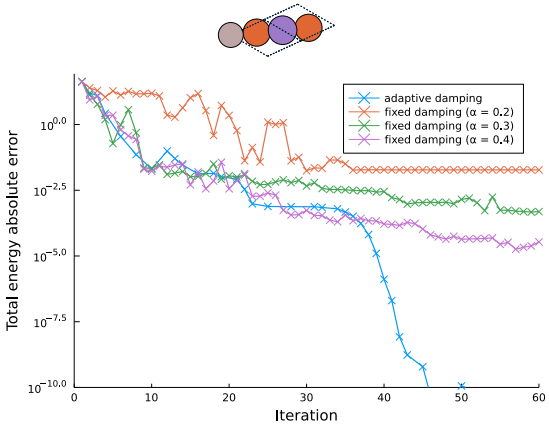
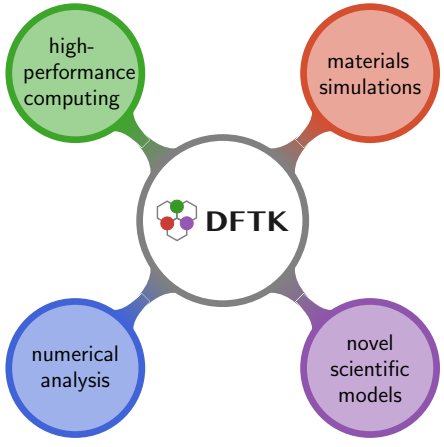
# Robust and efficient accelerated methods for density-functional theory (DFT)

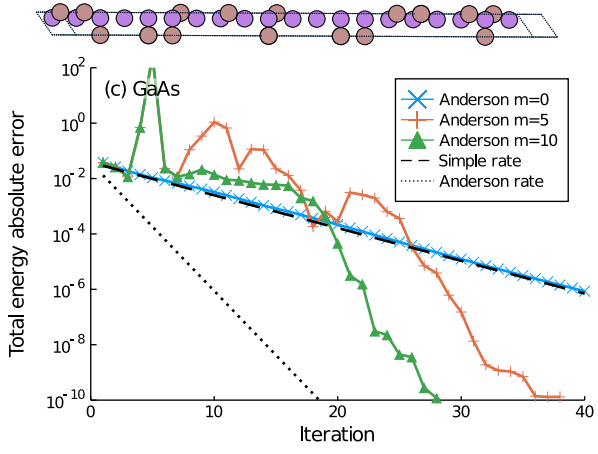
Michael F. Herbst*, Antoine Levitt

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

20th Mai 2021



# 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

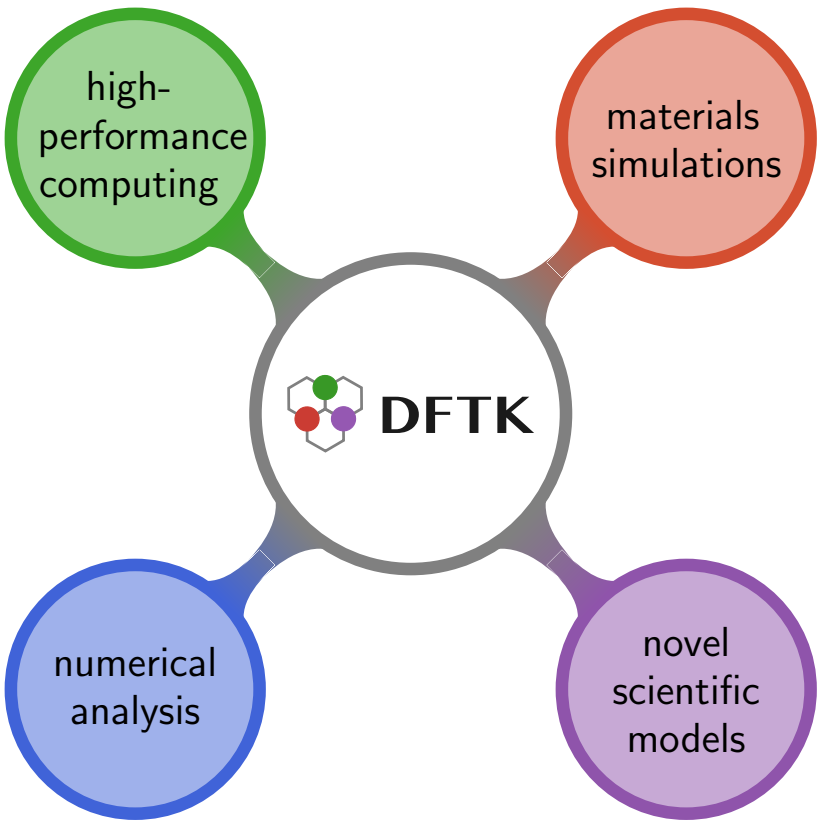
$\Rightarrow$ Reliability needs to be improved!

