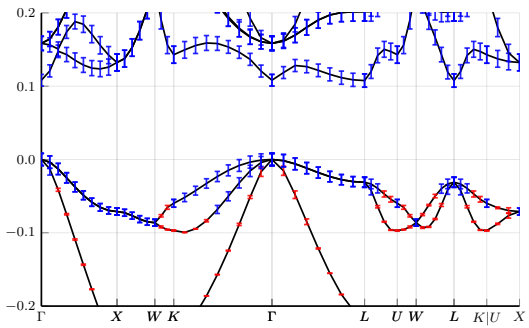
Efficient response property calculations for density-functional theory

Eric Cancès, Michael F. Herbst*, Gaspard Kemlin,
Antoine Levitt, Benjamin Stamm

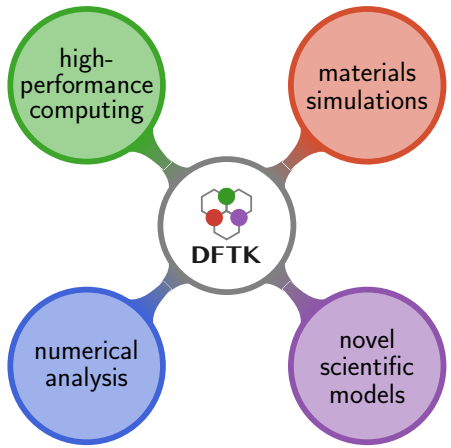
*Mathematics for Materials Modelling (matmat.org), EPFL

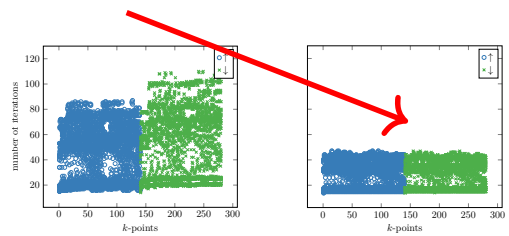
6 June 2023

Slides: https://michael-herbst.com/talks/2023.06.06_nmqc_response.pdf



Understanding errors for DFT



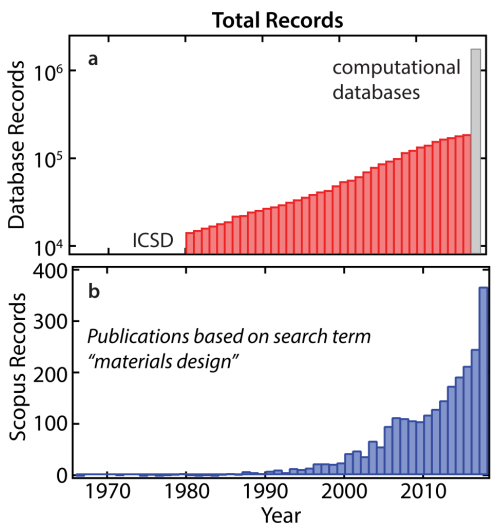


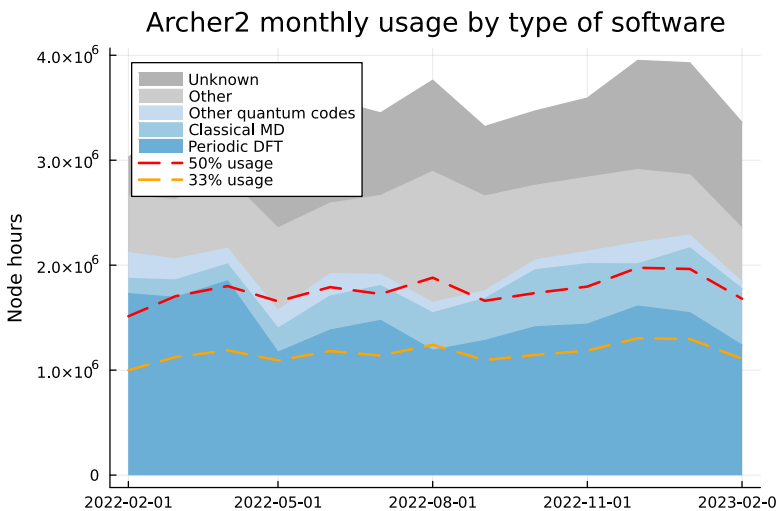
Schur complement approach to response

Tackling to 21st century challenges

- 21st century challenges:
  - Renewable energy, green chemistry, health care ...
- Current solutions limited by properties of available materials

