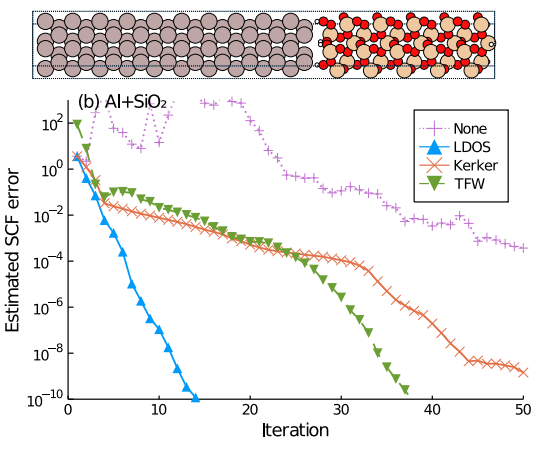
Reliable black-box self-consistent field schemes for high-throughput DFT calculations

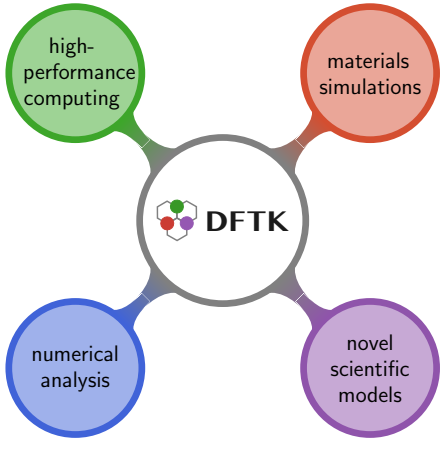
Michael F. Herbst

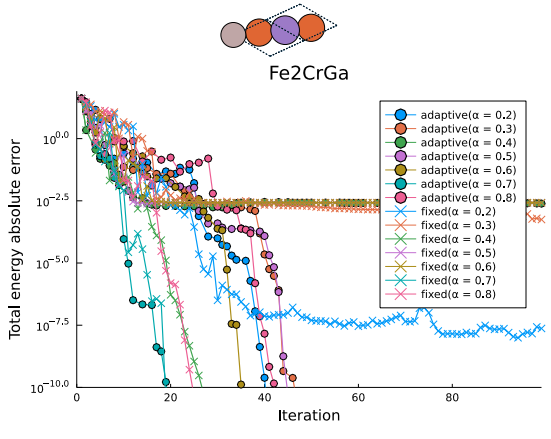
*Applied and Computational Mathematics, RWTH Aachen University
https://michael-herbst.com

29 June 2021

Slides: https://michael-herbst.com/talks/2021.06.29_qmb_reliable_scf.pdf







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

$\Rightarrow$ Starting point: Density-functional theory (DFT)

- Even beyond-DFT methods need DFT solutions

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

$\Rightarrow$ Starting point: Density-functional theory (DFT)