- 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)[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!

- International and interdisciplinary user base:
  - Analysis, mathematical physics, applications, ...

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

# Kohn-Sham DFT

- Minimisation of DFT energy functional under orthogonality.
- Euler-Lagrange: Coupled set of non-linear elliptic PDEs:

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

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

- 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 $V_{\text{xc}}$, nuclear attraction $V_{\text{Nuc}}$

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

- Density-dependent potential $\mathcal{V}(\rho)$
- Potential-to-density map $F$

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

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

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

- Numerically: Damped fixed-point scheme

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

# SCF convergence

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

- Near a fixed-point the error goes as

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

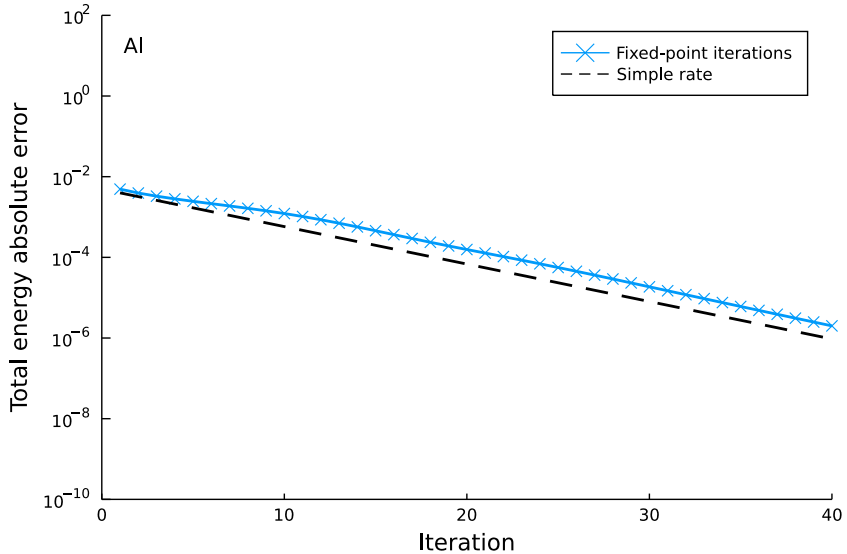
  where $\epsilon^\dagger = 1 - \chi_0 f_{\text{Hxc}}$ (dielectric matrix)

- $\chi_0$: Independent-particle susceptibility (derivative of $F$)
- $f_{\text{Hxc}}$: kernel (derivative of $\mathcal{V}$)

$\Rightarrow$ SCF convergence linked to dielectric properties

- Convergence rate depends on conditioning $\kappa\left(P^{-1}\epsilon^\dagger\right)$
- Preconditioner should capture dielectric properties

# SCF convergence: How bad does it get?




- Aluminium ("simple"), $\kappa(\varepsilon^\dagger) = 18$
- No preconditioning

$\Rightarrow$ Well, but just use a better preconditioner ...

# Choosing the preconditioner: Not always easy

- Standard preconditioners only treat a few simple cases
  - E.g. *bulk* insulators, metals, semiconductors
  - Need to choose *a priori*!
- Preconditioning challenging in important cases:
  - E.g. challenging magnetic alloys
  - Inhomogeneous materials (metal clusters, catalytic surfaces . . . )
  - Partial solution: LDOS Preconditioner[1]

$\Rightarrow$ Additionally need convergence acceleration techniques

$\Rightarrow$ Not always with the expected result (examples follow)

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

# Convergence acceleration (Anderson)

- Convergence acceleration as a black box:

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

- Define $f(\rho) = \alpha P^{-1}\left[F(\mathcal{V}(\rho)) - \rho\right], \quad g(\rho) = \rho + f(\rho)$
- $\mathrm{DIIS}(\,\cdot\,)$ accelerates $g(\rho) = \rho$ and returns

$$g(\rho_n) + \sum_i \beta_i \Big[g(\rho_i) - g(\rho_n)\Big] - \rho_n$$

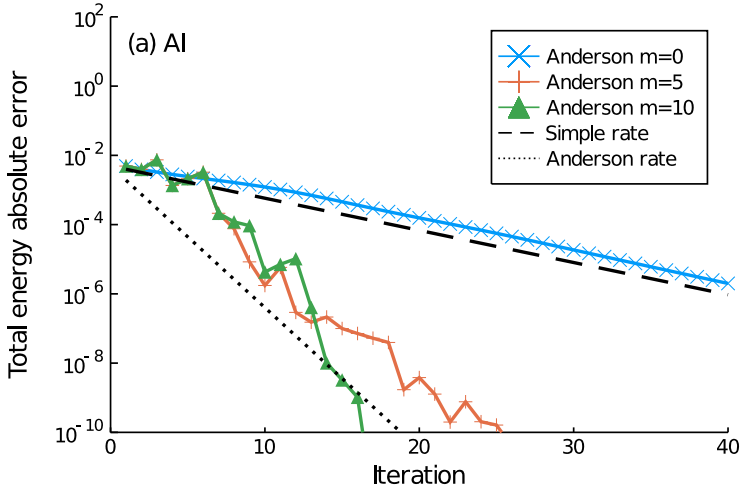
  where $\{\beta_i\}$ are the minimisers of

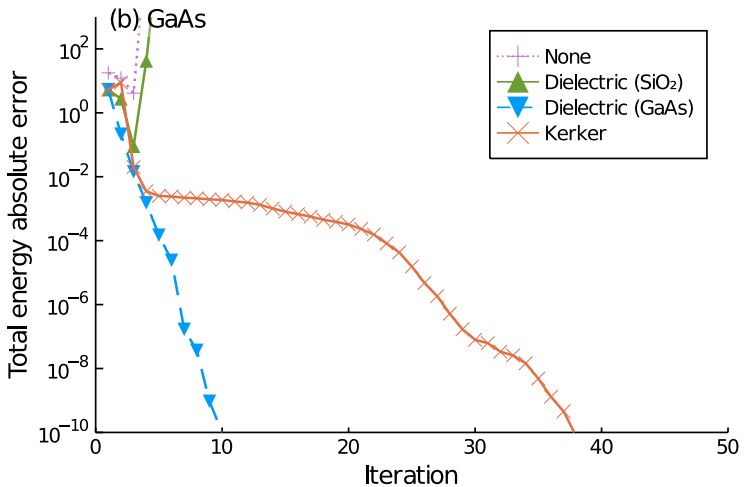
$$\left\| f(\rho_n) + \sum_i \beta_i \Big[f(\rho_i) - f(\rho_n)\Big] \right\|$$

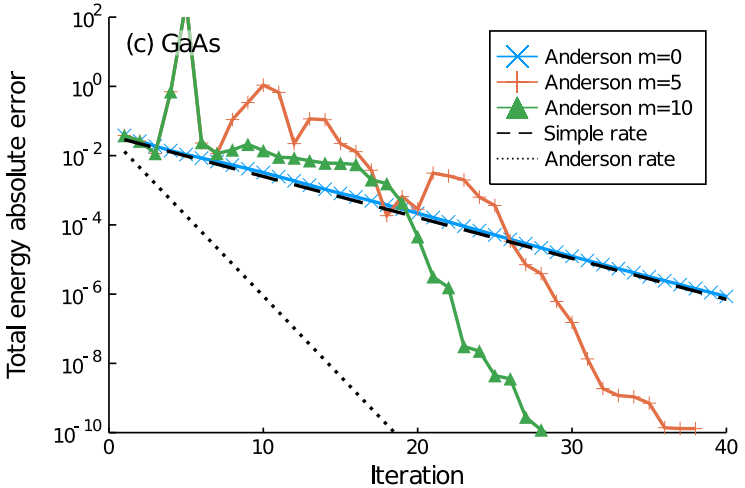
  and $\{\rho_i\}$ is a (truncated) history.

