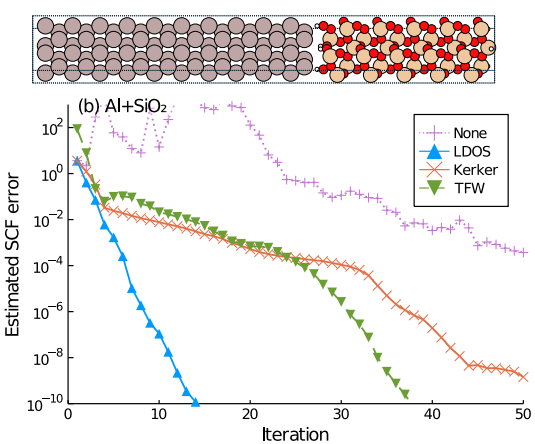
Towards error-controlled, black-box density-functional theory methods

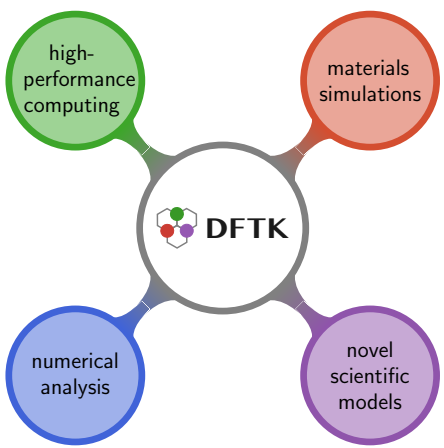
Michael F. Herbst*, Antoine Levitt

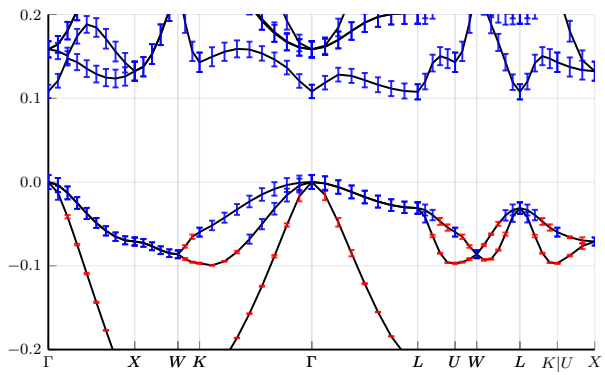
*Applied and Computational Mathematics, RWTH Aachen University
https://michael-herbst.com

21 July 2021

Slides: https://michael-herbst.com/slides/vmd2021







Motivation: Computational challenges

- Virtual materials design $\Rightarrow$ millions of calculations:
  - design space search, data generation for surrogates, ...
- Key requirements:
  - Automation (saves human time)
  - Efficiency (saves computer time)
  - Reliability (saves computer time & human time)
- State of the art:
  - Many parameters to choose (algorithms, tolerances, models)
  - Choice by experience
  - Workflow success rate: $\simeq 50\%$[1]
  - Little rigorous error control (basically trial and error)

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

Some questions and hardly explored techniques

- Leading questions:
  - Do we need the same accuracy everywhere?
  - Where can mathematical insight improve reliability?
  - Can we tune between accuracy and runtime using only a single parameter?
- Error estimation (*a posteriori*, UQ, sensitivity analysis)

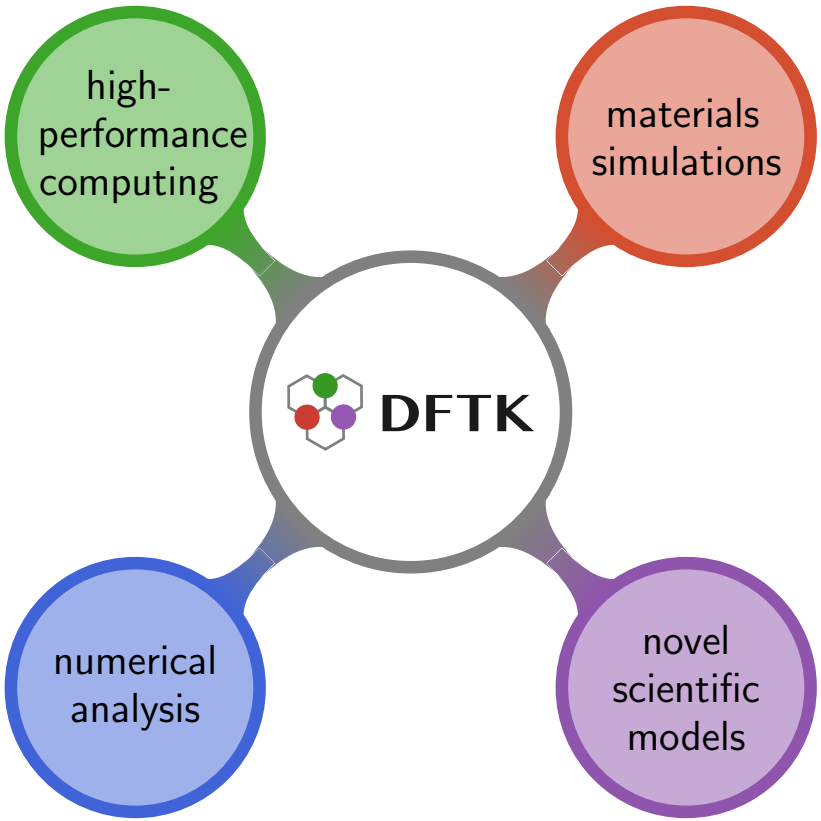
  $\Rightarrow$ Error-guided automatic selection of parameters
- Numerical analysis of simulation algorithms