- Even beyond-DFT methods need DFT solutions

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

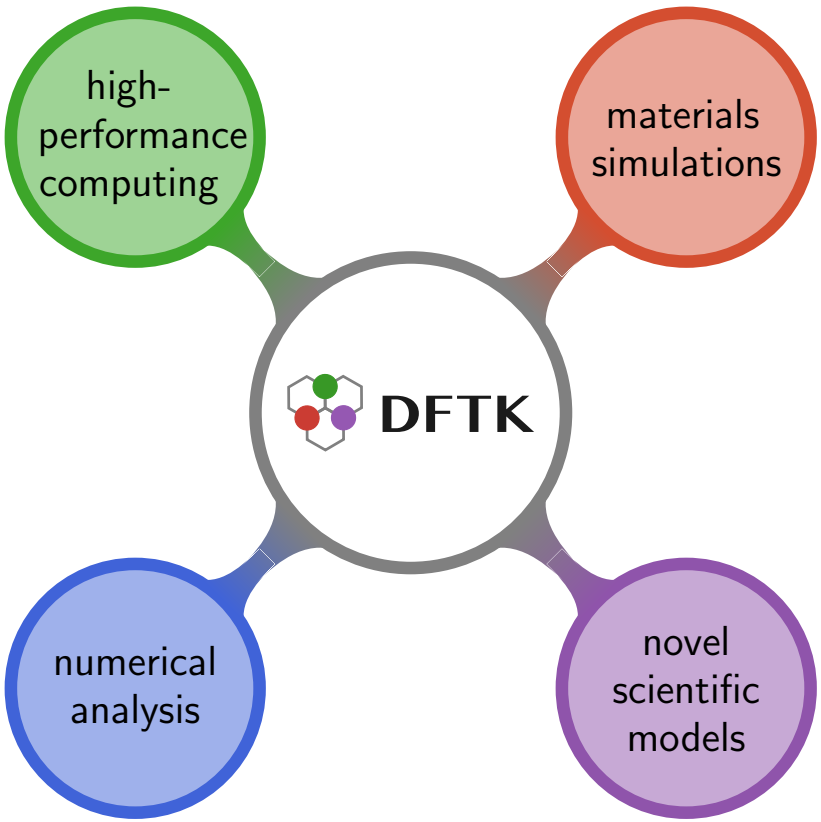
$\Rightarrow$ Reliability needs to be improved!

- 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)[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!

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

DFTK — https://dftk.org

- Ground state (LDA, GGA) and a bit of response theory
- Compose your model (e.g. analytic potentials, . . .)
- 1D / 2D / 3D systems
- Arbitrary floating point type

- Mixed MPI-Thread-based parallelism
- > 800 electrons possible
- Integration with materials-related python modules

- Performance: Within factor 2 of established codes
- Documentation and examples: `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 maths: Error analysis and automatic error balancing[1]
- Use physics: Reliable black-box SCF preconditioners[2]
- Better algorithms: Numerical analysis of SCF methods

⇒ Reliable automatic selection of parameters in DFT workflows

- 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).

[3] E. Cancès, G. Kemlin, A. Levitt. arXiv 2004.09088 (2020).

Upcoming Juliacon workshop

"A mathematical look at electronic structure theory"

22. July 2021 16:00 CEST

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

. . . and afterwards recorded on Youtube

Upcoming Juliacon workshop

"A mathematical look at electronic structure theory"

22. July 2021 16:00 CEST

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

. . . and afterwards recorded on Youtube

Perspective of rigorous error analysis in DFT

- Sources of error in DFT:
  - Model error
  - Discretisation error
  - Algorithm error
  - Arithmetic error
- *A posteriori* bound on errors:
  - 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

