

# MY regularisation meets DFTK: How to do KS forward and inverse

Michael F. Herbst\*, Vebjørn H. Bakkestuen, Andre Laestadius

\*Mathematics for Materials Modelling ([matmat.org](http://matmat.org)), EPFL

3 December 2024

[https://michael-herbst.com/talks/2024.12.03\\_Oslo\\_KS\\_Inversion.pdf](https://michael-herbst.com/talks/2024.12.03_Oslo_KS_Inversion.pdf)



# Energy consumption of materials discovery

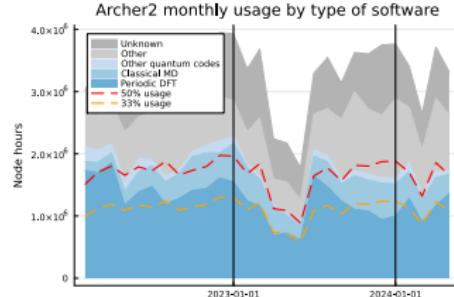
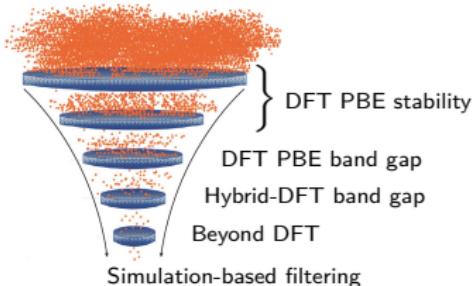


- Current solutions limited by properties of available materials  
    ⇒ Innovation driven by **discovering new materials**
  - **Experimental** research extremely **energy intensive**
    - 1 fume hood  $\simeq$  2-3 average households<sup>1</sup>
- ⇒ Complement experiment by **computational materials discovery**

---

<sup>1</sup>D. Wesolowski *et. al.* Int. J. Sustain. High. Edu. **11**, 217 (2010).

# 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
- Pressing open questions:
  - Robust numerical methods ?
  - Models mathematically justified ?
  - Same accuracy everywhere ?
  - Error estimation and propagation ?

# Density-functional theory (insulators)

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

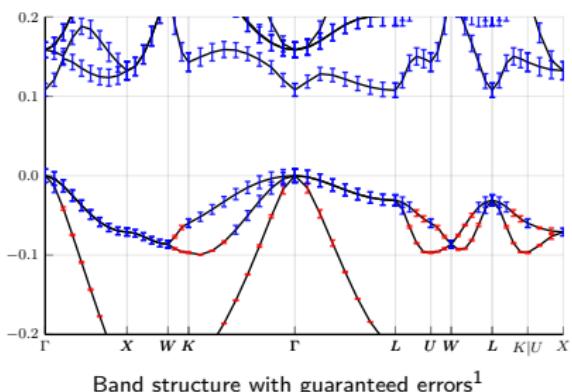
$$\left\{ \begin{array}{l} \forall i \in 1 \dots N : \left( -\frac{1}{2} \Delta + V(\rho_{\Phi}) \right) \psi_i = \varepsilon_i \psi_i, \\ V(\rho) = V_{\text{ext}} + V_{\text{Hxc}}(\rho), \\ \rho_{\Phi} = \sum_{i=1}^N |\psi_i|^2, \end{array} \right.$$

- nuclear attraction  $V_{\text{ext}}$
- Hartree-exchange-correlation  $V_{\text{Hxc}}$
- Plethora of DFT model: Different  $V_{\text{Hxc}}$
- Self-consistent field (SCF) fixed-point problem

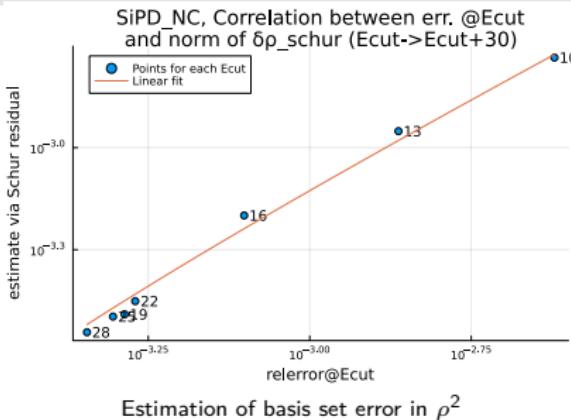
$$D(V(\rho)) = \rho$$

where  $D(V)$  is potential to density map

# Recent mathematical developments: Error estimation



Band structure with guaranteed errors<sup>1</sup>



Estimation of basis set error in  $\rho^2$

- Momentum towards **numerical error estimators** for DFT
  - Remarkable: Problems mathematically not simple !
  - Focus on discretisation basis error (some also floating-point, SCF convergence)
- Results promising, but many challenges & caveats remain
  - Mathematical research based on simplified models !
  - Appropriate software is crucial<sup>1-4</sup>

<sup>1</sup>MFH, A. Levitt, E. Cancès. Faraday Discuss. **223**, 227 (2020).

<sup>2</sup>E. Cancès, G. Dusson, G. Kemlin et. al. SIAM J. Sci. Comp., **44**, B1312 (2022).

<sup>3</sup>E. Cancès, G. Kemlin, A. Levitt. J. Matrix Anal. Appl., **42**, 243 (2021).

<sup>4</sup>E. Cancès, G. Kemlin, A. Levitt. J. Sci. Comput., **98**, 25 (2024)

# 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 ⇒ 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

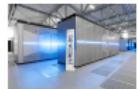
- **Goal:** Modelling physics
- **Scope:** All relevant systems
- **Mix of languages:**  
C, FORTRAN, python, . . .
- **Lifetime:** 100 manyears
- **Size:** 100k – 1M lines
- Obliged to write performant code

- Working with these codes requires different skillsets  
⇒ Orthogonal developer & user communities

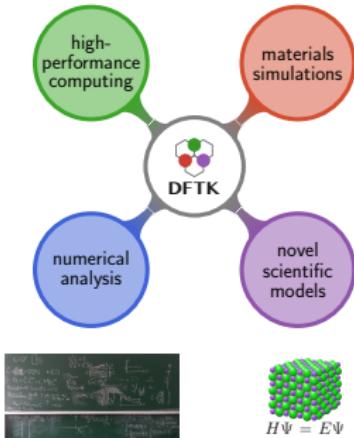
- Obstacle for knowledge transfer:
  - Mathematical methods never tried in practical setting  
(and may well not work well in the real world)
  - Some issues cannot be studied with mathematical codes  
(and mathematicians may never get to know of them)

- **Hypothesis:** People compose if software composes

# Difficulties of interdisciplinary research



- **Hypothesis:** People compose if software composes  
⇒ ~~Mat~~ goal: Software to foster cross-community research



- **DFTK**, the Density-Functional ToolKit
  - Allows restriction to relevant model problems,
  - and scale-up to application regime (1000 electrons)
  - Integrated with high-throughput:  MARVEL  AiiDA
- **Goal:** Demonstrate how **DFTK** can be useful for
  - Developing more robust algorithms for forward KS
  - Bringing error analysis to practitioners
  - Testing theoretical ideas (e.g. MY regularisation)
  - Developing inverse KS methods