  $\Rightarrow$ Towards black-box algorithms
- Multi-fidelity methods

  $\Rightarrow$ Combination of results of different quality (functionals, numerics, corrections, . . .)

$\Rightarrow$ Multidisciplinary research setting

Density-functional toolkit (DFTK)[1]



- `https://dftk.org`
- 2 years of development
- Open-source julia code
- Building on julia ecosystem

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

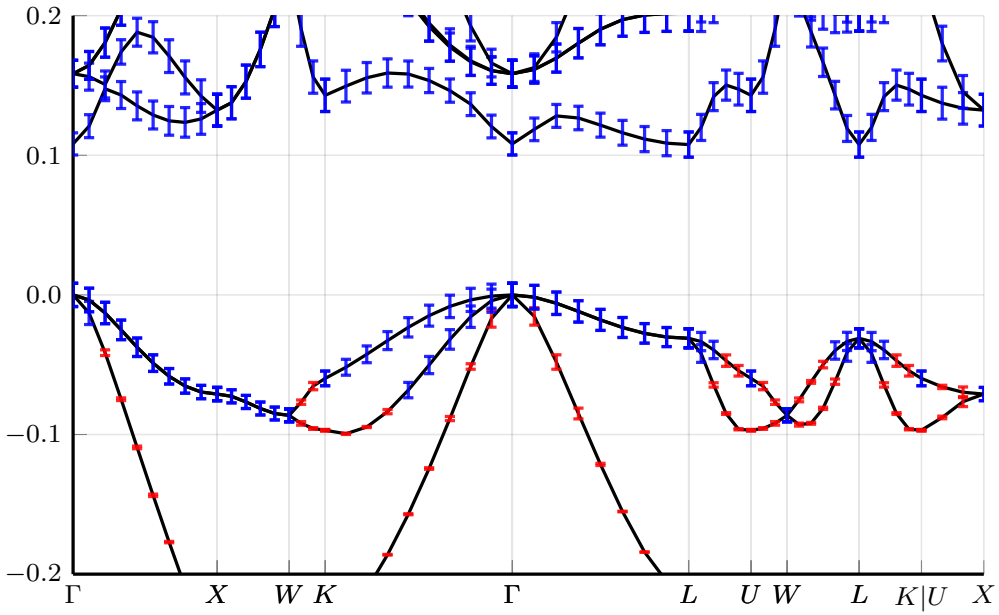
[1] M. F. Herbst, A. Levitt and E. Cancès. JuliaCon Proc., **3**, 69 (2021).

- Documentation and examples: `https://docs.dftk.org`
- Ground state (LDA, GGA) and a bit of response theory
- Compose your model (e.g. analytic potentials, 1D / 2D, . . .)
- Arbitrary floating point type / reduced precision
- Automatic differentiation (e.g. stresses, sensitivities)

- Mixed MPI-Thread-based parallelism
- > 800 electrons possible
- Performance: Within factor 2–4 of established codes

- Involved in multiple multidisciplinary collaborations:
  - ARPA-E's ACED-differentiate, ERC's EMC2 synergy,
    MIT's Center for the Exascale Simulation of Material Interfaces (CESMIX)