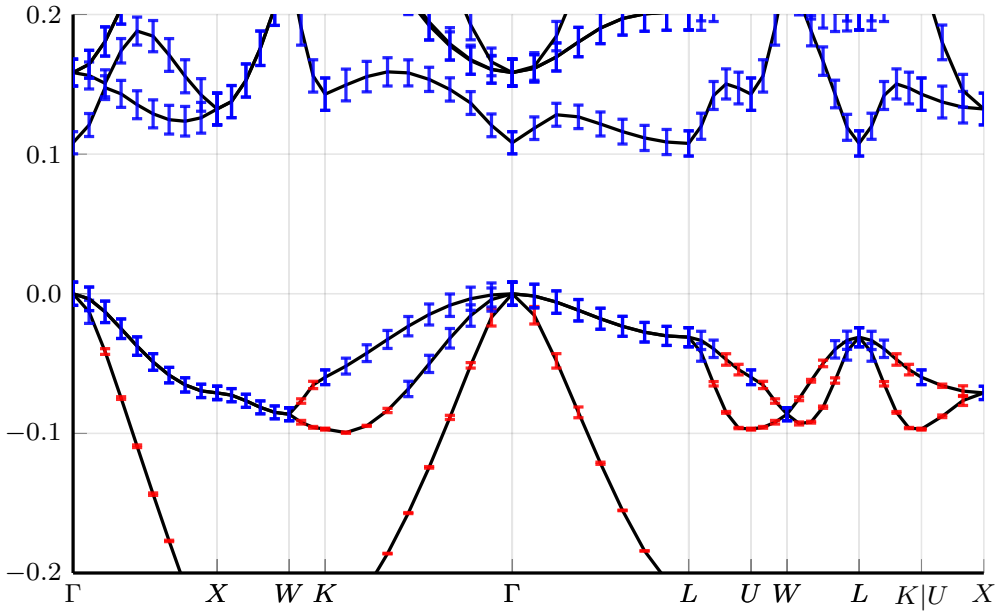
A posteriori error analysis with DFTK

- 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)

  $\Rightarrow$ 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, ...

    $\Rightarrow$ 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).

Kohn-Sham DFT

- Coupled set of non-linear elliptic partial differential equations:

$$\Big(-\frac{1}{2}\Delta + \mathcal{V}(\rho)\Big)\psi_i = \varepsilon_i \psi_i, \quad \int \psi_i^* \psi_j = \delta_{ij}$$

$$\rho = \sum_{i=1}^{\infty} f\left(\frac{\varepsilon_i - \varepsilon_F}{T}\right) |\psi_i|^2, \quad \text{with } \varepsilon_F \text{ such that } \int \rho = N$$

- Density-dependent potential

$$\mathcal{V}(\rho) = V_{\text{Nuc}} + \int (v_C \rho) + V_{\text{xc}}(\rho)$$

with Coulomb kernel $v_C(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') = \|\underline{\boldsymbol{r}} - \underline{\boldsymbol{r}}'\|^{-1}$

- Fermi-Dirac function $f(x) = 1/(1 + e^x)$
- Temperature T, electron count N
- Exchange-correlation potential V_{xc}, nuclear attraction V_{Nuc}

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

- Density-dependent potential

$$\mathcal{V}(\rho) = V_{\text{Nuc}} + \int (v_C \rho) + V_{\text{xc}}(\rho)$$

- 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 $-\frac{1}{2}\Delta + V$.

$\Rightarrow$ SCF solves $\rho = F(\mathcal{V}(\rho))$

- Numerically: Damped fixed-point scheme

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

with preconditioner ("mixing") P

SCF convergence (1)

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

- Near a fixed-point the error goes as

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

  where $\epsilon^\dagger = 1 - \chi_0 (v_C + f_{\text{xc}})$ (dielectric matrix)
- χ_0: Independent-particle susceptibility (derivative of F)
- f_{xc}: XC kernel (derivative of V_{xc})

SCF convergence (2)

$$e_{n+1} \simeq \left[1 - \alpha P^{-1} \epsilon^\dagger\right] e_n, \qquad \epsilon^\dagger = 1 - \chi_0 (v_C + f_{\text{xc}})$$

- Assume RPA: $\sigma(\epsilon^\dagger) > 0$ and real
- Assume $P > 0$
- 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