# Density-functional theory (insulators)

- DFT approximation: Effective single-particle model

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

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

$$D(V(\rho)) = \rho$$

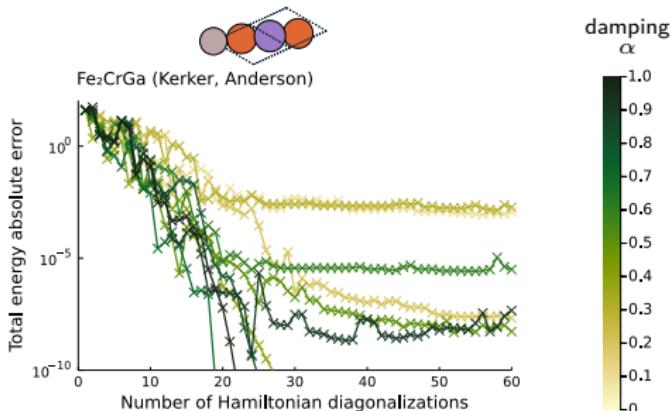
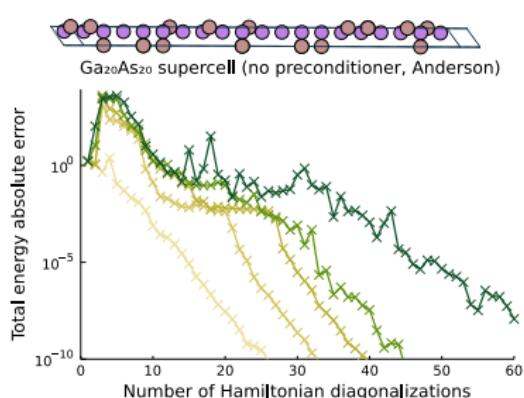
- Density mixing (preconditioner  $P$ , damping  $\alpha$ )

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

- Best  $P$  &  $\alpha$  highly system dependent (metal, insulator, ...)

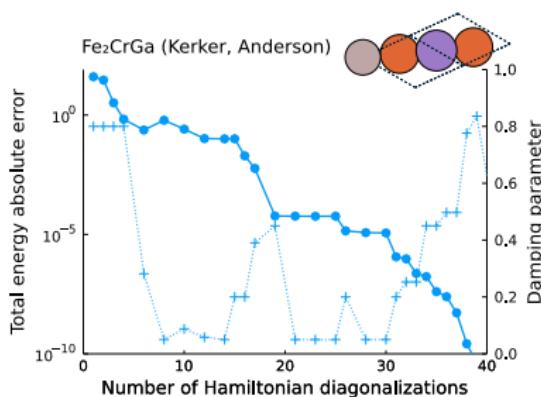
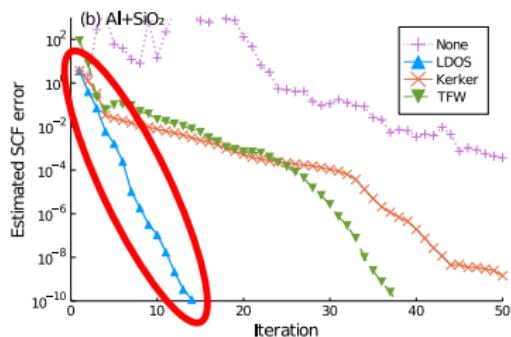
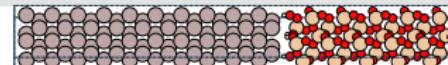
- Usually chosen by trial and error (Impact on energy consumption ...)

# Illustration: Guessing a suitable damping $\alpha$ can be hard



- Inefficient standard damping ( $0.6 - 0.8$ )
- Surprisingly small damping for smooth convergence
- Heusler alloy: Materials class with unusual magnetic properties  
⇒ Numerically challenging behaviour
- SCF irregular:  $\alpha$  versus convergence
- Usual heuristics breaks:  
Larger damping is better

# Self-adapting SCF algorithms



- Preconditioning inhomogeneous systems (surfaces, clusters, ...)
  - LDOS preconditioner<sup>1</sup>: Parameter-free and self-adapting
  - ca. 50% less iterations
- Damping  $\alpha$  adapted in each step (using tailored quadratic model)
  - Avoids trial and error (but may have a small overhead)
  - Safeguard with theoretical guarantees<sup>2</sup>

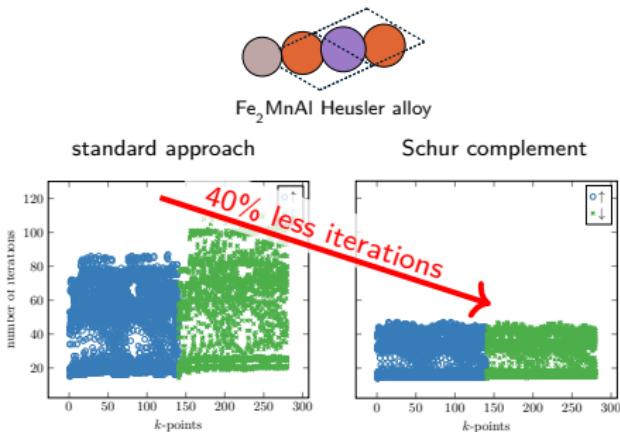
⇒ Maths / physics collaboration:

Exchange of ideas between simplified & practical settings crucial

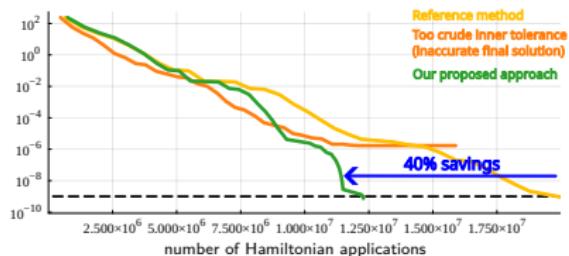
<sup>1</sup>MFH, A. Levitt. J. Phys. Condens. Matter **33**, 085503 (2021).

<sup>2</sup>MFH, A. Levitt. J. Comput. Phys. **459**, 111127 (2022).

# Aside: Our improvements for response & DFT properties<sup>1,2</sup>



- Schur-based approach tames CG
- ca. 40% less iterations
- Improvement comes for free
- Development guided using a “real material”



- Response: Nested linear problems
  - Tolerance of inner problem ?
  - Here<sup>2</sup>: Adaptive, provably correct strategy
  - Based on inexact Krylov methods
- ⇒ Superlinear convergence

<sup>1</sup>E. Cancès, MFH, G. Kemlin, et. al. Lett. Math. Phys. **113**, 21 (2023).

<sup>2</sup>B. Sun, MFH, *in preparation*.