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

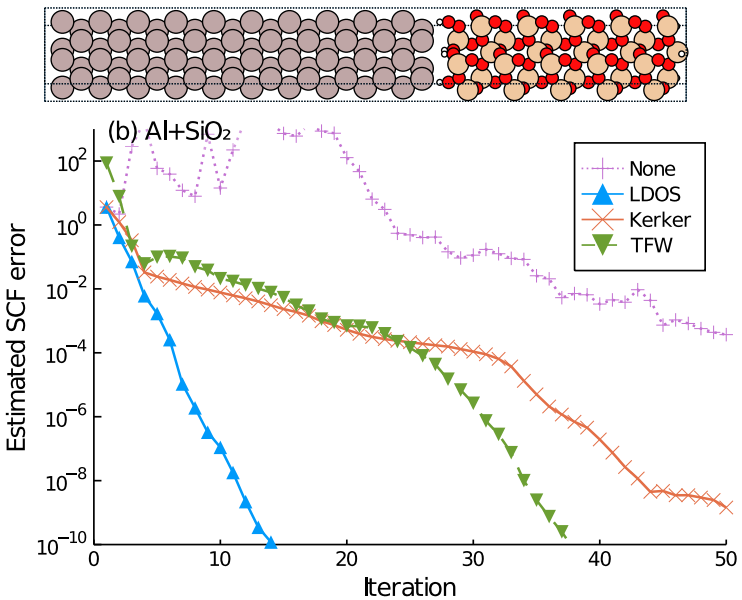


- Reduced model *without SCF*, otherwise representative setting
- Error bars guaranteed: Difference to analytical solution
- Hint what to improve: Tolerance, basis, floating-point type

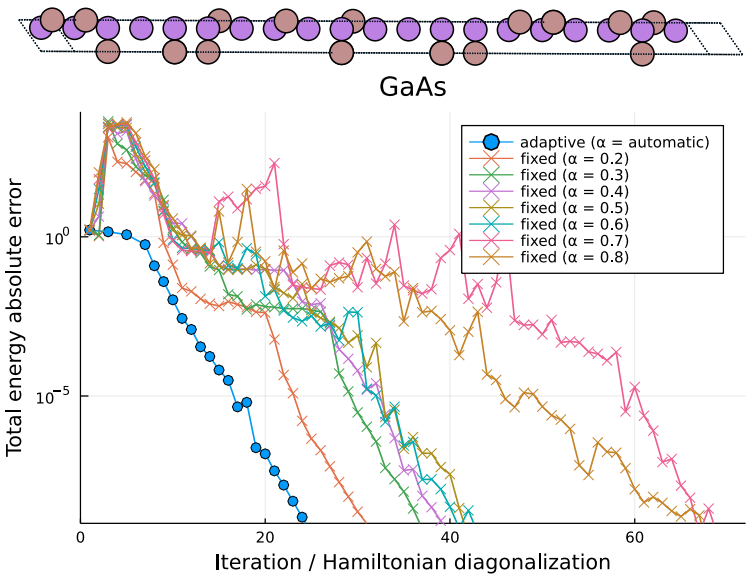
- Just a starting point . . .
- Model error: Combination of analytical and statistical approaches (BEEF)

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

Black-box SCF methods: Our recent ideas



- Parameter-free mixing for inhomogeneous systems[1]



- Adaptive and automatic selection of the damping parameter

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

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

- Self-consistent field equations:

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

$\Rightarrow$ Fixed-point problem $\rho = F(\rho)$

- Use damped update with mixing (preconditioner) P:

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

- Near a fixed-point the error goes as

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

  with $\epsilon^\dagger$ dielectric matrix

- Convergence iff $-1 < \left[1 - \alpha P^{-1} \epsilon^\dagger\right] < 1$
  - $\Rightarrow$ Need $P^{-1} \simeq \left(\epsilon^\dagger\right)^{-1}$ (matching preconditioner) or small α
  - $\Rightarrow$ $\kappa\left(P^{-1}\epsilon^\dagger\right)$ determines convergence rate

Drawback of established approaches

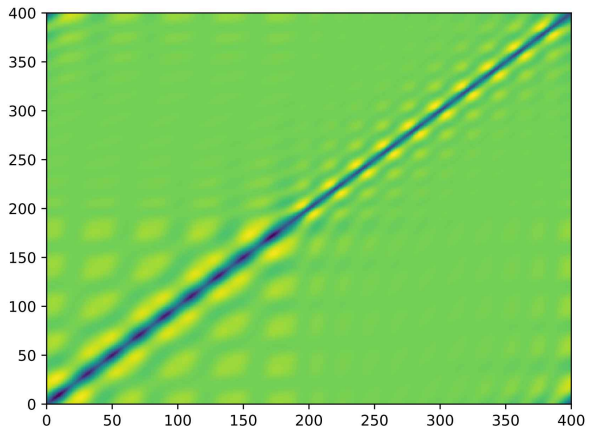
1. Ideal mixing P is system-dependent, but manually chosen
   - Rough idea of dielectric properties needed *a priori*
   - Good preconditioners only known for bulk materials
   - Misses important applications (e.g. inhomogeneous systems)
   - Examples: Metal clusters, passivated surfaces, heterogeneous catalysis, ...