SCF instabilities

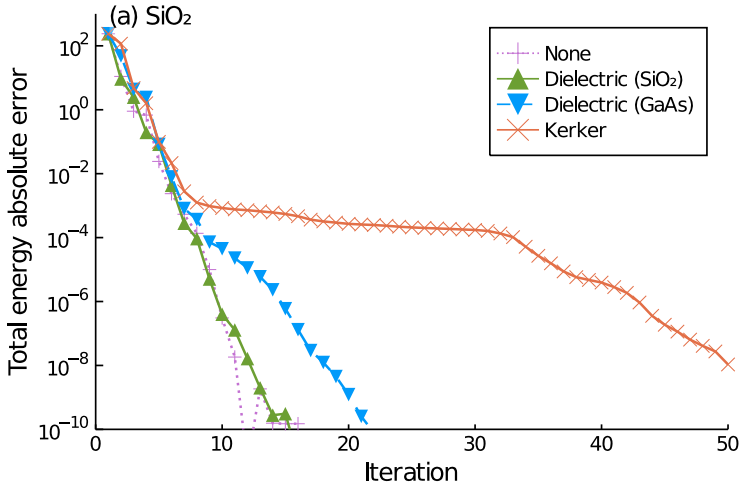
$$e_{n+1} \simeq \left[1 - \alpha P^{-1} \epsilon^\dagger\right] e_n, \qquad \epsilon^\dagger = 1 - \chi_0 (v_C + f_{\text{xc}})$$

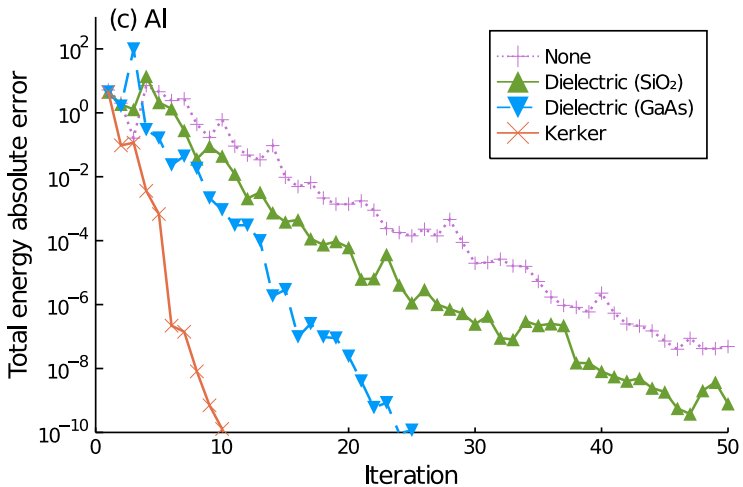
- SCF instabilities increase condition number κ:
  - $\epsilon^\dagger$ has small eigenvalues (e.g. symmetry breaking)
  - χ_0 has large eigenvalues (localised states)
  - Large charge-sloshing modes of v_C are uncompensated by χ_0 (metals)

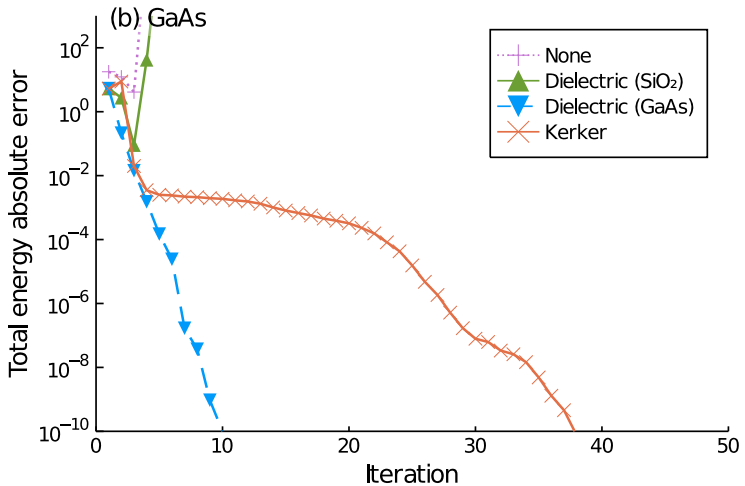
$\Rightarrow$ Need infeasibly small α or good P

- Physics where a good mixing P is known:
  - *Bulk* insulators ($P = I$)
  - *Bulk* metals (Kerker mixing)
  - *Bulk* semiconductors (e.g. Resta's dielectric model)

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).