# DEMO

# DEMO

Custom SCF algorithms in  DFTK



→ [https://michael-herbst.com/talks/2024.12.03\\_Oslo\\_KS\\_Inversion\\_1\\_scf.html](https://michael-herbst.com/talks/2024.12.03_Oslo_KS_Inversion_1_scf.html)

# Exact DFT for periodic problems

- Regular lattice  $\mathbb{L} = \mathbb{Z}a_1 + \mathbb{Z}a_2 + \mathbb{Z}a_3$
  - Unit cell  $\Omega = \{x \in \mathbb{R}^3 \mid |x - R| > |x| \quad \forall R \in \mathbb{L}\}$
  - Hilbert space  $L_{\text{per}}^2(\mathbb{R}^3) = \{f \in L_{\text{loc}}^2(\mathbb{R}^3) \mid f \text{ is } \mathbb{L}\text{-periodic}\}$
  - Derived Sobolev spaces  $H_{\text{per}}^s$  accordingly
- We take

- Density space  $\rho \in \mathcal{D} = H_{\text{per}}^{-1}(\mathbb{R}^3)$
- Potential space  $\rho \in \mathcal{V} = \mathcal{D}^* = H_{\text{per}}^1(\mathbb{R}^3)$
- Duality mapping  $J : \mathcal{D} \rightarrow \mathcal{V}$  (analytic; see later)

- Exact DFT problem:

$$E(v) = \inf_{\rho \in \mathcal{D}} (F(\rho) + \langle v | \rho \rangle)$$

(with  $v$  external potential,  $F$  universal functional)

- Universal functional  $F$  is not differentiable, but convex & lower semi-continuous

## Exact DFT for periodic problems

- Regular lattice  $\mathbb{L} = \mathbb{Z}a_1 + \mathbb{Z}a_2 + \mathbb{Z}a_3$
- Unit cell  $\Omega = \{x \in \mathbb{R}^3 \mid |x - R| > |x| \quad \forall R \in \mathbb{L}\}$
- Hilbert space  $L_{\text{per}}^2(\mathbb{R}^3) = \{f \in L_{\text{loc}}^2(\mathbb{R}^3) \mid f \text{ is } \mathbb{L}\text{-periodic}\}$
- Derived Sobolev spaces  $H_{\text{per}}^s$  accordingly
- We take
  - Density space  $\rho \in \mathcal{D} = H_{\text{per}}^{-1}(\mathbb{R}^3)$
  - Potential space  $\rho \in \mathcal{V} = \mathcal{D}^* = H_{\text{per}}^1(\mathbb{R}^3)$
  - Duality mapping  $J : \mathcal{D} \rightarrow \mathcal{V}$  (analytic; see later)
- Exact DFT problem:
$$E(v) = \inf_{\rho \in \mathcal{D}} (F(\rho) + \langle v | \rho \rangle)$$
(with  $v$  external potential,  $F$  universal functional)
- Universal functional  $F$  is not differentiable, but convex & lower semi-continuous

# Moreau-Yosida regularisation

## Definition

$\mathcal{D}$  uniformly convex &  $\mathcal{F} : \mathcal{D} \rightarrow \mathbb{R}$  convex lsc. For  $\varepsilon > 0$  the MY regularisation is

$$\mathcal{F}^\varepsilon(\rho) = \inf_{\mu \in \mathcal{D}} \left( \mathcal{F}(\mu) + \frac{1}{2\varepsilon} \|\mu - \sigma\|_{\mathcal{D}}^2 \right)$$

- Define  $E^\varepsilon(\rho) = \inf_{\rho \in \mathcal{D}} (\mathcal{F}^\varepsilon(\rho) + \langle v | \rho \rangle)$  to note

$$E(v) = E^\varepsilon(v) + \frac{\varepsilon}{2} \|v\|_{\mathcal{V}}^2$$

$\Rightarrow$  Lossless regularisation (original minimum can be recovered)

- $\mathcal{F}^\varepsilon$  is convex lsc. and has unique minimiser at proximal point

$$\rho^\varepsilon = \arg \min_{\rho \in \mathcal{D}} \mathcal{F}^\varepsilon(\rho)$$

# Kohn-Sham inversion

- Given ground-state density  $\rho_{GS}$  (assumed  $v$ -representable)
- What is the corresponding potential  $v$  ?
- Guiding functional (Misses  $E_{xc}$  compared to  $F$ , but has potential)  
$$\mathcal{F}(\rho) = T(\rho) + E_H(\rho) + \langle v | \rho \rangle$$

- Consider computing proximal point of  $\rho_{GS}$ :

$$\begin{aligned}\rho_{GS}^\varepsilon &= \arg \min_{\rho \in \mathcal{D}} \mathcal{E}(\rho; \rho_{GS}) \\ \text{with } \mathcal{E}(\rho; \rho_{GS}) &= \mathcal{F}(\rho) + \frac{1}{2\varepsilon} \|\rho - \rho_{GS}\|_{\mathcal{D}}^2\end{aligned}$$

- Notably  $\rho_{GS} = \lim_{\varepsilon \rightarrow 0} \rho_{GS}^\varepsilon$  and (since  $\underline{\partial} \frac{1}{2} \|\rho\|_{\mathcal{D}}^2 = J(\rho)$ ) the stationarity conditions imply

$$0 \in \underline{\partial} \mathcal{F}(\rho_{GS}^\varepsilon) + \frac{1}{\varepsilon} J(\rho_{GS}^\varepsilon - \rho_{GS})$$

$$\Rightarrow v_{xc} = \lim_{\varepsilon \rightarrow 0} v_{xc}^\varepsilon \quad \text{with } v_{xc}^\varepsilon = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} J(\rho_{GS}^\varepsilon - \rho_{GS})$$

(using  $\rho_{xc}$  is  $v$ -representable & comparison of  $F^\varepsilon$  with  $\mathcal{F}^\varepsilon$ )

# Kohn-Sham inversion

- Given ground-state density  $\rho_{GS}$  (assumed  $v$ -representable)
- What is the corresponding potential  $v$  ?
- Guiding functional** (Misses  $E_{xc}$  compared to  $F$ , but has potential)  
$$\mathcal{F}(\rho) = T(\rho) + E_H(\rho) + \langle v | \rho \rangle$$

- Consider computing proximal point of  $\rho_{GS}$ :

$$\begin{aligned}\rho_{GS}^\varepsilon &= \arg \min_{\rho \in \mathcal{D}} \mathcal{E}(\rho; \rho_{GS}) \\ \text{with } \mathcal{E}(\rho; \rho_{GS}) &= \mathcal{F}(\rho) + \frac{1}{2\varepsilon} \|\rho - \rho_{GS}\|_{\mathcal{D}}^2\end{aligned}$$

- Notably  $\rho_{GS} = \lim_{\varepsilon \rightarrow 0} \rho_{GS}^\varepsilon$  and (since  $\underline{\partial} \frac{1}{2} \|\rho\|_{\mathcal{D}}^2 = J(\rho)$ ) the stationarity conditions imply

$$0 \in \underline{\partial} \mathcal{F}(\rho_{GS}^\varepsilon) + \frac{1}{\varepsilon} J(\rho_{GS}^\varepsilon - \rho_{GS})$$

$$\Rightarrow v_{xc} = \lim_{\varepsilon \rightarrow 0} v_{xc}^\varepsilon \quad \text{with } v_{xc}^\varepsilon = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} J(\rho_{GS}^\varepsilon - \rho_{GS})$$

(using  $\rho_{xc}$  is  $v$ -representable & comparison of  $F^\varepsilon$  with  $\mathcal{F}^\varepsilon$ )

# Kohn-Sham inversion

- **Guiding functional** (Misses  $E_{xc}$  compared to  $F$ , but has potential)

$$\mathcal{F}(\rho) = T(\rho) + E_H(\rho) + \langle v | \rho \rangle$$

- Consider computing proximal point of  $\rho_{GS}$ :

$$\begin{aligned}\rho_{GS}^\varepsilon &= \arg \min_{\rho \in \mathcal{D}} \mathcal{E}(\rho; \rho_{GS}) \\ \text{with } \mathcal{E}(\rho; \rho_{GS}) &= \mathcal{F}(\rho) + \frac{1}{2\varepsilon} \|\rho - \rho_{GS}\|_{\mathcal{D}}^2\end{aligned}$$