2. No good mixing P known
   - Damping α found by trial and error

- Our motivation: Making these cases more black-box

Solving 1: LDOS mixing[1]

- Bulk preconditioning models tackle directly $P^{-1} \approx \left(\varepsilon^{\dagger}\right)^{-1}$
- But we have $\epsilon^{\dagger} = (1 - \chi_0 K_{\text{Hxc}})$
- Plot of χ_0 (Chain of 10 Sodium atoms and 10 helium atoms):



$\Rightarrow$ Diagonal-dominant, try to approx. $\chi_0(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') \simeq \widetilde{\chi_0}(\underline{\boldsymbol{r}})\delta(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}')$:

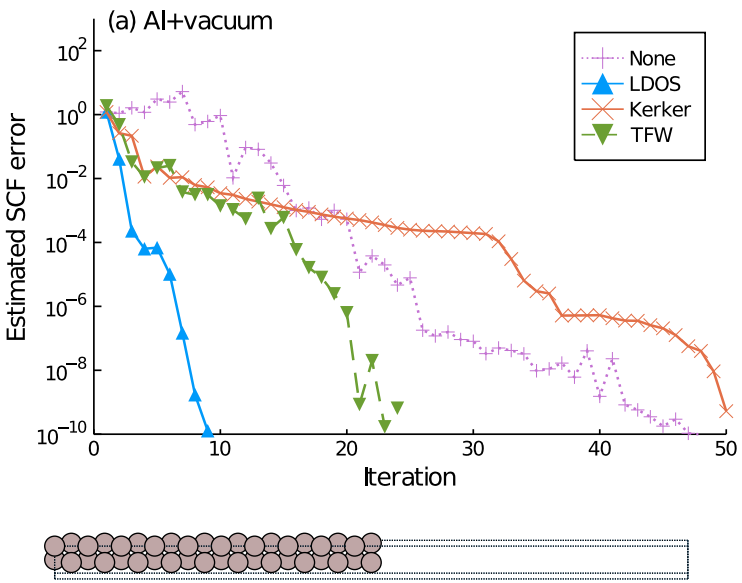
$$P^{-1}\rho_n = \left(1 - \widetilde{\chi_0} K_{\text{Hxc}})\right)^{-1} \rho_n \qquad \text{(iteratively)}$$

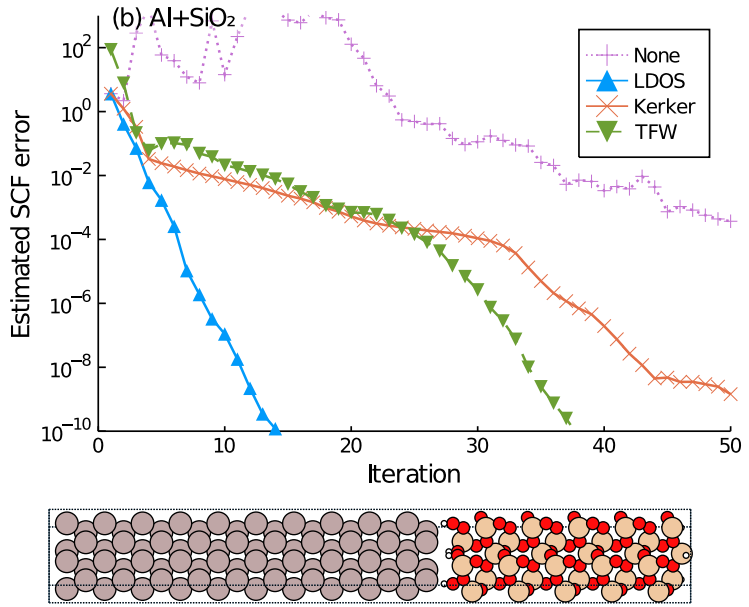
- In the case of LDOS mixing:

$$\chi_0(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}') \simeq -\text{LDOS}(\underline{\boldsymbol{r}})\delta(\underline{\boldsymbol{r}}, \underline{\boldsymbol{r}}')$$

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

LDOS preconditioning (examples)[2]





- 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).

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

Solving 2: Adaptive damping

- Potential mixing:

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

- 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:

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

$\Rightarrow$ Use model to find damping automatically!