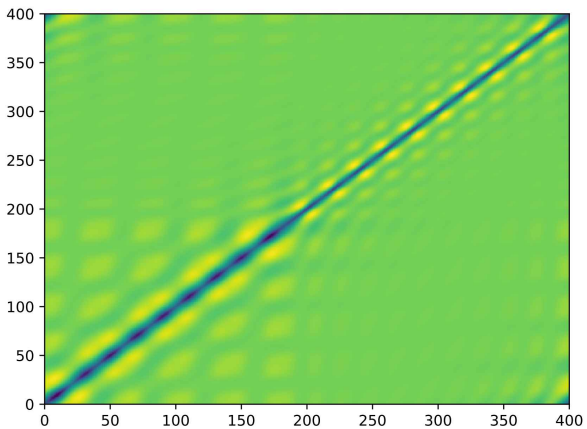
Problems with established mixing methods

- Preconditioner P manually chosen
  - Rough idea of dielectric properties needed *a priori*
- Which P for inhomogeneous materials?
  - Either bulk or non-trivial parameters
  - E.g. metal clusters, metallic surfaces, passivated surfaces
- Which P for localised states?
  - None I know of ...
  - E.g. d-metal alloys (Heuslers), surface states
- Damping α manually chosen
  - Trial and error ... (especially if P unsuitable)

Preconditioning inhomogeneous systems

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

- 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:

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

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":

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

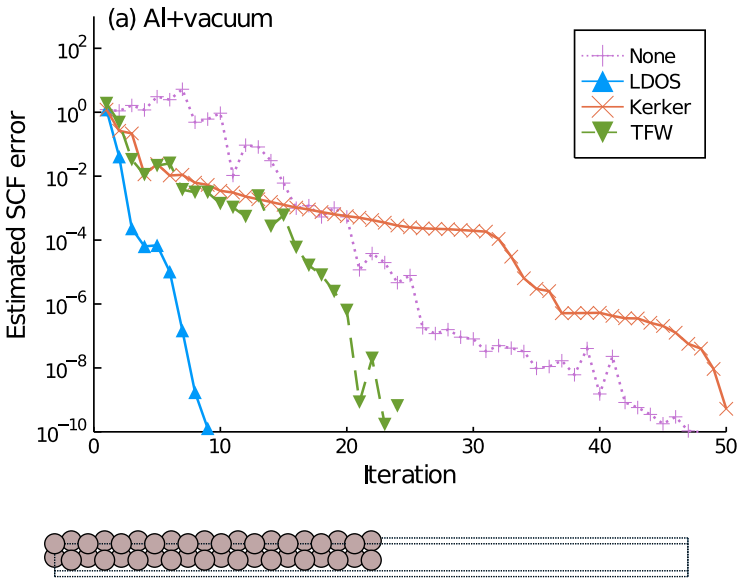
with local density of states

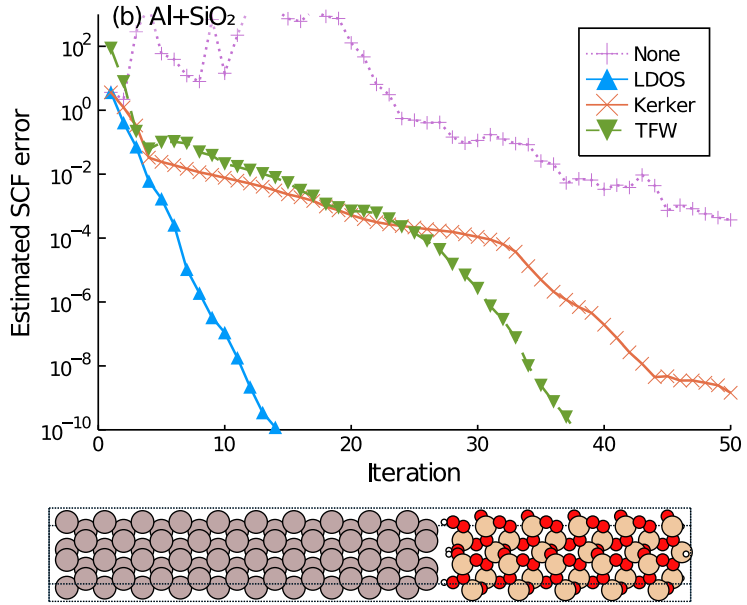
$$D_{\mathsf{loc}}(\underline{\boldsymbol{r}}) = \frac{1}{T} \sum_i f'\left(\frac{\varepsilon_i - \varepsilon_F}{T}\right) |\psi_i(\underline{\boldsymbol{r}})|^2$$