- Notably  $\rho_{GS} = \lim_{\varepsilon \rightarrow 0} \rho_{GS}^\varepsilon$  and (since  $\underline{\partial} \frac{1}{2} \|\rho\|_{\mathcal{D}}^2 = J(\rho)$ ) the stationarity conditions imply

$$\begin{aligned}0 &\in \underline{\partial} \mathcal{F}(\rho_{GS}^\varepsilon) + \frac{1}{\varepsilon} J(\rho_{GS}^\varepsilon - \rho_{GS}) \\ \Rightarrow v_{xc} &= \lim_{\varepsilon \rightarrow 0} v_{xc}^\varepsilon \quad \text{with } v_{xc}^\varepsilon = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} J(\rho_{GS}^\varepsilon - \rho_{GS})\end{aligned}$$

(using  $\rho_{GS}$  is  $v$ -representable & comparison of  $F^\varepsilon$  with  $\mathcal{F}^\varepsilon$ )

# KS inversion in our setting<sup>1</sup>

- KS inversion:

$$\rho_{\text{GS}}^{\varepsilon} = \arg \min_{\rho} T(\rho) + E_H(\rho) + \langle v | \rho \rangle + \frac{1}{2\varepsilon} \|\rho - \rho_{\text{GS}}\|_{\mathcal{D}}^2$$

- Forward Kohn-Sham DFT:

$$\rho_* = \arg \min_{\rho} T(\rho) + E_H(\rho) + \langle v | \rho \rangle + \tilde{E}_{\text{xc}}(\rho)$$

- ⇒ Use forward-like implementation to perform inverse KS
- ⇒ Parametrisation of  $\rho$  in orthonormal orbitals

$$\rho(r) = \sum_{i=1}^N |\psi_i(r)|^2 \quad \text{with} \quad \langle \psi_i | \psi_j \rangle = \delta_{ij}$$

---

<sup>1</sup>MFH, V. Bakkestuen, A. Laestadius. *Kohn-Sham Inversion with Mathematical Guarantees* arXiv:2409.04372

## Why on earth should this work ?

- In the periodic setting & using plane-wave discretisations all key quantities are easily computable
- E.g. duality map  $J$  is a convolution using a Yukawa kernel:

$$J(\rho) = \int_{\mathbb{R}^3} \frac{\rho(x)}{|r-x|} e^{-|r-x|} dx = \frac{1}{4\pi} \sum_{G \in \mathbb{L}} \frac{\hat{\rho}(G)}{(1 + |G|)^2}$$

- Similarly for e.g. Sobolev norms:

$$\|u\|_{H_{\text{per}}^s}^2 = \sum_{G \in \mathbb{L}} (1 + |G|^2)^s |\hat{u}_G|^2$$

# DEMO

# DEMO

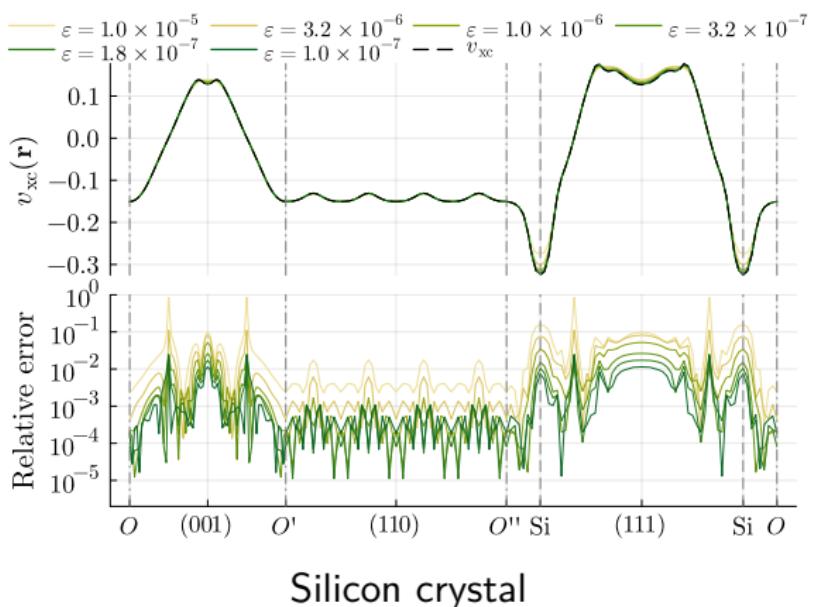
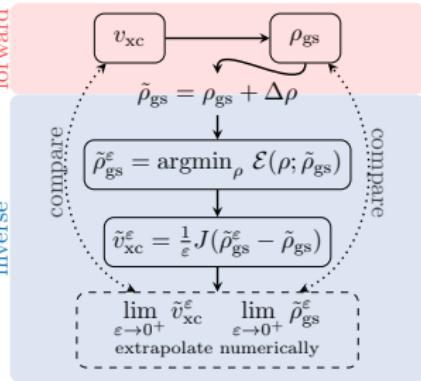
Implementing inverse KS in  DFTK



→ [https://michael-herbst.com/talks/2024.12.03\\_Oslo\\_KS\\_Inversion\\_2\\_inversion.html](https://michael-herbst.com/talks/2024.12.03_Oslo_KS_Inversion_2_inversion.html)

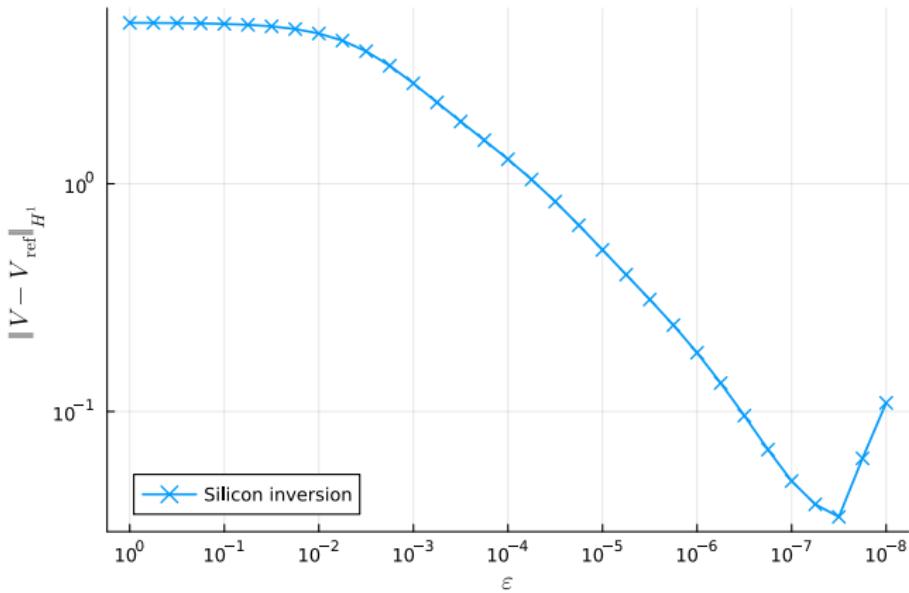
# Inversion algorithm flowchart and performance

forward  
inverse



Silicon crystal

# Convergence of silicon



## Error bounds

- Within this framework we can say something about errors
- Consider inexact references

$$\tilde{\rho}_{\text{gs}} = \rho_{\text{gs}} + \Delta\rho$$

- Triangle-inequality decomposition of the total error

$$\|v_{\text{xc}} - \tilde{v}_{\text{xc}}^{\varepsilon}\|_{\mathcal{V}} \leq \underbrace{\|v_{\text{xc}} - v_{\text{xc}}^{\varepsilon}\|_{\mathcal{V}}}_{\text{terminating at finite } \varepsilon} + \underbrace{\|v_{\text{xc}}^{\varepsilon} - \tilde{v}_{\text{xc}}^{\varepsilon}\|_{\mathcal{V}}}_{\text{Use of inexact } \rho_{\text{gs}}}$$