Solving 2: Adaptive damping

- Potential mixing:

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

- Quadratic model for DFT energy:

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

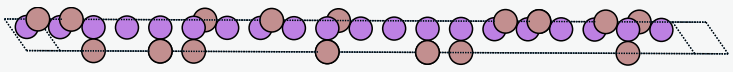
- After some algebra:

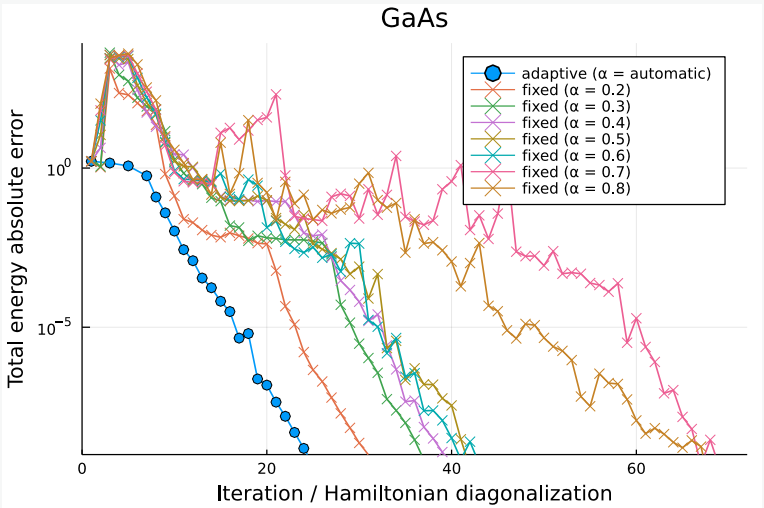
$$\nabla E_{|V=V_n} = -\chi_0 \left[\mathcal{V}(\mathcal{D}(V_n)) - V_n \right]$$
$$\nabla^2 E_{|V=V_n} \simeq -\chi_0 \left[1 - K_{\mathsf{Hxc}} \chi_0 \right]$$

$\Rightarrow$ Use model to find damping automatically!

Adaptive damping

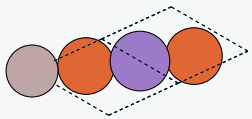
GaAs (with Anderson)

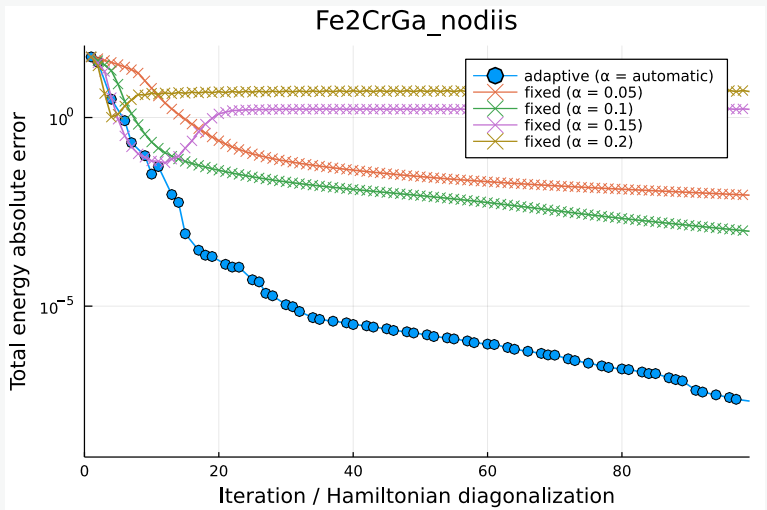




- No preconditioner ($P = I$)
- Non-linear SCF behaviour in initial steps

Fe_2CrGa Heusler (without Anderson)



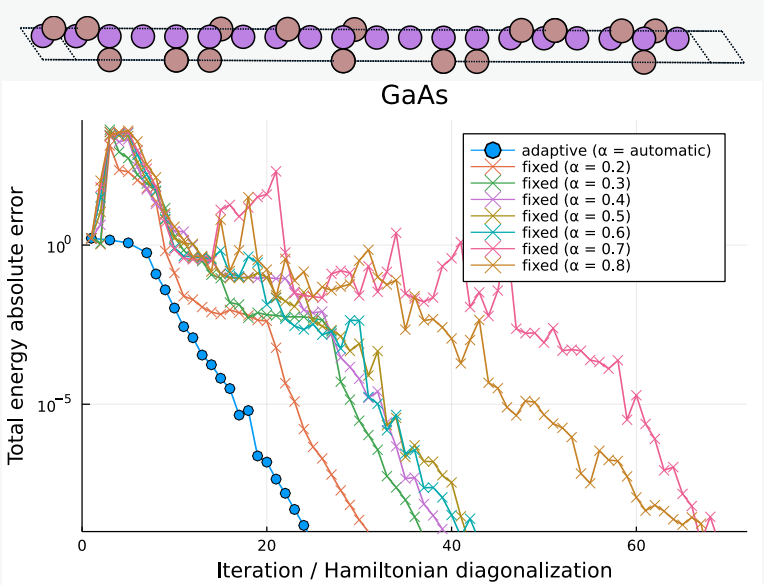


- Unsuitable Kerker preconditioner
- Localised states, spin

- Adaptive damping as black-box safeguard
- Ensures energy / residual decrease
- Interplay with Anderson tricky to interpret