using Adler-Wiser formula.

[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)

$\Rightarrow$ Parameter-free and black-box

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

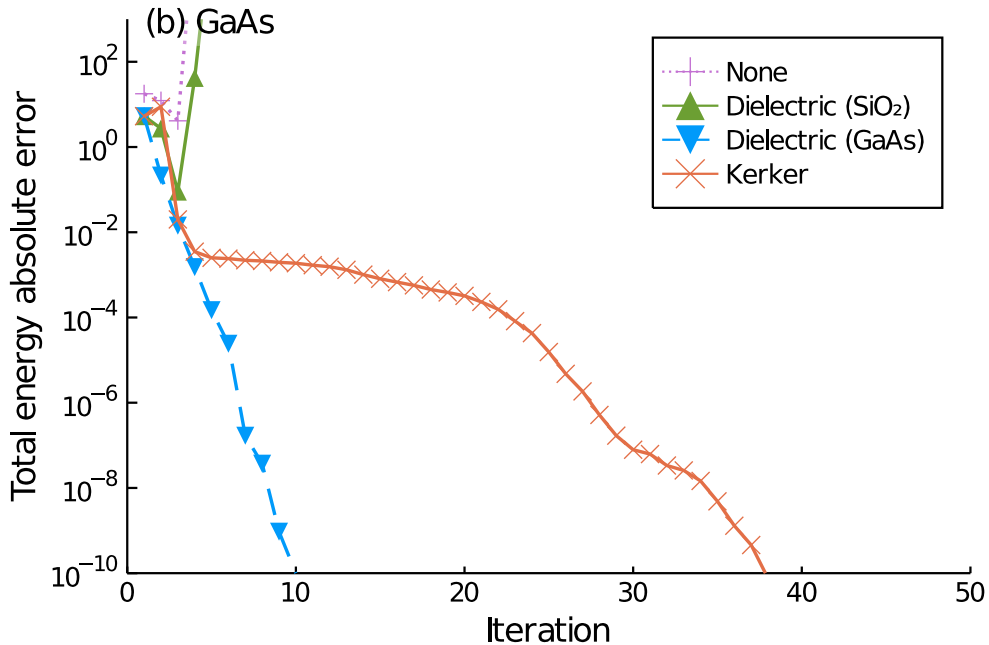
LDOS preconditioning results[1]

		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).

Revisiting a "simple" case



- In initial phase exhibits strong non-linearity

$\Rightarrow$ Anderson extrapolates very far off

- Practitioners trial and error with damping α
- Convergence is guaranteed if damping α small enough

$\Rightarrow$ Adaptive damping strategy

Adaptive damping (1)

- Potential mixing:

$$V_{n+1} = V_n + \delta V_n$$

$$\delta V_n = \mathrm{DIIS}\Big(\alpha P^{-1}\left[\mathcal{V}(F(V_n)) - V_n\right]\Big)$$

- Quadratic model for DFT energy:

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

- Following[1]

$$\nabla E_{|V=V_n} = -\chi_0 \left(V_n^{\mathrm{out}} - V_n\right)$$

$$\nabla^2 E_{|V=V_n} \simeq -\chi_0 \left[1 - (v_C + f_{\mathrm{xc}})\chi_0\right]$$

[1] X. Gonze Phys. Rev. B **54**, 4383 (1996).