- Does this work? (Remember: $F(\mathcal{V}(\,\cdot\,))$ is not linear)

# Accelerated SCF convergence (Examples)[1]




(a) $\kappa(\epsilon^\dagger) \simeq 18$

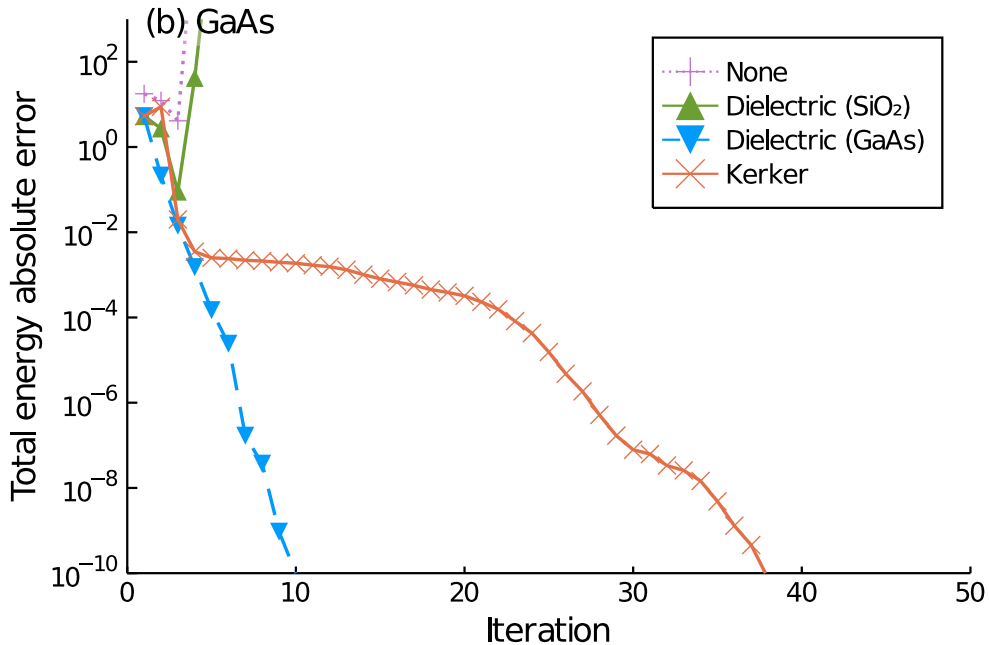
(b) 20 repeats, $\kappa(\epsilon^\dagger) \simeq 18$; $\kappa(P_{\text{Kerker}}^{-1}\epsilon^\dagger) \simeq 100$

(c) 10 repeats, $\kappa(\epsilon^\dagger) \simeq 15$

- Can we do better?

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

# Thoughts from the gallium arsenide case




- In initial phase exhibits strong non-linearity

$\Rightarrow$ Anderson extrapolates very far off

- Practitioners trial and error with damping $\alpha$
- Convergence is guaranteed if damping $\alpha$ 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 $V_n \to F(V_n), V_n^{\mathrm{out}} \to \delta V_n, V_{n+1} \to F(V_{n+1})$ then find optimal damping $\alpha$

- (Simplified) sketch of adaptive damping algorithm:
  - Choose trial $\tilde{\alpha} = \alpha$
  - Accept if energy or residual decreases
  - Else: Find optimal damping $\alpha$, recompute $V_{n+1}$ and $F(V_{n+1})$

# Convergence acceleration with adaptive damping

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

- Adaptive damping $\alpha$ may change between iterations:
  - $f(V) = P^{-1}\left[\mathcal{V}(F(V_n)) - V_n\right]$
  - $g(V) = V + \tilde{\alpha} f(V)$
- DIIS$(\cdot)$ accelerates $g(V) = V$ and returns

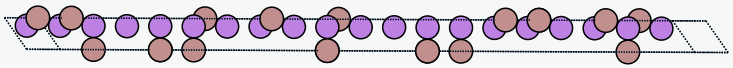
$$\tilde{\alpha}\delta V_n = g(V_n) + \sum_i \beta_i \Big[g(V_i) - g(V_n)\Big] - V_n$$