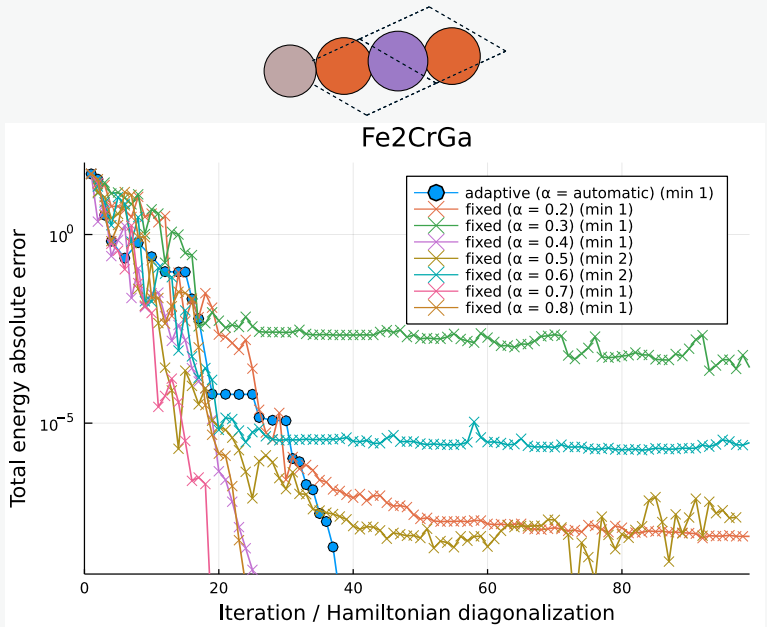
Adaptive damping

GaAs (with Anderson)



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

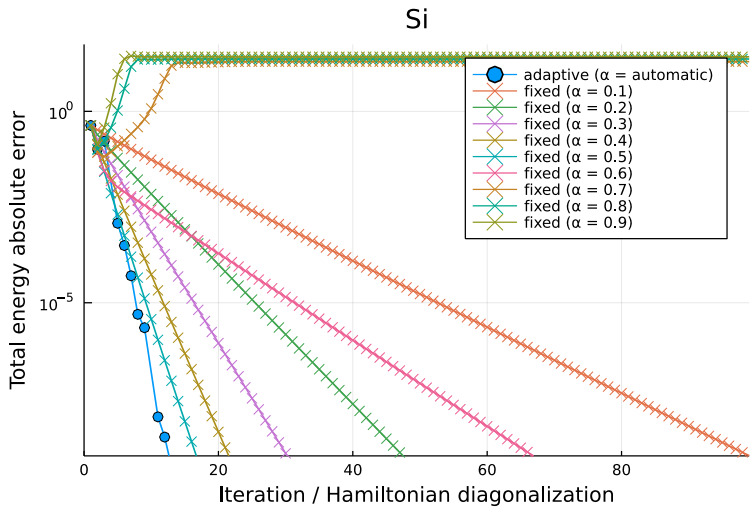
Fe_2CrGa Heusler (with Anderson)