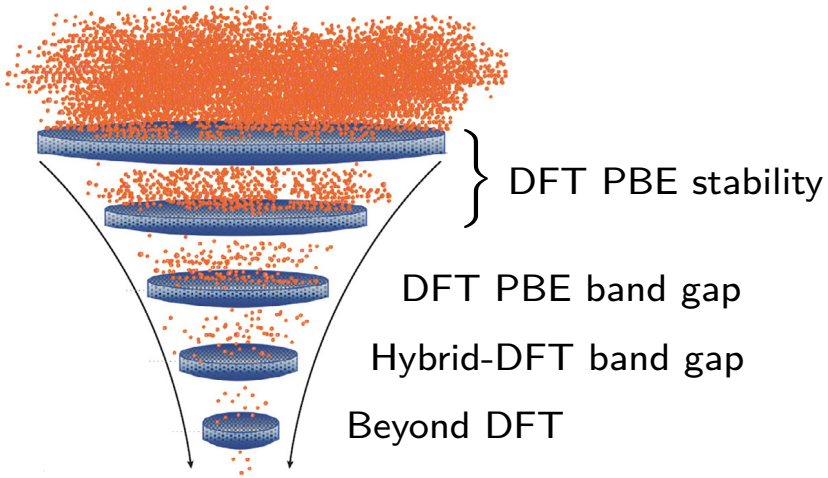
  $\Rightarrow$ Innovation driven by discovering new materials
- Crucial tool: Computational materials discovery
  - Systematic simulations on $\simeq 10^4 - 10^6$ compounds
  - Complemented by data-driven approaches
  - Noteworthy share of world's supercomputing resources



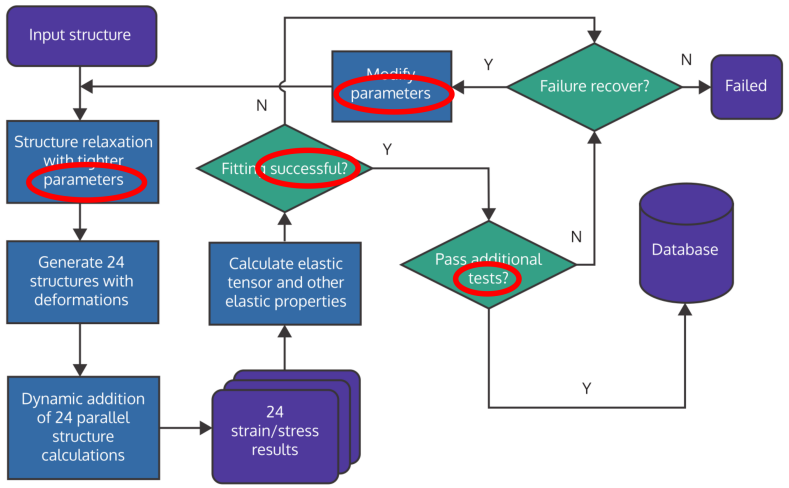


K. Alberi *et. al.* J. Phys. D, **52**, 013001 (2019).

Sketch of high-throughput workflows



Design funnel for photovoltaic materials



Workflow for computing elasticity tensors

- Many parameters to choose (algorithms, tolerances, models)
  - Elaborate heuristics: Failure rate $\simeq 1\%$
  - Still: Thousands of failed calculations

  $\Rightarrow$ Wasted resources & increased human attention (limits througput)

- **Goal:** Self-adapting black-box algorithms
  - Transform empirical wisdom to built-in convergence guarantees
  - Requires: Uncertainty quantification & error estimation

  $\Rightarrow$ Understand where and how to spend efforts best

G. Hautier Comput. Mater. Sci. **164**, 108 (2019); L. Himanen *et. al.* Adv. Science **6**, 1900808 (2019).

Error sources in DFT simulations

- Model error: Selection of DFT model
- Computational approach:
  - Discretisation error: Basis size, k-point mesh
  - Algorithm error: Convergence thresholds (SCF, eigensolver)
  - Floating-point error: Floating-point arithmetic
- Additionally: Programming error, hardware error (not discussed further)

- Error control: Link parameter selection $\leftrightarrow$ simulation error
  - Remarkable progress in mathematical research on DFT
  - **Goal of this work:** Reliable computation of DFT sensitivities

  $\Rightarrow$ Understand influence of DFT model on predicted properties

Density-functional theory

- Self-consistent field procedure: Fixed-point problem

$$F(V_{\text{ext}} + V_{\text{Hxc}}(\rho_{\text{SCF}})) = \rho_{\text{SCF}}$$
$$\text{where} \qquad V_{\text{Hxc}}(\rho) = v_C \rho + V_{\text{XC}}(\rho)$$

