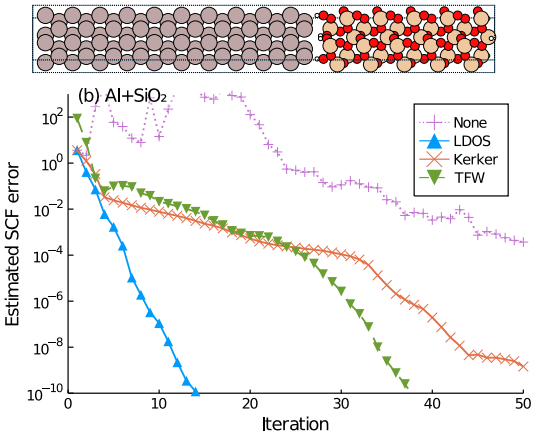
Accelerating the discovery of tomorrow's materials by robust & error-controlled simulations

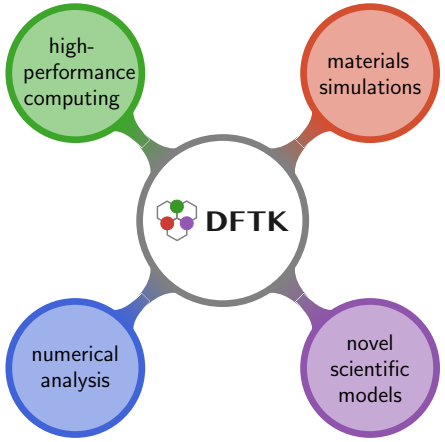
Michael F. Herbst

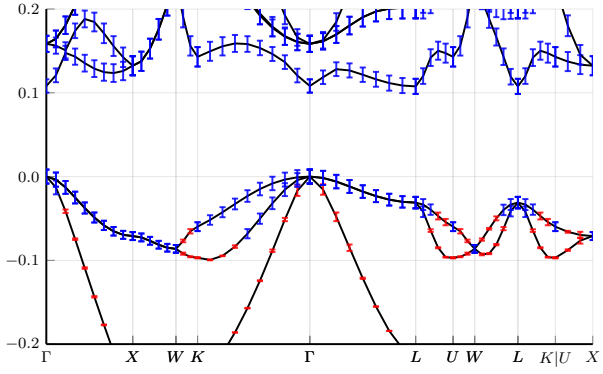
*Applied and Computational Mathematics, RWTH Aachen University
https://michael-herbst.com

12 July 2021

Slides: https://michael-herbst.com/talks/2021.07.12_ssd_materials_discovery.pdf







Electronic structure theory

- Modelling and understanding behaviour of electrons in matter

Electronic structure theory

- Modelling and understanding behaviour of electrons in matter

Materials and semiconductors

Chemical and pharmaceutical industry

Domains and societal challenges

- Renewable energies
- Catalyst design
  - Fertilisation, cleaner chemical processes, ...
- Battery materials
  - Electric cars, energy storage, ...
- Materials for data storage and communication
- Drug discovery
- ...

$\Rightarrow$ Materials discovery as source for innovation

Aren't experiments good enough?

- Experiments are expensive (money, people, time)
- 1 droplet water[1]: $1.7 \cdot 10^{21}$ particles
- Experiments only measure averages
- Sometimes hard to link to physical laws

$\Rightarrow$ Cooperative research of experiment and theory

$\Rightarrow$ Standard practice in industry and research

[1] Assume $0.05\,\mathrm{ml}$.

The need for high-throughput

- Design space is absolutely vast

- E.g. to propose new catalyst, need to consider possibilities in
  - Host materials — 30 transition metals
  - Dopants — 30 transition metals
  - Surface terminations — $\simeq 3 - 5$
  - Reaction intermediates — $\simeq 10$
  - Adsorption configurations — $\simeq 30$

  $\Rightarrow$ $\simeq 10^5 - 10^6$ possibilities!

- State-of-the-art: Density-functional theory (DFT)
- One calculation: $\mathcal{O}$(hours) to $\mathcal{O}$(days)

So can't I just do machine learning?

- Yes (ongoing research), but ...

[1] L. Chanussot *et. al.* ACS Catal. **11** 6059 (2021).

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

So can't I just do machine learning?

- Yes (ongoing research), but ...

- ...I still need DFT as training data
- 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!

[1] L. Chanussot *et. al.* ACS Catal. **11** 6059 (2021).

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

So can't I just do machine learning?

- Yes (ongoing research), but ...

- ...I still need DFT as training data
- 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!

[1] L. Chanussot *et. al.* ACS Catal. **11** 6059 (2021).

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

Why is electronic structure theory so hard?

- Regime of quantum mechanics
- System: Hamiltonian $\hat{\mathcal{H}}$, differential operator
- Minimisation problem: Ground state Ψ with energy

$$E = \min_{\Psi} \int_{\mathbb{R}^{3N}} \Psi\left(\underline{\boldsymbol{r}}_1, \underline{\boldsymbol{r}}_2, \ldots, \underline{\boldsymbol{r}}_N\right) \hat{\mathcal{H}} \Psi\left(\underline{\boldsymbol{r}}_1, \underline{\boldsymbol{r}}_2, \ldots, \underline{\boldsymbol{r}}_N\right) \mathrm{d}\underline{\boldsymbol{r}}_1 \cdots \mathrm{d}\underline{\boldsymbol{r}}_N$$

- Electronic properties: Derivatives of the energy

- Challenge: Size of N, e.g. 2 silicon atoms: $N = 28$
- 2 quadrature points per DOF $\Rightarrow \mathbf{2}^{84} \approx 2 \cdot 10^{25}$ integrand evals

$\Rightarrow$ Finished in 1 year:

Why is electronic structure theory so hard?