## Error bounds: Inexact densities<sup>1</sup>

- Our analysis uses non-expensiveness of  $\rho \mapsto \rho^\varepsilon$ :

$$\|\rho_{\text{gs}}^\varepsilon - \tilde{\rho}_{\text{gs}}^\varepsilon\|_{\mathcal{D}} \leq \|\rho_{\text{gs}} - \tilde{\rho}_{\text{gs}}\|_{\mathcal{D}} = \|\Delta\rho\|_{\mathcal{D}}$$

- Therefore  $0 \leq Q_\varepsilon(\Delta\rho) \leq 1$  with

$$Q_\varepsilon(\Delta\rho) \frac{\|\rho_{\text{gs}}^\varepsilon - \tilde{\rho}_{\text{gs}}^\varepsilon\|_{\mathcal{D}}}{\|\Delta\rho\|_{\mathcal{D}}}$$

and using the linearity of  $J$  we get:

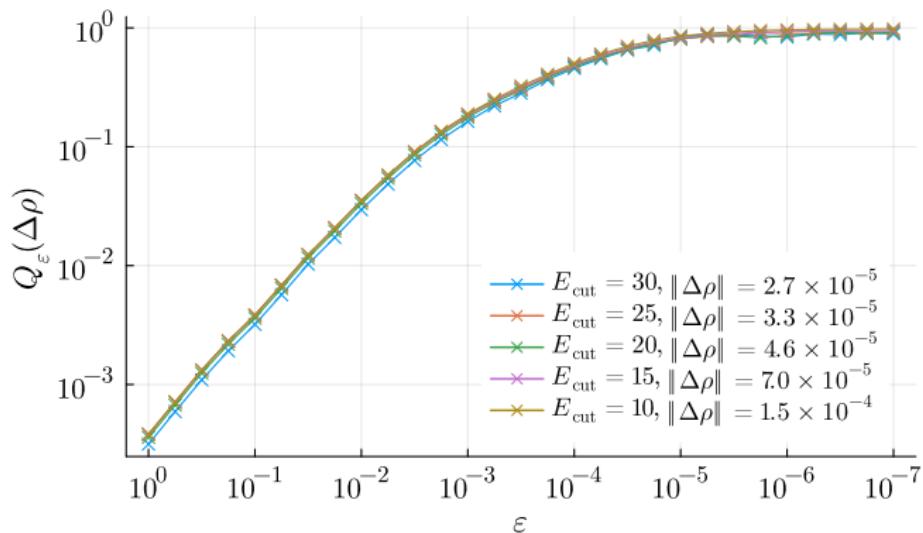
$$\|\tilde{v}_{\text{xc}}^\varepsilon - v_{\text{xc}}^\varepsilon\|_{\mathcal{V}} \leq \frac{1 + Q_\varepsilon(\Delta\rho)}{\varepsilon} \|\Delta\rho\|_{\mathcal{D}}$$

---

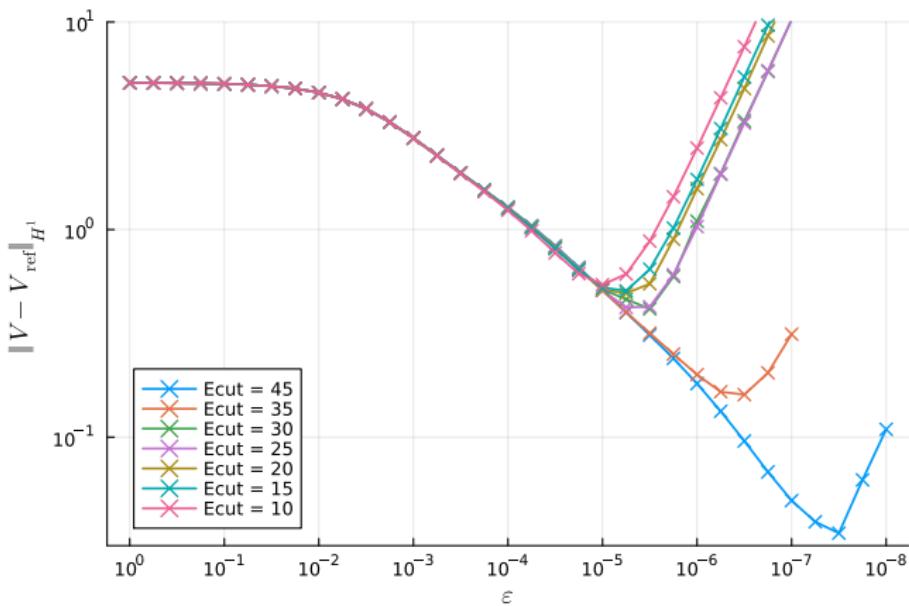
<sup>1</sup>MFH, V. Bakkestuen, A. Laestadius. *Kohn-Sham Inversion with Mathematical Guarantees* arXiv:2409.04372

## Non-expansiveness verified in practice

$$Q_\varepsilon(\Delta\rho) \frac{\|\rho_{\text{gs}}^\varepsilon - \tilde{\rho}_{\text{gs}}^\varepsilon\|_{\mathcal{D}}}{\|\Delta\rho\|_{\mathcal{D}}}$$



## Convergence behaviour with errors





- High-throughput materials discovery
  - Need to understand errors & mathematical properties of models
-  **DFTK**: A DFT software for cross-disciplinary research
  - Reduced settings (error analysis) *and* high-throughput testing
  - Stimulated research of robust algorithms for forward DFT
  - Numerical experiments to accompany mathematical research
  - ⇒ Overcome barriers: **People compose if software composes**
- Moreau-Yosida formulation of DFT
  - Route to Kohn-Sham inversion with mathematical guarantees
  -  **DFTK** provides first practical implementation

# Acknowledgements

MY-based KS inversion

- Vebjørn H. Bakkestuen (OsloMet)
- Andre Laestadius (OlsoMet)

MatMat group

- Bruno Ploumhans (~~Mat Mat~~)
- Anna Paulish (~~Mat Mat~~)
- Ivan Maliyov (~~Mat Mat~~)
- Niklas Schmitz (~~Mat Mat~~)
- Cédric Travelletti (~~Mat Mat~~)



MARVEL



# Questions?

~~M~~~~X~~t Mat <https://matmat.org>

 mfherbst

 michael.herbst@epfl.ch



[https://michael-herbst.com/talks/2024.12.03\\_Oslo\\_KS\\_Inversion.pdf](https://michael-herbst.com/talks/2024.12.03_Oslo_KS_Inversion.pdf)

 DFTK <https://dftk.org>

 MFH, V. Bakkestuen, A. Laestadius. *Kohn-Sham Inversion with Mathematical Guarantees* arXiv:2409.04372

I also have stickers ...