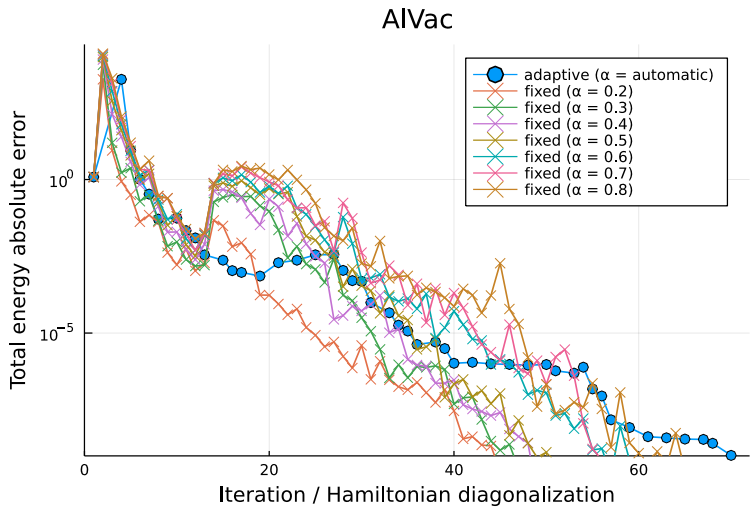


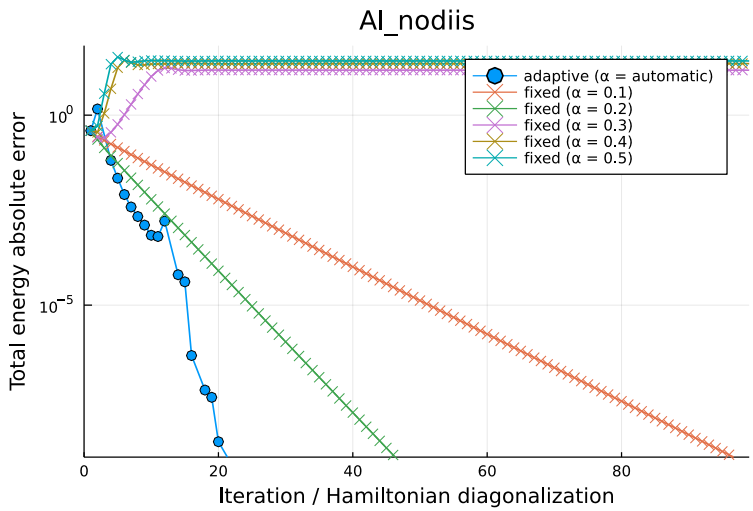
- Unsuitable Kerker preconditioner
- Localised states, spin

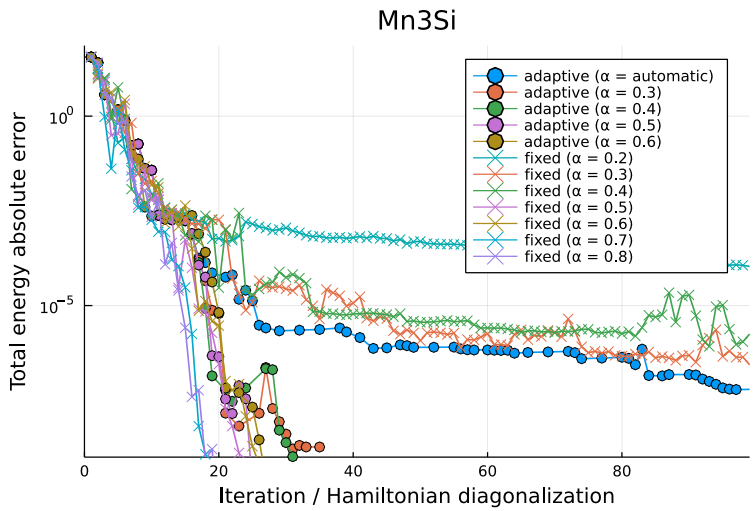
- Adaptive damping as black-box safeguard
- Ensures energy / residual decrease
- Interplay with Anderson tricky to interpret

Adaptive damping







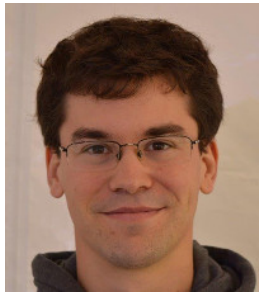


Summary

-  **DFTK** usage:
  - First develop LDOS scheme on test systems (1D, toy problems)
  - Tests on > 800 electrons (in the same code!)
  - Key quantities (χ_0, K_{xc}) fully accessible
- Towards error-controlled methods:
  - Mathematically guided numerics matching the error of models
- 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/slides/vmd2021



Antoine Levitt

Benjamin Stamm

Eric Cancès

all DFTK contributors

Questions?