- $F(V)$ is the potential-to-density map (i.e. diagonalisation)

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

- ε_F chosen such that $\int F(V) = N$ (number of electrons)
- nuclear attraction V_{nuc}, exchange-correlation V_{XC}, Hartree potential $-\Delta\left(v_C \rho\right) = 4\pi\rho$, ψ_i orthogonal, f: Occupation function between 0 and 2

DFT properties and sensitivities

- SCF defines mapping $V_{\text{ext}} \mapsto \rho_{\text{SCF}}$ $(F(V_{\text{ext}} + V_{\text{Hxc}}(\rho_{\text{SCF}})) = \rho_{\text{SCF}})$
- DFT properties: Response of system to external perturbation

  $\Rightarrow$ (Higher-order) derivative of some function of ρ_{SCF}

  - Examples: Forces (energy wrt. position), dipole moment (energy wrt. el. field), elasticity (energy cross-response to lattice deformation)

$\Rightarrow$ Goal: Understand derivative of SCF mapping

- Density-functional perturbation theory (CP-SCF, . . .)

- Link to DFT model sensitivities: Consider the V_{ext} parameters:
  - a: Lattice constant
  - θ: DFT exchange-correlation parameters

Stress
$$S(a, \theta) = \frac{\partial \mathcal{E}(\rho_{\text{SCF}}(a, \theta))}{\partial a}$$

Model sensitivity
$$\frac{dS}{d\theta} = \frac{\partial S}{\partial \rho_{\text{SCF}}} \frac{\partial \rho_{\text{SCF}}}{\partial \theta}$$

Density-functional perturbation theory

$$F(V_{\text{ext}} + V_{\text{Hxc}}(\rho_{\text{SCF}})) = \rho_{\text{SCF}}$$

- δV: Perturbation to V_{ext}, by chain rule

$$\delta\rho = F'(V_{\text{ext}} + V_{\text{Hxc}}(\rho_{\text{SCF}})) \cdot (\delta V + K_* \delta\rho)$$

$$\Leftrightarrow \quad \delta\rho = (1 - \chi_0 K)^{-1} \chi_0 \delta V$$

where $K_* = V'_{\text{Hxc}}(\rho_{\text{SCF}})$, $\chi_0 = F'(V_{\text{ext}} + V_{\text{Hxc}}(\rho_{\text{SCF}}))$

- Dyson equation: Solved by iterative methods
- Adler-Wiser formula (using $f_n = f(\varepsilon_n)$):

$$\delta\rho(r) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m} \psi_n^*(r)\psi_m(r) \left(\delta V_{mn} - \delta\varepsilon_F \delta_{nm}\right)$$

under the convention

$$\frac{f_n - f_n}{\varepsilon_n - \varepsilon_n} = \frac{1}{T} f'\left(\frac{\varepsilon_n - \varepsilon_F}{T}\right) = f'_n$$

and where $\delta V_{mn} = \langle \psi_m | \delta V \psi_n \rangle$, $\delta\varepsilon_F$ has an explicit formula

Getting rid of infinities (1)

- Represent $\delta\rho$ by variations $\delta\psi_n$ and δf_n[1] (our new unknowns)

$$\delta\rho(r) = \sum_{n=1}^{N} 2 f_n \operatorname{Re}\left(\psi_n^*(r)\delta\psi_n(r)\right) + \delta f_n \left|\psi_n(r)\right|^2$$

  where $\delta f_n = f_n'(\delta V_{nn} - \delta\varepsilon_F)$

- Define:
  - $P = \operatorname{span}\{\psi_n \,|\, n = 1, \ldots, N\}$: Space spanned by N lowest eigenpairs (ε_n, ψ_n) of H (occupied subspace)
  - $\Pi_Q = 1 - \Pi_P$ with Π_P projector onto P.
- Separate the contributions:

$$f_n\delta\psi_n = f_n\delta\psi_n^P + f_n\delta\psi_n^Q$$

- Note: We deal with the setting of *many* basis functions
  (Plane waves, wavelets, finite elements, real-space, . . .)

  $\Rightarrow$ We cannot compute all eigenpairs of H

[1] E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023).

Getting rid of infinities (2)

- occupied-occupied $\delta\psi_n^P$: Use sum over states

$$f_n \delta\psi_n^P = \sum_{m=1, m\neq n}^{N} \Gamma_{mn}\psi_m$$

where we need $\Gamma_{nn} = 0$ and

$$\Gamma_{mn} + \Gamma_{nm}^* = \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m}\delta V_{mn}$$

- Question 1: This is not unique. How to choose Γ_{nm}?