# Contents

② SCF convergence and LDOS

③ Schur and inexact Krylov

# SCF convergence analysis

- Density mixing (preconditioner  $P$ , damping  $\alpha$ )

$$\rho_{n+1} = F(\rho_n) \quad \text{with} \quad F(\rho_n) = \rho_n + \alpha P^{-1} [D(V(\rho_n)) - \rho_n]$$

- Convergent iff  $1 < F'(\rho_*) < 1$

$$\begin{aligned} F'(\rho_*) &= 1 + \alpha P^{-1} \left[ \underbrace{D'(V(\rho_*))}_{=\chi_0} \underbrace{V'(\rho_*) - 1}_{=K} \right] \\ &= 1 - \alpha P^{-1} [1 - \chi_0 K] \\ &= 1 - \alpha P^{-1} \varepsilon^\dagger \end{aligned}$$

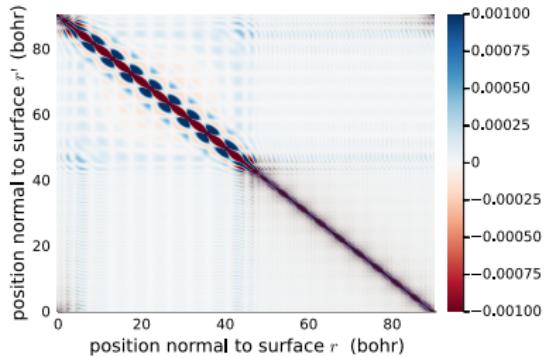
- Dielectric matrix  $\varepsilon^\dagger$ : **Depends on physics** (conduction, screening)
- By second-order conditions:  $\varepsilon^\dagger \geq 0$  (near fixed point)

⇒ Ideal preconditioner has  $P^{-1}\varepsilon^\dagger \approx I$

- Note:  $P$  needs to **adapt to physics** of unknown system!
- No such  $P$  available: Choose  $\alpha$  appropriately (**Trial and error**)

# Black-box $P$ : Local density of states (LDOS) mixing<sup>1</sup>

- Bulk preconditioner (e.g. Kerker) neglect local structure of  $\varepsilon^\dagger$
- We propose to employ  $\varepsilon^\dagger = (1 - \chi_0 K)$
- $\chi_0(r, r')$  unit-cell internal fluctuations, diagonal dominant:



- Tackle **charge sloshing**: Consider large-scale variations of  $\chi_0$ :  
$$\chi_0(r, r') \simeq -\text{LDOS}(r)\delta(r, r') \quad (\text{homogenisation } \text{LDOS}(r) \approx \int \chi_0(r, r') dr')$$
- Apply preconditioner **iteratively**:  
$$P^{-1}\rho_n = [1 - \widetilde{\chi}_0 K]^{-1} \rho_n, \quad \widetilde{\chi}_0(r, r') = -\text{LDOS}(r)\delta(r, r')$$

<sup>1</sup>MFH, A. Levitt. J. Phys. Condens. Matter **33**, 085503 (2021).

# Local density of states (LDOS) approximation for $\chi_0$ <sup>1</sup>

- Adler-Wiser formula (with occupation function  $f$ ):

$$\chi_0(\underline{\mathbf{r}}, \underline{\mathbf{r}'}) = \sum_{n,m} \frac{f(\varepsilon_n) - f(\varepsilon_m)}{\varepsilon_n - \varepsilon_m} \psi_n(\underline{\mathbf{r}}) \psi_m^*(\underline{\mathbf{r}}) \psi_m(\underline{\mathbf{r}'}) \psi_n^*(\underline{\mathbf{r}'})$$

- Large-scale variations from  $V_n$  to  $V_{n+1}$ :

⇒ Assume  $\underline{\mathbf{r}} \mapsto \chi_0(\underline{\mathbf{r}}, \underline{\mathbf{r}'})$  more localised around  $\underline{\mathbf{r}'}$  than  $V_i(\underline{\mathbf{r}'})$ .

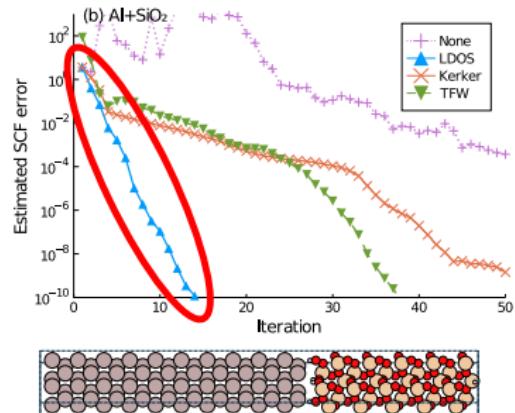
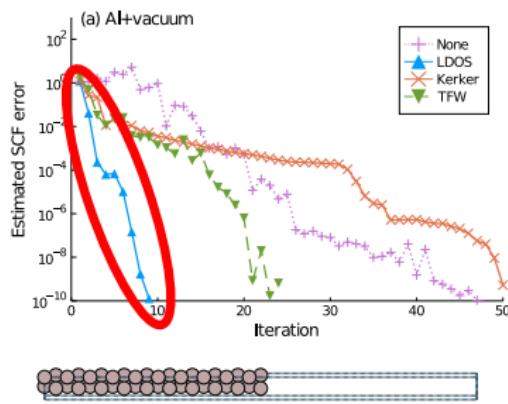
- “Row-sum mass lumping”:

$$\begin{aligned} \int \chi_0(\underline{\mathbf{r}}, \underline{\mathbf{r}'}) V(\underline{\mathbf{r}'}) \mathrm{d}\underline{\mathbf{r}'} &\simeq V(\underline{\mathbf{r}}) \int \chi_0(\underline{\mathbf{r}}, \underline{\mathbf{r}'}) \mathrm{d}\underline{\mathbf{r}'} \\ &= V(\underline{\mathbf{r}}) \sum_{n,m} \frac{f(\varepsilon_n) - f(\varepsilon_m)}{\varepsilon_n - \varepsilon_m} \psi_n(\underline{\mathbf{r}}) \psi_m^*(\underline{\mathbf{r}}) \delta_{mn} \\ &= V(\underline{\mathbf{r}}) \sum_n f'(\varepsilon_n) |\psi_n(\underline{\mathbf{r}})|^2 \\ &= V(\underline{\mathbf{r}}) \cdot \left( -\text{LDOS}(\underline{\mathbf{r}}) \right) \end{aligned}$$

---

<sup>1</sup>MFH, A. Levitt. J. Phys. Condens. Matter 33, 085503 (2021).

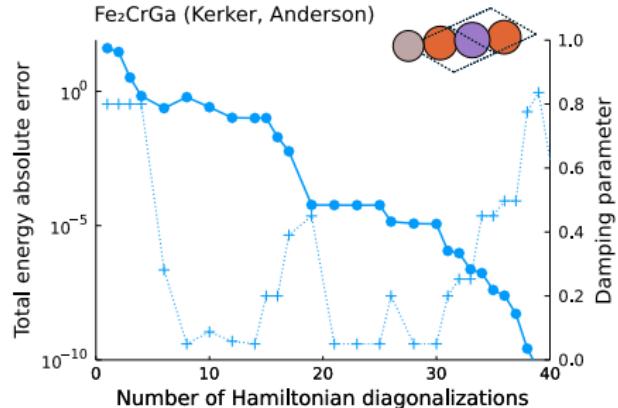
# Black-box $P$ : LDOS preconditioning<sup>1</sup>



- Inhomogeneous material: Aluminium metal + Insulator
- LDOS automatically interpolates between Kerker mixing (suitable for metals) and no mixing (suitable for insulators)
  - ⇒ Based on mathematical understanding of screening
  - ⇒ Parameter-free and black-box
- > 50% less SCF iterations

<sup>1</sup>MFH, A. Levitt. J. Phys. Condens. Matter **33**, 085503 (2021).