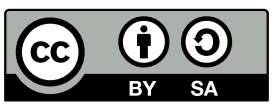
DFTK https://dftk.org

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

mfherbst

https://michael-herbst.com/blog

herbst@acom.rwth-aachen.de



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

- Self-consistent field equations:

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

- Potential-to-density map $\mathcal{D}$

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

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

$\Rightarrow$ Fixed-point problem

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

Analysis of SCF convergence

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

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

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

  - χ_0: Susceptibility (derivative of $\mathcal{D}$), K_{Hxc}: Kernel (deriv. of $\mathcal{V}$)

- Convergence iff $-1 < \left[1 - \alpha P^{-1} \epsilon^\dagger\right] < 1$

  $\Rightarrow$ Need $P^{-1} \simeq \left(\epsilon^\dagger\right)^{-1}$ (matching preconditioner) or small α

  $\Rightarrow$ $\kappa\left(P^{-1}\epsilon^\dagger\right)$ determines convergence rate

SCF instabilities

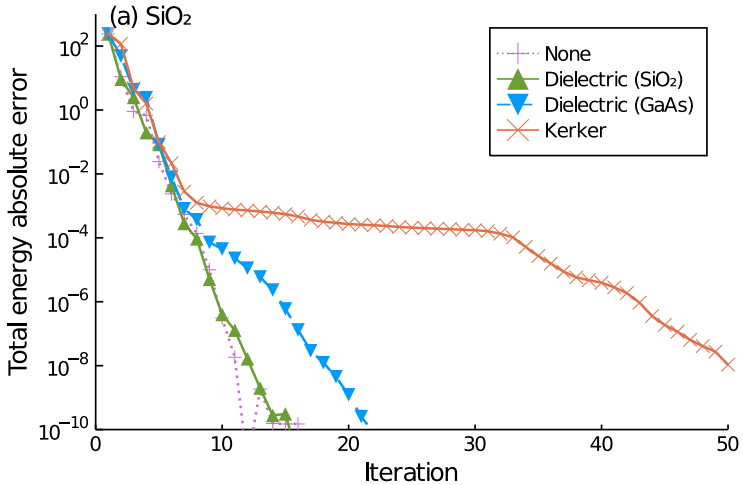
$$e_{n+1} \simeq \left[1 - \alpha P^{-1} \epsilon^\dagger\right] e_n, \qquad \epsilon^\dagger = 1 - \chi_0 (v_C + K_{\text{xc}})$$

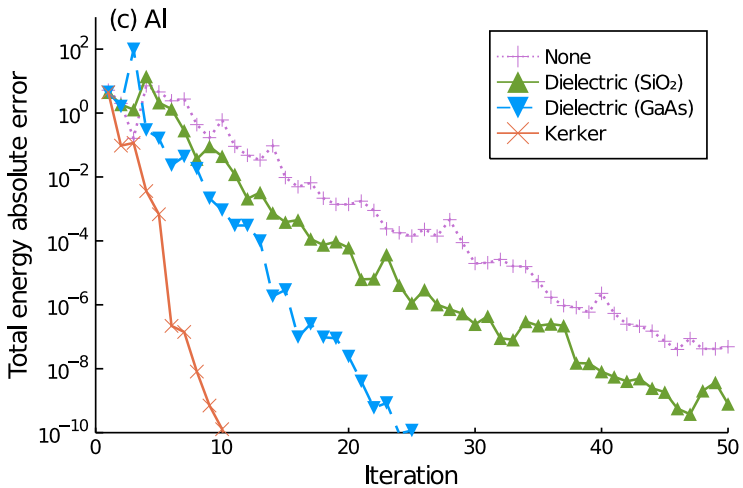
- SCF instabilities increase condition number κ:
  - $\epsilon^\dagger$ has small eigenvalues (e.g. symmetry breaking)
  - χ_0 has large eigenvalues (localised states)
  - Large charge-sloshing modes of v_C are uncompensated by χ_0 (metals)

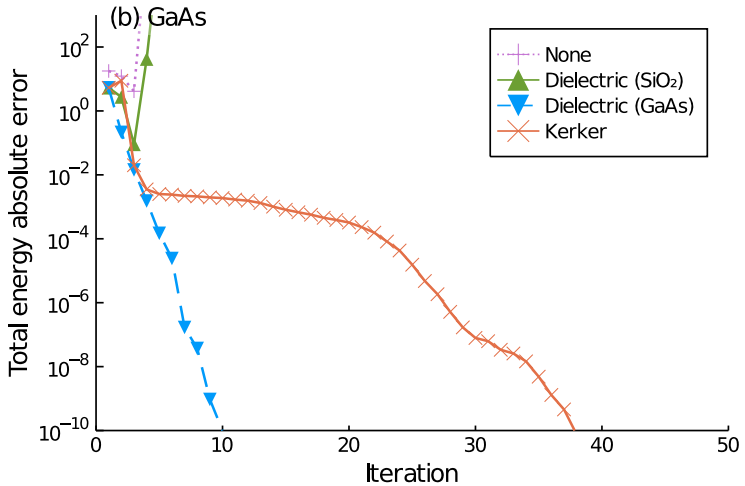
$\Rightarrow$ Need infeasibly small α or good P

- 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).

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 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).