- Regime of quantum mechanics
- System: Hamiltonian $\hat{\mathcal{H}}$, differential operator
- Minimisation problem: Ground state Ψ with energy

$$E = \min_{\Psi} \int_{\mathbb{R}^{3N}} \Psi\left(\underline{\boldsymbol{r}}_1, \underline{\boldsymbol{r}}_2, \ldots, \underline{\boldsymbol{r}}_N\right) \hat{\mathcal{H}} \Psi\left(\underline{\boldsymbol{r}}_1, \underline{\boldsymbol{r}}_2, \ldots, \underline{\boldsymbol{r}}_N\right) \mathrm{d}\underline{\boldsymbol{r}}_1 \cdots \mathrm{d}\underline{\boldsymbol{r}}_N$$

- Electronic properties: Derivatives of the energy

- Challenge: Size of N, e.g. 2 silicon atoms: $N = 28$
- 2 quadrature points per DOF $\Rightarrow \mathbf{2}^{84} \approx 2 \cdot 10^{25}$ integrand evals

$\Rightarrow$ Finished in 1 year: ≈ 1.5 attoseconds per eval

Density-functional theory

⇒ Brute force won't cut it

⇒ Cost / quality balance of approximate models ($\simeq 10^3 - 10^4$)

⇒ E.g. density-functional theory (DFT) approximation

- Amongst the most-used family of models
- Effective one-particle model ($N = 3$)
- May construct DFT model for specific context
- Discretisation basis: Build known physics into model
- But: Non-convex, non-linear minimisation

Self-consistent field procedure

- Euler-Lagrange equations (DFT):

$$\begin{cases} \hat{\mathcal{F}}_\rho = -\frac{1}{2}\Delta + \mathcal{V}(\boldsymbol{\rho}) \\ \boldsymbol{\rho}(\underline{\boldsymbol{r}}) = \sum_i f\left(\frac{\varepsilon_i - \varepsilon_F}{T}\right) |\psi_i(\underline{\boldsymbol{r}})|^2 \quad \text{with } \hat{\mathcal{F}}_\rho \psi_i = \varepsilon_i \psi_i, \\ \qquad \text{with } \varepsilon_F \text{ chosen such that } \int \rho \, \mathrm{d}\underline{\boldsymbol{r}} = N, \\ \qquad \text{and Fermi-Dirac function } f(x) = 1/(1+\exp(x)) \end{cases}$$

- Self-consistent field procedure (SCF):
  (1) Guess initial density $\boldsymbol{\rho}$
  (2) Build Kohn-Sham operator $\hat{\mathcal{F}}_\rho$
  (3) Diagonalise it to get new $\{\psi_i\}_i$
  (4) Build new $\boldsymbol{\rho}$ go to (2).

Obstacles for (high-throughput) DFT calculations

- SCF requires nested layers of solvers:
  - Eigenproblem inside fixed-point problem (details later)
  - Algorithms? Preconditioning? Tolerances?
- Accuracy-related parameters chosen by experience
  - Even in automated workflows![1]

⇒ Empirical balance: Accuracy *versus* speed *versus* reliability

- Promising aspects, hardly explored:
  - Error analysis, uncertainty quantification, robust algorithms
  - Modern software engineering techniques
  - Precision reduction, GPGPU, algorithmic differentiation

[1] L. Chanussot *et. al.* ACS Catal. **11** 6059 (2021).

Interdisciplinary field ⇒ Multidisciplinary community

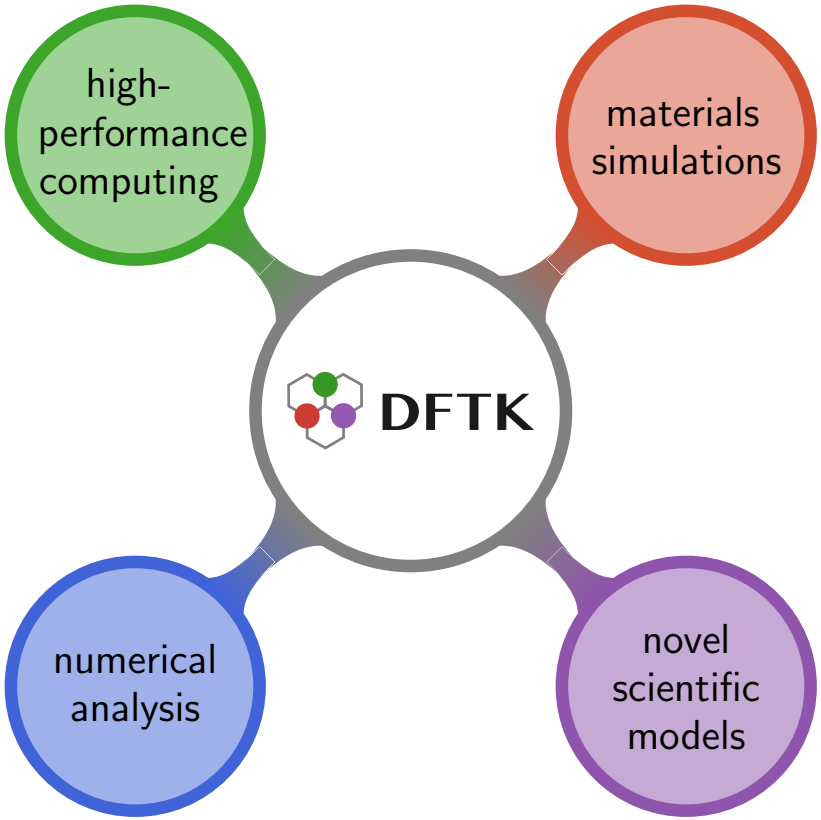
- **Mathematicians:** Toy models and unphysical edge cases
- **High-performance person:** Exploit hardware specialities
- **Scientist:** Design new models, not tweak numerics
- **Practitioner:** Reliable, black-box code, high-level interface

