IN RESEARCH

CESMIX

EPFL MatMat

Questions?

https://matmat.org


mfherbst


michael.herbst@epfl.ch


https://michael-herbst.com/slides/inaugural2024

https://dftk.org

EPFL M t Mat