Adaptive damping (1)

- Potential mixing:

$$V_{n+1} = V_n + \alpha\, \delta V_n$$

$$\delta V_n = \mathrm{DIIS}\Big(\tilde{\alpha} P^{-1} \left[V_n^{\mathrm{out}} - V_n\right]\Big)/\tilde{\alpha}, \qquad V_n^{\mathrm{out}} = \mathcal{V}(F(V_n))$$

- Quadratic model for DFT energy:

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

- Following[1]

$$\nabla E_{|V=V_n} = -\chi_0 \left(V_n^{\mathrm{out}} - V_n\right)$$

$$\nabla^2 E_{|V=V_n} \simeq -\chi_0 \left[1 - \left(v_C + f_{\mathrm{xc}}\right)\chi_0\right]$$

[1] X. Gonze Phys. Rev. B **54**, 4383 (1996).

Adaptive damping (2)

$$V_{n+1} = V_n + \alpha\,\delta V_n, \quad \delta V_n = \mathrm{DIIS}\Big(\tilde{\alpha} P^{-1}\left[V_n^{\mathrm{out}} - V_n\right]\Big)/\tilde{\alpha}, \quad V_n^{\mathrm{out}} = \mathcal{V}(F(V_n))$$

- Quadratic model (after some algebra):

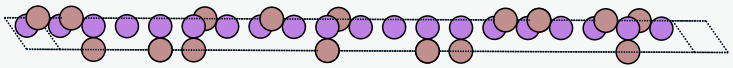
$$E(V_n + \alpha\,\delta V_n) \simeq E(V_n) - \alpha\,\langle V_n^{\mathrm{out}} - V_n | \delta\rho_n\rangle + \frac{\alpha^2}{2}\Big[- \langle \delta V_n | \delta\rho_n\rangle + \Big\langle \delta\rho_n \Big| \Big(v_C + f_{\mathrm{xc}}\Big)\,\delta\rho_n \Big\rangle \Big]$$

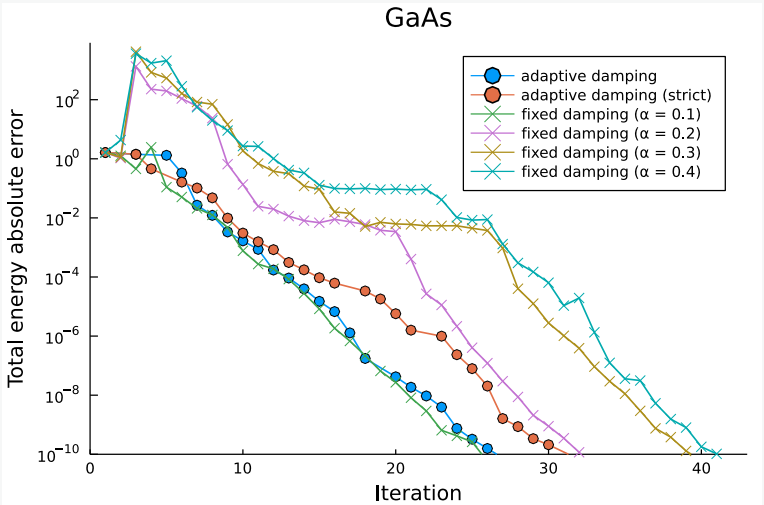
  where $\delta\rho_n = F(V_{n+1}) - F(V_n)$.