- State-of-the-art DFT codes:
  - Difficult problem ⇒ Complex codes
  - Hard-coded: Workflow / algorithms / hardware optimisations
  - Huge code bases (1M lines and beyond)
  - Non-standard input syntax and API
  - Two-language problem: Algorithmic code hardly accessible

Contents

Contents

Density-functional toolkit (DFTK)[1]



- `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!

- Examples to get going: `https://docs.dftk.org`

[1] M. F. Herbst, A. Levitt and E. Cancès. JuliaCon Proc., **3**, 69 (2021).

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

  ⇒ First get it to work then get it to work *fast*

  ⇒ Write code once generically, then re-use (GPU, sparse data structures, low-precision floats, AD, symbolic methods, . . .)
- `https://michael-herbst.com/learn-julia`

Current research with DFTK

⇒ Vision: Improve high-throughput workflows:

- Use maths: Error analysis and automatic error balancing[1]
- Use physics: Reliable black-box SCF preconditioners[2]

⇒ Reduction of parameters in workflows

⇒ Mathematically motivated / proven algorithms

- International and interdisciplinary user base:
  - ENPC Paris, CMU, MIT, RWTH Aachen, TUM, ...
- Involved in multidisciplinary research projects:
  - ARPA-E's ACED-differentiate, ERC's EMC2 synergy, DOE's Center for the Exascale Simulation of Material Interfaces (CESMIX)

[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).

Contents

Perspective of rigorous error analysis in DFT

- DFT is a complex workflow
  - $\Rightarrow$ Multiple sources of error
    - Choice of model, discretisation, SCF tolerance, diagonalisation tolerance, floating-point arithmetic ...
- Rigorous error bounds:
  - Total error known $\Rightarrow$ Error bars
  - Error-guided adaptive numerics $\Rightarrow$ Efficiency gain
  - Adaptive multi-fidelity methods (Kriging etc.)
- For DFT: Error estimation not fully developed

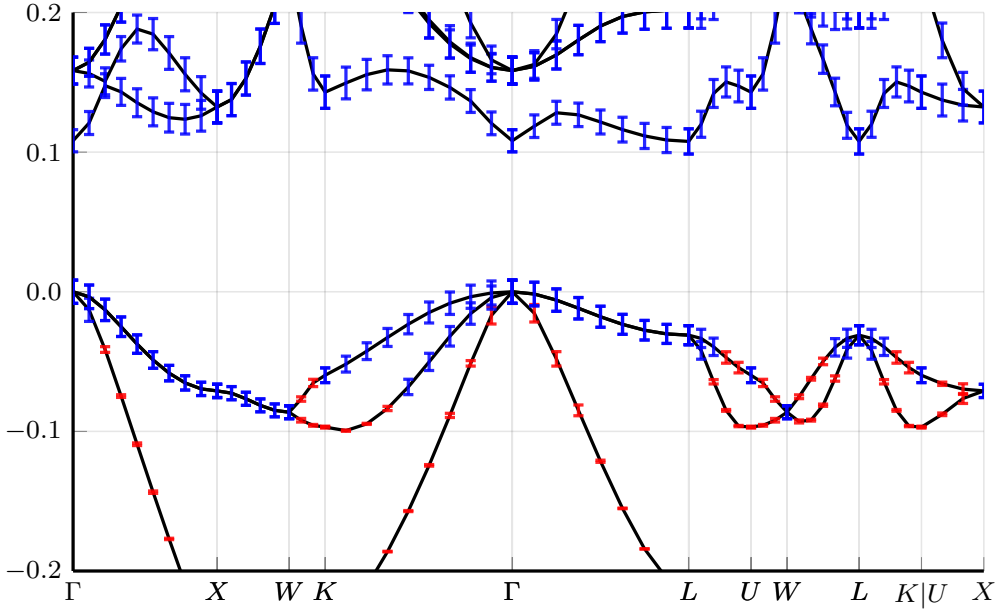
A posteriori error analysis with DFTK

- Requirements:
  - Mathematical theory can only treat reduced models
  - Step-by-step expansion (with guidance from physics)
  - Ingredients not yet clear (e.g. form of integrals for bounds)

  ⇒ Need toolbox for experimentation
- DFTK offers:
  - Fully customisable model
  - Support for arbitrary floating-point types
  - Use julia ecosystem on DFTK datastructures:
    - Numerical quadrature, forward-mode AD, ...