# Black-box $\alpha$ : Adaptive damping<sup>1</sup>



- **Theorem:** SCF convergence guaranteed if  $\alpha$  small enough (see paper)
- $\alpha$  adapted *in each step* using line search & quadratic model
- Novelty: Reuse of expensive quantities in next SCF step  
⇒ No overhead if line search immediately successful
- For tricky systems: Adaptive damping has an overhead
  - But: **Avoids trial and error**
  - **Mathematically motivated safeguard mechanism**

<sup>1</sup>MFH, A. Levitt. J. Comput. Phys. **459**, 111127 (2022).

# Contents

② SCF convergence and LDOS

③ Schur and inexact Krylov

# What about properties ?

- DFT properties: Response of system to external changes:

- Connection Theory  $\Leftrightarrow$  Experiment
- Modelling: Potential  $V(\theta, \rho)$  depends on parameters  $\theta$   
(e.g. atomic positions, el. field)

- SCF procedure yields fixed-point density  $\rho_{SCF}$

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

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

- 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<sup>n</sup>)

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

## What about inverse design ?

- $\theta$ : Design parameters (e.g. stress, dopant concentration)
- SCF procedure yields fixed-point density  $\rho_{SCF}$

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

- From these compute quantities of interest:  $Q(\rho)$
- **Question:** How to chose  $\theta$  to reach target ?

$$\frac{dQ}{d\rho} = \frac{\partial Q}{\partial \theta} + \frac{\partial Q}{\partial \rho} \frac{\partial \rho}{\partial \theta}$$

⇒ Need again Density-functional perturbation theory

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

- Algorithmic differentiation:
  - Avoid coding up all these derivatives by hand !
  - Accessible to non-DFT experts !

# Solving density-functional perturbation theory

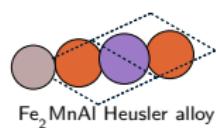
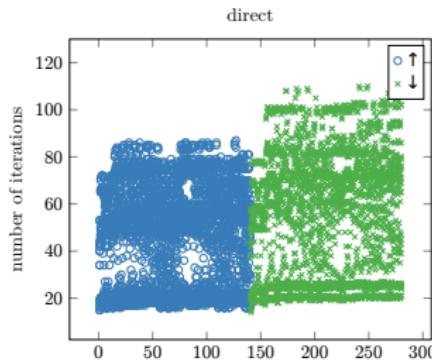
$$\frac{\partial \rho_{SCF}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta}$$

- Inversion performed iteratively
  - ⇒ Requires many products  $\chi_0 \delta V$
- Nested problem:  $\chi_0 \delta V$  solves  $N$  Sternheimer equations

$$(\tilde{H} - \varepsilon_i) \delta \psi_i = -P \delta V \psi_i \quad \forall i = 1, \dots, N$$

$H = -\frac{1}{2}\Delta + V$ ,  $\tilde{H} = PHP$  and  $P$  some projector;  $(\varepsilon_i, \psi_i)$  eigenpairs of  $H$

- ⇒ How to choose inner tolerance ?
- ⇒ Badly conditioned for metallic systems ( $\varepsilon_i$  near eigenvalue of  $\tilde{H}$ )



# Schur complement approach to response<sup>1</sup>

- SCF diagonalisations yield  $N_{\text{ex}}$  “extra” orbitals  
 $\Phi = (\psi_{N+1}, \dots, \psi_{N+N_{\text{ex}}})$  spanning  $T$ .
- Not converged eigenvector, but use to partition  $\tilde{H}$ :

$$\tilde{H} = \begin{pmatrix} E_{\text{ex}} & \mathbf{C} \\ \mathbf{C}^\dagger & \mathbf{R} \end{pmatrix} \quad \text{where} \quad \begin{aligned} E_{\text{ex}} &= \text{diag}(\varepsilon_{N+1}, \dots, \varepsilon_{N+N_{\text{ex}}}) \\ \mathbf{C} &= \Phi \Phi^\dagger \tilde{H} (1 - \Phi \Phi^\dagger) \\ \mathbf{R} &= (1 - \Phi \Phi^\dagger) \tilde{H} (1 - \Phi \Phi^\dagger) \end{aligned}$$

⇒ Use Schur complement:

- Solve for  $\Phi \Phi^\dagger \delta \psi_n$  exactly
- $x = (1 - \Phi \Phi^\dagger) \delta \psi_n$  obtained by (b some RHS)

$$(\mathbf{R} - \mathbf{C}^\dagger E_{\text{ex}}^{-1} \mathbf{C} - \varepsilon_n) x = b$$

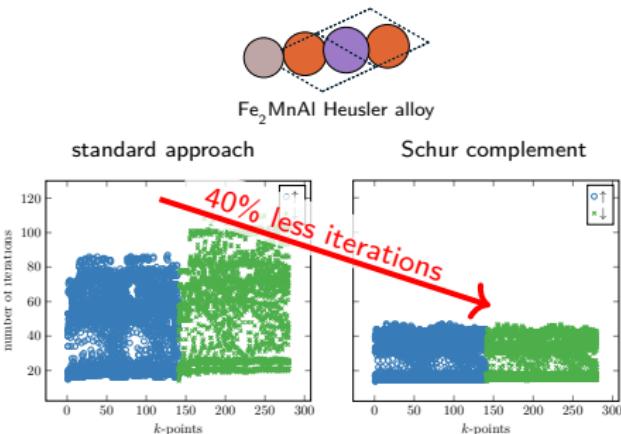
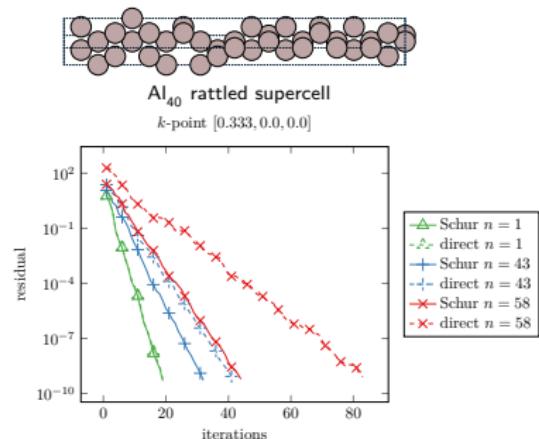
- Smallest eigenvalue about  $\varepsilon_{N+N_{\text{ex}}} - \varepsilon_N$

⇒ Conditioning improved, savings on CG iterations

---

<sup>1</sup>E. Cancès, MFH, G. Kemlin, et. al. Lett. Math. Phys. 113, 21 (2023).