- Given $\left[\tilde{\alpha}, V_n\right] \to \left[F(V_n), V_n^{\mathrm{out}}\right] \to \left[\delta V_n, V_{n+1}\right] \to F(V_{n+1})$ then find optimal damping α
- (Simplified) sketch of adaptive damping algorithm:
  - Choose trial $\tilde{\alpha} = \alpha$
  - Accept if energy or residual decreases
  - Else: Find optimal damping α, recompute V_{n+1} and $F(V_{n+1})$

Adaptive damping (WIP examples)

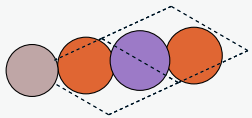
GaAs (Gallium arsenide)

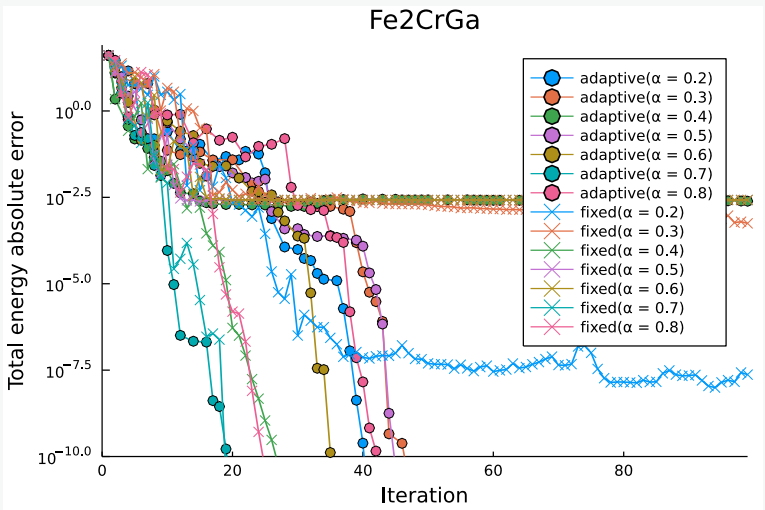




- Non-linear SCF behaviour in initial steps
- Anderson extrapolation fails

Fe_2CrGa Heusler alloy structure





- Localised states, spin
- No suitable preconditioner P

- Adaptive damping as safeguard
- Ensures energy / residual decrease

Adaptive damping (WIP overview)

- Convergence in the energy to 10^{-8} for tricky Heuslers

System	adaptive							standard						
	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.2	0.3	0.4	0.5	0.6	0.7	0.8
CoFeMnGa	40	30	31	36	20	22	36	n.c.	26	23	21	19	20	26
Fe_2CrGa	39	44	n.c.	44	35	17	41	94	n.c.	24	n.c.	n.c.	19	23
Fe_2MnAl	30	n.c.	n.c.	52	22	24	32	41	n.c.	n.c.	n.c.	18	19	15
Mn_2RuGa	n.c.	51	38	27	22	23	26	n.c.	n.c.	n.c.	25	21	21	20
Mn_3Si	121	27	54	76	48	22	28	n.c.	n.c.	n.c.	22	24	17	18

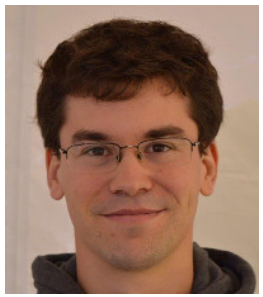
Summary

https://michael-herbst.com/talks/2021.06.29_qmb_reliable_scf.pdf

- **DFTK** usage:
  - First develop LDOS scheme on test systems (1D, toy problems)
  - Tests on > 800 electrons (in the same code!)
  - Key quantities (χ_0, f_{xc}) fully accessible
  - E.g. routine computation of λ_{min} and λ_{max}
- LDOS preconditioner:
  - Parameter-free $\Rightarrow$ 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

Acknowledgements

https://michael-herbst.com/talks/2021.06.29_qmb_reliable_scf.pdf



Antoine Levitt

Eric Cancès

Benjamin Stamm

all DFTK contributors

Questions? https://michael-herbst.com/talks/2021.06.29_qmb_reliable_scf.pdf


DFTK https://dftk.org


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


mfherbst


https://michael-herbst.com/blog


herbst@acom.rwth-aachen.de