    ⇒ Rapid prototyping in numerical linear algebra

A posteriori error analysis: First results with DFTK[1]



- Reduced model: Non-self-consistent Kohn-Sham
- Estimation of arithmetic error (`IntervalArithmetic.jl`)
- Time to submission: 10 weeks
- Further work currently ongoing . . .

[1] M. F. Herbst, A. Levitt and E. Cancès. Faraday Discuss. **224**, 227 (2020).

Contents

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

- Self-consistent field equations (reminder):

$$\begin{cases} \hat{\mathcal{F}}_\rho = -\frac{1}{2}\Delta + \mathcal{V}(\boldsymbol{\rho}) \\ \boldsymbol{\rho}(\underline{\boldsymbol{r}}) = \sum_i f\left(\frac{\varepsilon_i - \varepsilon_F}{T}\right) |\psi_i(\underline{\boldsymbol{r}})|^2 \end{cases} \quad \text{with } \hat{\mathcal{F}}_\rho \psi_i = \varepsilon_i \psi_i,$$

- Potential-to-density map F

$$F(V) = \sum_{i=1}^{\infty} f\left(\frac{\varepsilon_i - \varepsilon_F}{T}\right) |\psi_i|^2$$

with (ε_i, ψ_i) eigenpairs of $-\frac{1}{2}\Delta + V$.

$\Rightarrow$ Fixed-point problem

$$\rho = F(\mathcal{V}(\rho))$$

Analysis of SCF convergence (1)

- Fixed-point problem $\rho = F(\mathcal{V}(\rho)) \Rightarrow$ Use damped update

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

  with preconditioner ("mixing") P

- Near a fixed-point the error goes as

$$e_{n+1} \simeq \left[1 - \alpha P^{-1} \epsilon^{\dagger} \right] e_n$$

  where $\epsilon^{\dagger} = 1 - \chi_0 K_{\mathsf{Hxc}}$ (dielectric matrix)

- χ_0: Independent-particle susceptibility (derivative of F)
- K_{Hxc}: kernel (derivative of $\mathcal{V}$)

Analysis of SCF convergence (2)

$$e_{n+1} \simeq \left[1 - \alpha P^{-1}\epsilon^{\dagger}\right] e_n$$

- Assume $\sigma(P^{-1}\epsilon^{\dagger}) > 0$ and real (reasonable)
- λ_{min}, λ_{max}: Extremal eigenvalues of $P^{-1}\epsilon^{\dagger}$
- Optimal damping $\alpha = \dfrac{2}{\lambda_{\text{min}} + \lambda_{\text{max}}}$
- Optimal rate $R = \dfrac{\kappa - 1}{\kappa + 1}, \quad \kappa = \dfrac{\lambda_{\text{max}}}{\lambda_{\text{min}}}$

$\Rightarrow$ SCF convergence linked to dielectric properties

- Either $P^{-1} \simeq \left(\epsilon^{\dagger}\right)^{-1}$ or infeasibly small α needed

Analysis of SCF convergence (2)

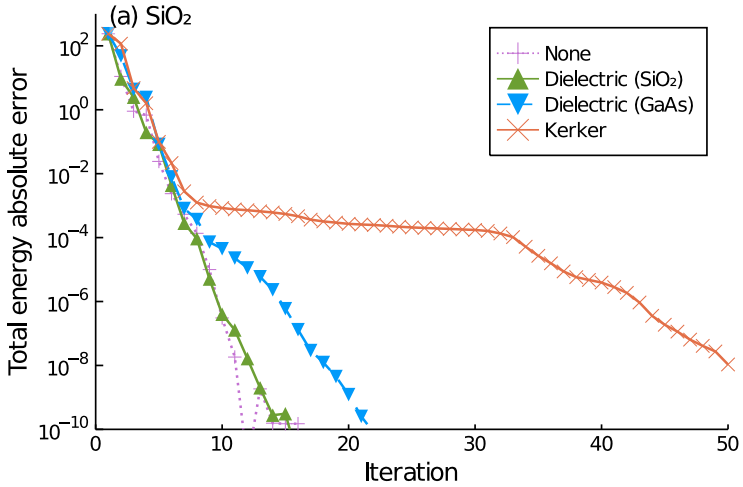
$$e_{n+1} \simeq \left[1 - \alpha P^{-1}\epsilon^{\dagger}\right] e_n$$

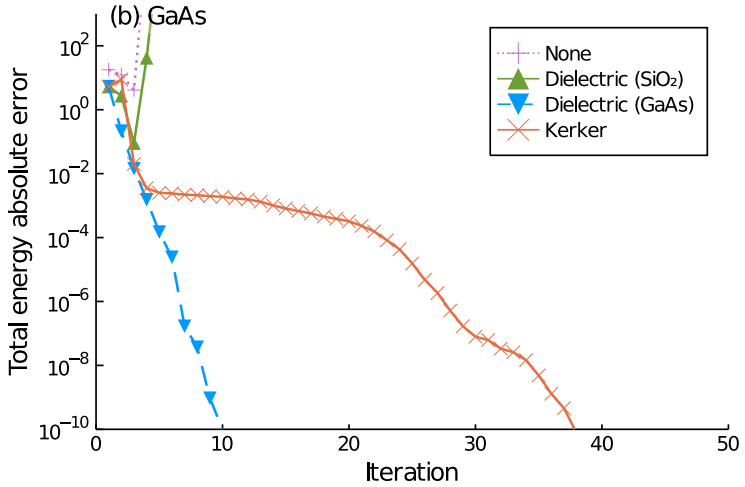
- Assume $\sigma(P^{-1}\epsilon^{\dagger}) > 0$ and real (reasonable)
- λ_{min}, λ_{max}: Extremal eigenvalues of $P^{-1}\epsilon^{\dagger}$
- Optimal damping $\alpha = \dfrac{2}{\lambda_{\text{min}} + \lambda_{\text{max}}}$
- Optimal rate $R = \dfrac{\kappa - 1}{\kappa + 1}, \quad \kappa = \dfrac{\lambda_{\text{max}}}{\lambda_{\text{min}}}$