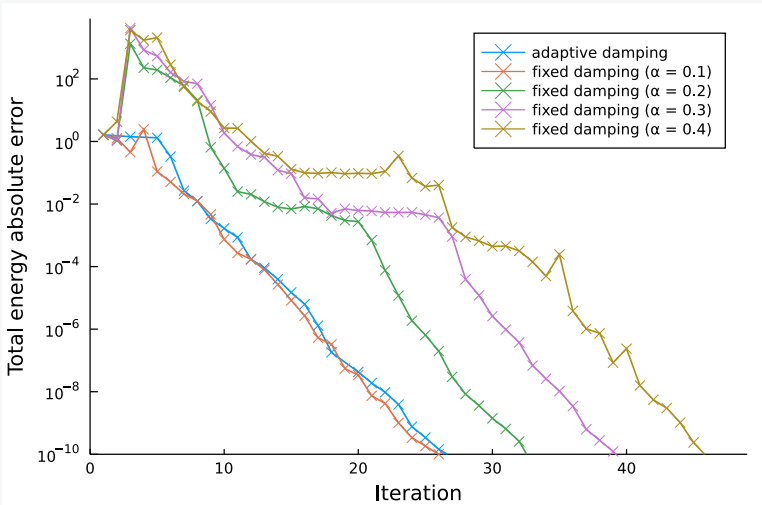
where $\{\beta_i\}$ are the minimisers of

$$\left\| f(\rho_n) + \sum_i \beta_i \Big[f(\rho_i) - f(\rho_n)\Big] \right\|$$

# Adaptive damping (WIP examples)

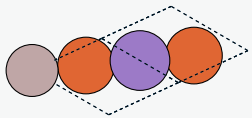
## GaAs (Gallium arsenide)

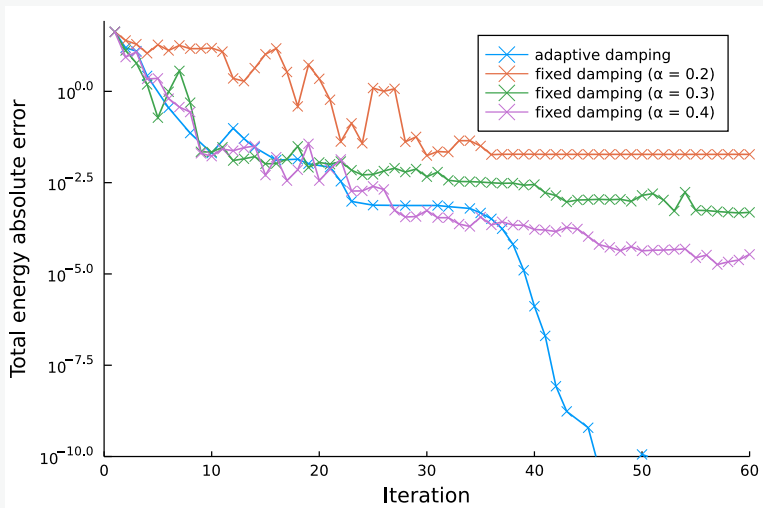





- Non-linear SCF behaviour in initial steps
- Anderson extrapolation fails

## $Fe_2MnAl$ Heusler alloy structure






- Localised states, spin
- No suitable preconditioner $P$

- Adaptive damping as safeguard
- But: Tends to be slightly more expensive for "simple" cases

# Summary

https://michael-herbst.com/slides/siamla21

- DFTK
  - Toolbox for playing with SCF and DFT methods
  - Reduced models and tests on $> 800$ electrons

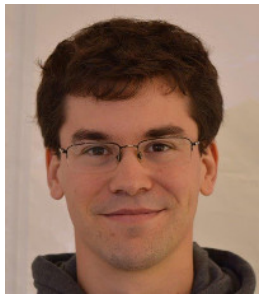
  $\Rightarrow$ One code for mathematical prototyping and applications

- SCF convergence:
  - Interplay between preconditioner, acceleration and damping
  - Good numerical setup needs to look at the physics

- Adaptive damping scheme:
  - Safe guard for strong non-linearities / no preconditioner
  - Reduction of the human factor in parameter selection

# Acknowledgements

https://michael-herbst.com/slides/siamla21



Antoine Levitt

Benjamin Stamm

Eric Cancès

all DFTK contributors

# Questions?

https://michael-herbst.com/slides/siamla21

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

mfherbst

https://michael-herbst.com/blog

herbst@acom.rwth-aachen.de