# Schur-based response: Numerical examples<sup>1</sup>



- Largest reduction in iterations near Fermi level ( $n = 58$ ) (where gap is smallest)
- Overall 17% less iterations
- ⇒ Improvement comes for free (extra bands needed during SCF)

- Relevant materials class with unusual magnetic properties
- Translates to challenging numerical behaviour
- Schur-based approach tames CG
- ca. 40% less iterations

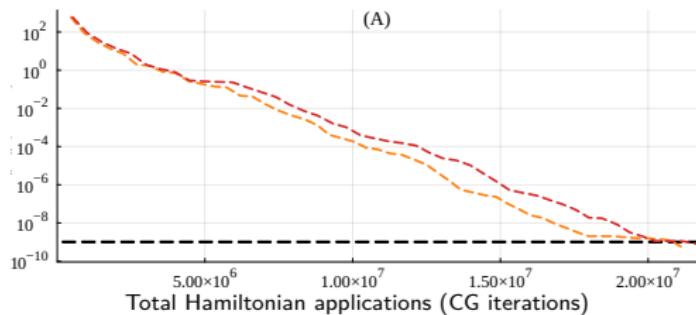
<sup>1</sup>E. Cancès, MFH, G. Kemlin, et. al. Lett. Math. Phys. **113**, 21 (2023).

# Choosing the Sternheimer tolerance

$$\frac{\partial \rho_{\text{SCF}}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta} \quad (\tilde{H} - \varepsilon_n) \delta \psi_n = -P \delta V \psi_n \quad \forall i = 1, \dots, N$$

GMRES tolerance  $\tau$                                     CG tolerance  $\tau_{i,n}^{\text{CG}}$

- Dyson + Sternheimer: Nested iteratively solved problems
  - Tolerance for CGs when applying  $\chi_0$  ?
  - Naive strategies:  $\tau_{i,n}^{\text{CG}} = \tau/100$  and  $\tau_{i,n}^{\text{CG}} = \tau/10$  for  $\tau = 10^{-9}$



- Dashed: GMRES estimated residual norm
  - Solid: Actual residual norm

- Fail by 3 orders ( $\text{Al}_{40}$  supercell)

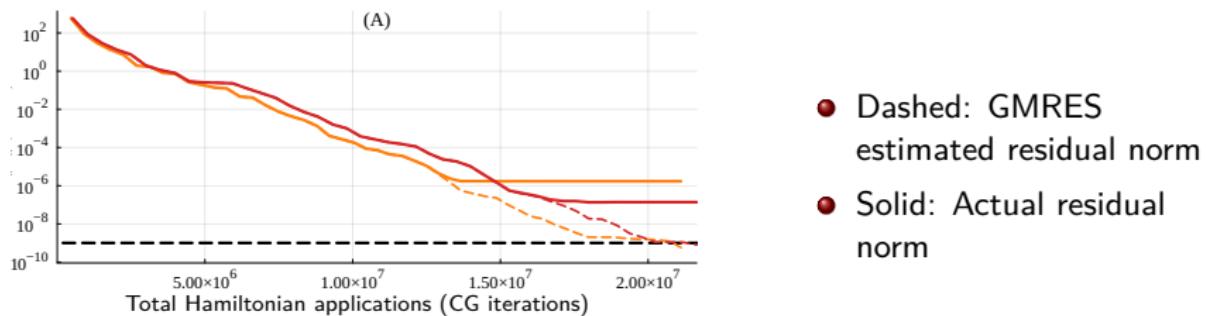
⇒ Need adaptive & guaranteed strategy for  $\tau_{i,n}^{\text{CG}}$

# Choosing the Sternheimer tolerance

$$\frac{\partial \rho_{\text{SCF}}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta} \quad (\tilde{H} - \varepsilon_n) \delta \psi_n = -P \delta V \psi_n \quad \forall i = 1, \dots, N$$

GMRES tolerance  $\tau$     CG tolerance  $\tau_{i,n}^{\text{CG}}$

- Dyson + Sternheimer: Nested iteratively solved problems
- Tolerance for CGs when applying  $\chi_0$  ?
- Naive strategies:  $\tau_{i,n}^{\text{CG}} = \tau/100$  and  $\tau_{i,n}^{\text{CG}} = \tau/10$  for  $\tau = 10^{-9}$



- Fail by 3 orders ( $\text{Al}_{40}$  supercell)
- ⇒ Need adaptive & guaranteed strategy for  $\tau_{i,n}^{\text{CG}}$

# Outlook: Inexact Krylov methods (1)

- **Inexact GMRES:** Tolerable errors for  $1 - \chi_0 K$
- ⇒ **Theorem:**<sup>1</sup> Guaranteed convergence of GMRES to  $\tau$  when

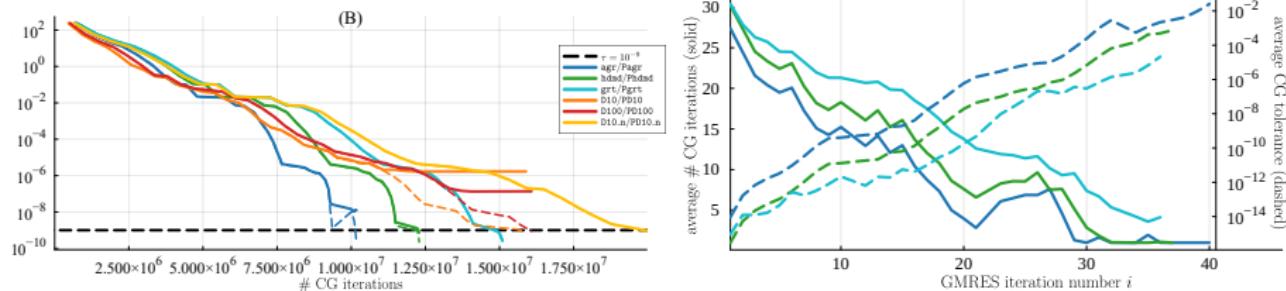
$$\tau_{i,n}^{\text{CG}} \lesssim \frac{1}{k \mathcal{C}} \frac{\sqrt{|\Omega|}}{N \text{Ecut}^{3/4}} \frac{1}{f_n} \frac{s}{3m \|\tilde{r}_{i-1}\|} \tau$$

- $\|\tilde{r}_{i-1}\|$ : GMRES estimated residual norm
- $s$ : Estimate for cond. num. of GMRES Hessenberg matrix  
(updated on the fly)
- $m$ : GMRES maximal subspace size
- $k$ : Constants of order 1
- $\mathcal{C}$ : System size-indep. const.
- Main features:
  - Looser tolerance closer to convergence (as  $\tilde{r}_{i-1} \rightarrow 0$ )
  - Looser tolerance for small  $f_n$  (when Sternheimer worst conditioned)
  - Tighter tolerance for larger systems (as  $\frac{\sqrt{|\Omega|}}{N} \searrow$ )

---

<sup>1</sup>B. Sun, MFH, *in preparation*.

# Outlook: Inexact Krylov methods (2)<sup>1</sup>



- Guaranteed (grt) computes  $C$  exactly
  - Balanced (hdmd) sets  $C = 1$ 
    - Requires a good preconditioner for metals (since  $\|Kv_i\|$  dropped)  
⇒ We employ standard Kerker preconditioner also in GMRES
  - Aggressive (agr) drops even more constants
    - Even faster than hdmd, but can be a factor 10 off
  - From about 20M to 12M Hamiltonian applications  
(the expensive step)
- ⇒ Superlinear convergence

<sup>1</sup>B. Sun, MFH, *in preparation.*