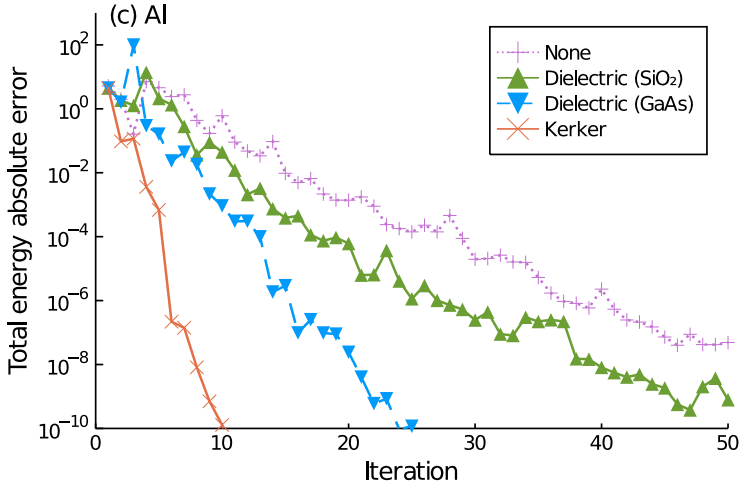
$\Rightarrow$ SCF convergence linked to dielectric properties

- Either $P^{-1} \simeq \left(\epsilon^{\dagger}\right)^{-1}$ or infeasibly small α needed

Convergence results for bulk materials[1]







- silica (SiO_2) insulator
- gallium arsenide (GaAs) semiconductor
- aluminium (Al) metal

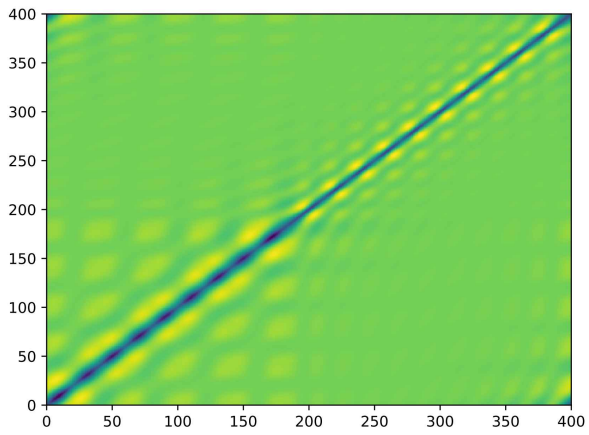
[1] M. F. Herbst, A. Levitt. J. Phys. Condens. Matter **33**, 085503 (2021).

Problems with established mixing methods

- Preconditioner P is system-dependent, but manually chosen
  - Rough idea of dielectric properties needed *a priori*
- Good preconditioners only known for bulk materials
  - Misses important applications (e.g. inhomogeneous systems)
  - Examples: Metal clusters, surfaces, d-metal alloys, ...
- Damping α manually chosen
  - Trial and error ... (especially if P unsuitable)
- Can we do better?

Preconditioning inhomogeneous systems

- Bulk preconditioning models tackle directly $P^{-1} \approx \left(\varepsilon^{\dagger}\right)^{-1}$
- Plot of χ_0 (Chain of 10 Sodium atoms and 10 helium atoms):



$\Rightarrow$ Diagonal-dominant, try an approximation $\chi_0(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') \simeq \widetilde{\chi_0}(\underline{\boldsymbol{r}})$

- Apply preconditioner iteratively:

$$P^{-1}\rho_n = \left(1 - \widetilde{\chi_0} K_{\mathsf{Hxc}}\right)\right)^{-1} \rho_n$$

Local density of states (LDOS) approximation for χ_0[1]

- Main interest: Large-scale variations from ρ_n to ρ_{n+1}

$\Rightarrow$ Assume $\underline{\boldsymbol{r}} \mapsto \chi_0(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}')$ more localised around $\underline{\boldsymbol{r}}'$ than $V(\underline{\boldsymbol{r}}')$.

- "Row-sum mass lumping":

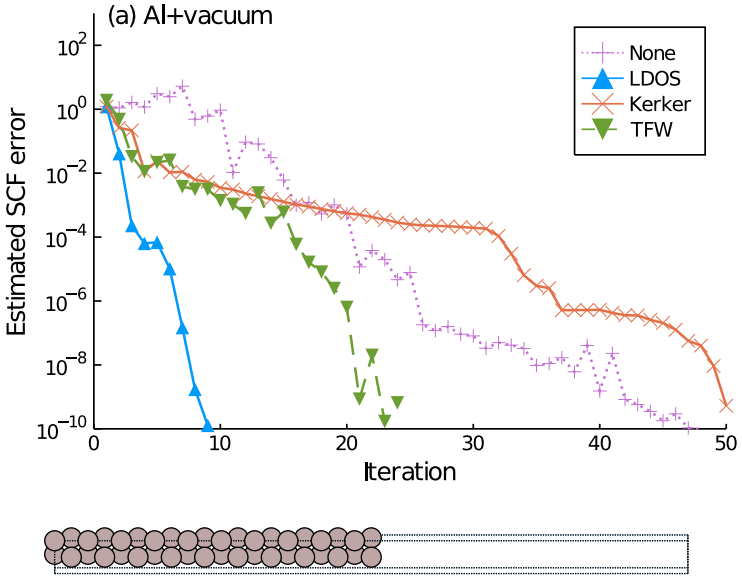
$$\int \chi_0(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') V(\underline{\boldsymbol{r}}') \, \mathrm{d}\underline{\boldsymbol{r}}' \simeq V(\underline{\boldsymbol{r}}) \int \chi_0(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') \, \mathrm{d}\underline{\boldsymbol{r}}'$$
$$= -V(\underline{\boldsymbol{r}}) D_{\mathsf{loc}}(\underline{\boldsymbol{r}})$$

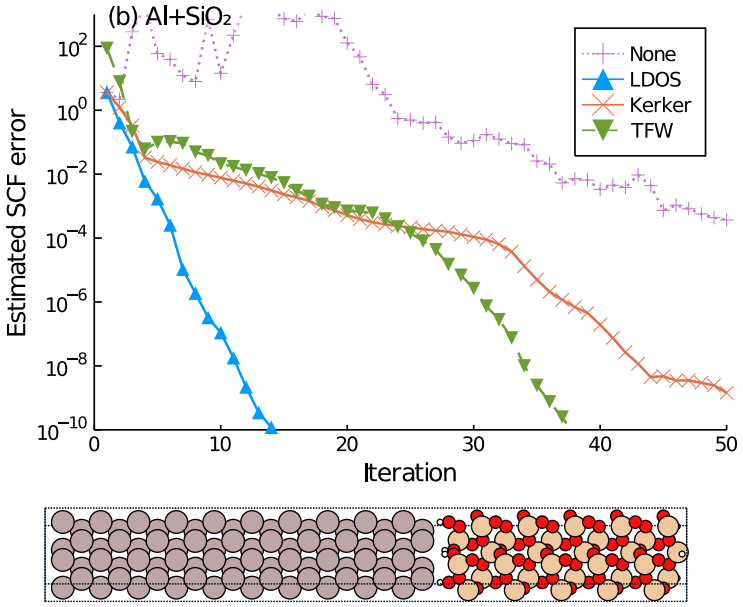
with local density of states

$$D_{\mathsf{loc}}(\underline{\boldsymbol{r}}) = \frac{1}{T} \sum_i f'\left(\frac{\varepsilon_i - \varepsilon_F}{T}\right) |\psi_i(\underline{\boldsymbol{r}})|^2$$

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

LDOS preconditioning (examples)





- 20 repeats of aluminium + 20 repeats vacuum / silica
- TFW: local Thomas-Fermi-von Weizsäcker mixing[1]
- LDOS automatically interpolates between Kerker mixing (in the metallic region) and no mixing (insulating region)

⇒ Parameter-free and black-box

[1] D. Raczkowski, A. Canning, L. W. Wang, Phys. Rev. B. **64**, 121101 (2001).

LDOS preconditioning results[1]

		None		Dielectric		Kerker		LDOS		LDOS+ Dielectric	
	$\mathcal{N}$	it	κ	it	κ	it	κ	it	κ	it	κ
SiO_2+vacuum	10	11	3.3	26	19.7	50	95.7	11	3.3	26	19.7
	20	12	3.4	30	24.4	n.c.	351.5	12	3.4	30	21.7
GaAs+vacuum	10	17	13.4	18	6.2	23	67.0	17	12.4	18	10.4
	20	20	15.5	22	12.9	n.c.	312.2	20	15.5	22	12.9
Al+vacuum	10	19	51.5	24	44.3	22	64.4	9	3.7	16	10.3
	20	47	170.8	49	168.5	n.c.	323.9	9	3.5	20	10.5
GaAs+SiO_2[a]	10	45	13.7	19	8.9	34	52.4	45	13.4	19	8.8
	20	n.c.	18.2	20	10.2	n.c.	170.1	n.c.	18.2	20	10.2
Al+SiO_2	10	43	93.1	29	33.6	30	50.9	17	6.1	20	9.2
	20	n.c.	316.6	n.c.	118.4	n.c.	159.4	14	5.4	20	10.1
Al+GaAs	10	n.c.	144.0	24	22.4	16	9.0	15	7.2	11	3.5
	20	n.c.	485.0	40	59.0	26	28.8	26	21.4	13	5.0
Al+GaAs+SiO_2	10	n.c.	149.5	34	50.4	36	62.9	26	21.5	19	9.0

- Coloured: Condition number κ less than doubled on doubling system size

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

Now . . . how about the damping?

- LDOS solves:
  - SCF issues for large inhomogeneous metal-insulator systems
  - Does *not* prevent all causes of bad SCF conditioning
  - (more research on this is needed!)
- Practitioners trial and error with damping α
- Convergence is guaranteed if damping α small enough

$\Rightarrow$ Adaptive damping strategy

Adaptive damping

- Potential mixing:

$$V_{n+1} = V_n + \alpha \delta V_n$$
$$\delta V_n = P^{-1}\left[\mathcal{V}(F(V_n)) - V_n\right]$$

- Quadratic model for DFT energy:

$$E(V_n + \alpha\,\delta V_n) \simeq E(V_n) + \alpha \left\langle \nabla E_{|V=V_n} \middle| \delta V_n \right\rangle + \frac{\alpha^2}{2} \left\langle \delta V_n \middle| \nabla^2 E_{|V=V_n} \delta V_n \right\rangle$$

- After some algebra:

$$\nabla E_{|V=V_n} = -\chi_0 \left(V_n^{\text{out}} - V_n\right)$$
$$\nabla^2 E_{|V=V_n} \simeq -\chi_0 \left[1 - K_{\text{Hxc}} \chi_0\right]$$

$\Rightarrow$ Use model to find damping automatically!

Adaptive damping

- Potential mixing:

$$V_{n+1} = V_n + \alpha\, \delta V_n$$
$$\delta V_n = P^{-1}\left[V_n^{\mathsf{out}} - V_n\right], \qquad V_n^{\mathsf{out}} = \mathcal{V}(F(V_n))$$

- Quadratic model for DFT energy:

$$E(V_n + \alpha\, \delta V_n) \simeq E(V_n) + \alpha \left\langle \nabla E_{|V=V_n} \middle| \delta V_n \right\rangle + \frac{\alpha^2}{2} \left\langle \delta V_n \middle| \nabla^2 E_{|V=V_n} \delta V_n \right\rangle$$

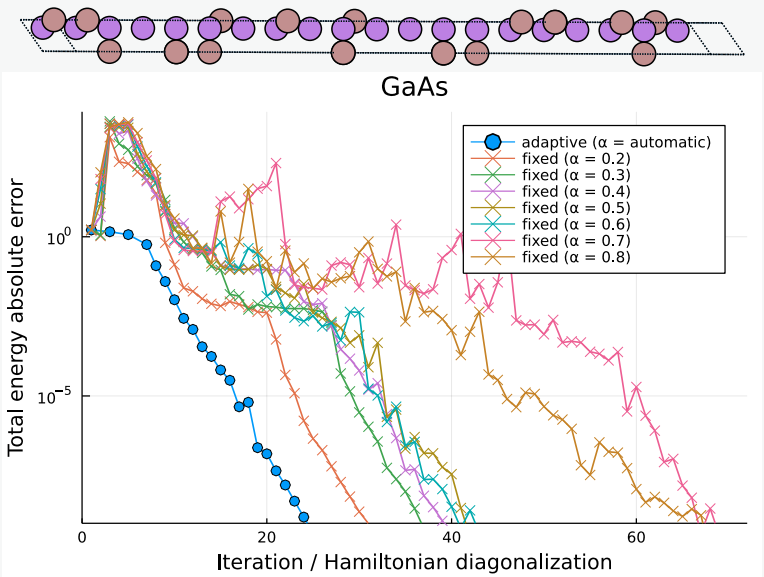
- After some algebra:

$$\nabla E_{|V=V_n} = -\chi_0 \left(V_n^{\mathsf{out}} - V_n\right)$$
$$\nabla^2 E_{|V=V_n} \simeq -\chi_0 \left[1 - K_{\mathsf{Hxc}} \chi_0\right]$$

$\Rightarrow$ Use model to find damping automatically!

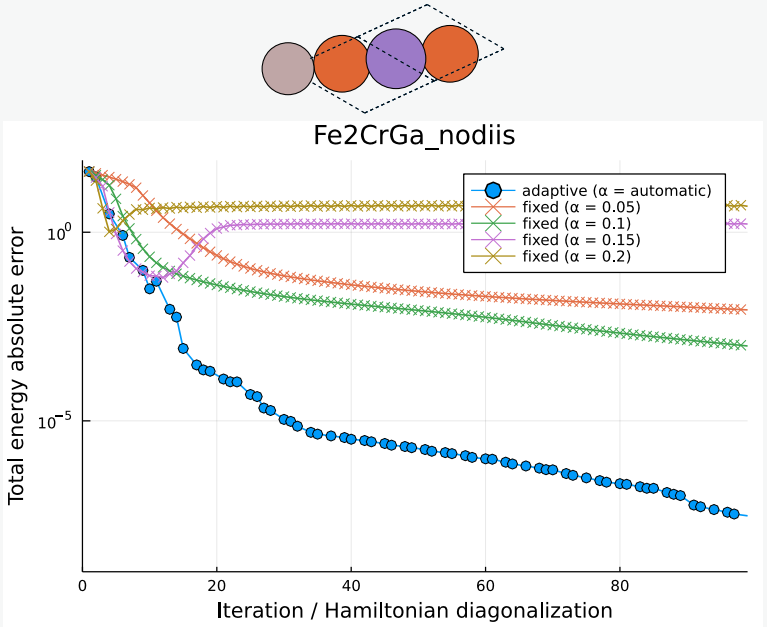
Adaptive damping (WIP examples)

GaAs (Gallium arsenide)



- No preconditioner ($P = I$)
- Non-linear SCF behaviour in initial steps

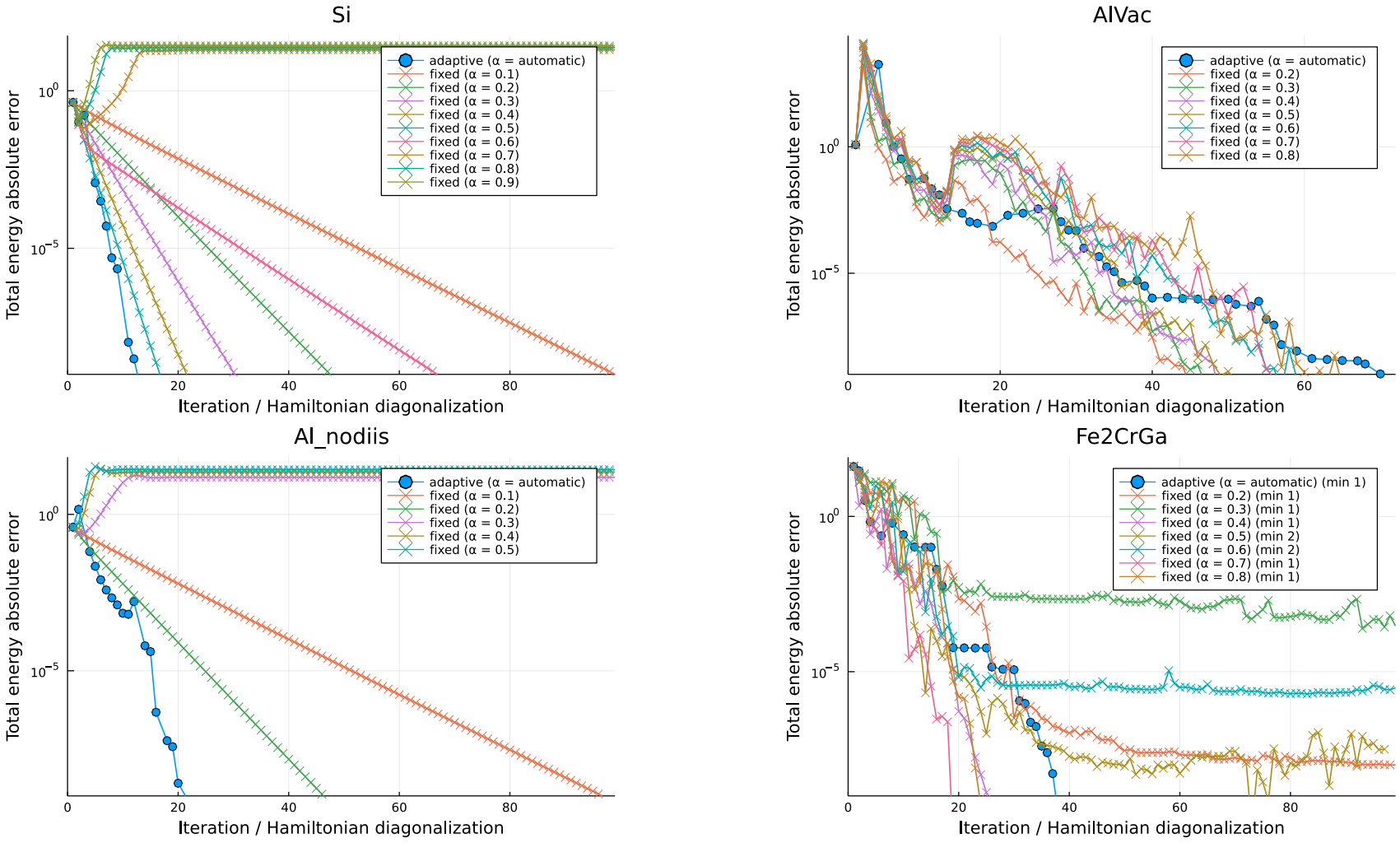
Fe_2CrGa Heusler alloy (no Anderson)



- Unsuitable Kerker preconditioner
- Localised states, spin

- Adaptive damping as safeguard
- Ensures energy / residual decrease

Adaptive damping (WIP overview)



Summary

- High-throughput screening
  - Key Ingredients to design the materials of tomorrow
  - Main obstacle: Large number of parameters
  - Chosen empirically ⇒ Reliability limited
- **DFTK**: Multidisciplinary software development
  - julia-based framework for new DFT algorithms
  - Toy problems and scale-up to realistic applications
- LDOS preconditioner:
  - Parameter-free ⇒ Highly suitable for high-throughput
  - Adaptive preconditioning for inhomogeneous systems
- Adaptive damping scheme:
  - Safe guard if preconditioner not perfect / tricky system
  - Reduction of the human factor

Upcoming Juliacon workshop

"A mathematical look at electronic structure theory"

22. July 2021 16:00 CEST

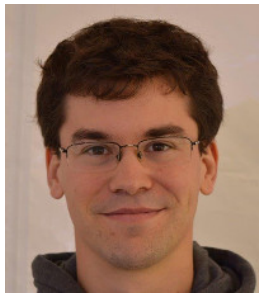
juliacon

`https://juliacon.org/2021/` (free)

. . . and afterwards recorded on Youtube

Acknowledgements

https://michael-herbst.com/talks/2021.07.12_ssd_materials_discovery.pdf



Antoine Levitt

Eric Cancès

Benjamin Stamm

all DFTK contributors

Questions?

https://michael-herbst.com/talks/2021.07.12_ssd_materials_discovery.pdf

DFTK https://dftk.org

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

mfherbst

https://michael-herbst.com/blog

herbst@acom.rwth-aachen.de