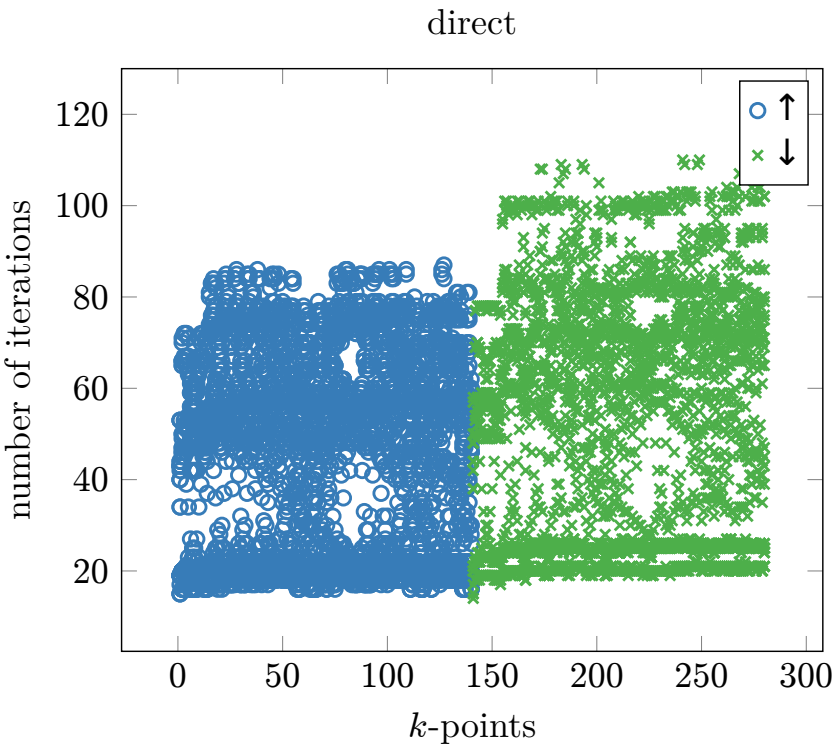
Getting rid of infinities (3)

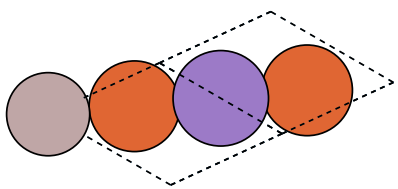
- unocc-occ $\delta\psi_n^Q$: Use Sternheimer equation

$$\Pi_Q(H-\varepsilon_n)\Pi_Q\delta\psi_n = -\Pi_Q\delta V\psi_n \qquad \forall n = 1,\ldots,N \quad (*)$$

- Question 2: $(*)$ is badly conditioned if gap $\varepsilon_{N+1} - \varepsilon_N$ small

  $\Rightarrow$ How can we make response cheaper for metals?





Fe_2MnAl Heusler alloy

Contents

EPFL MXtMat

The bad choice: Orthogonal gauge

- Recall, we need

$$\Gamma_{mn} + \Gamma^*_{nm} = \Delta_{mn} = \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m} \delta V_{mn}$$

  and additionally $\Gamma_{mn} = \langle \psi_m | f_n \delta\psi_n \rangle$ by construction

- Zero temperature (insulators): $\delta\psi^P = 0$

$\Rightarrow$ Orbitals can be kept orthogonal under response (for insulators)

- **Orthogonal gauge**: Enforce orthogonality in all cases, i.e.

$$\begin{aligned} 0 &= \delta \langle \psi_m | \psi_n \rangle = \langle \delta\psi_m | \psi_n \rangle + \langle \psi_m | \delta\psi_n \rangle \\ \Rightarrow \qquad 0 &= \Gamma_{mn}/f_n + \Gamma^*_{nm}/f_m \\ \Rightarrow \qquad \Gamma^{\text{orth}}_{mn} &= \frac{f_n}{\varepsilon_n - \varepsilon_m} \delta V_{mn} \end{aligned}$$

- Problem: This can lead to a large contribution as $\varepsilon_n \to \varepsilon_m$ which is almost compensated by $\Gamma^{\text{orth},*}_{nm}$

$\Rightarrow$ Loss of numerical precision

The optimal choice: Minimal gauge

- Minimise the size of all contributions to $\delta\psi_n$, i.e.

$$\min \sum_{m,n} \frac{1}{f_n^2} |\Gamma_{mn}|^2$$

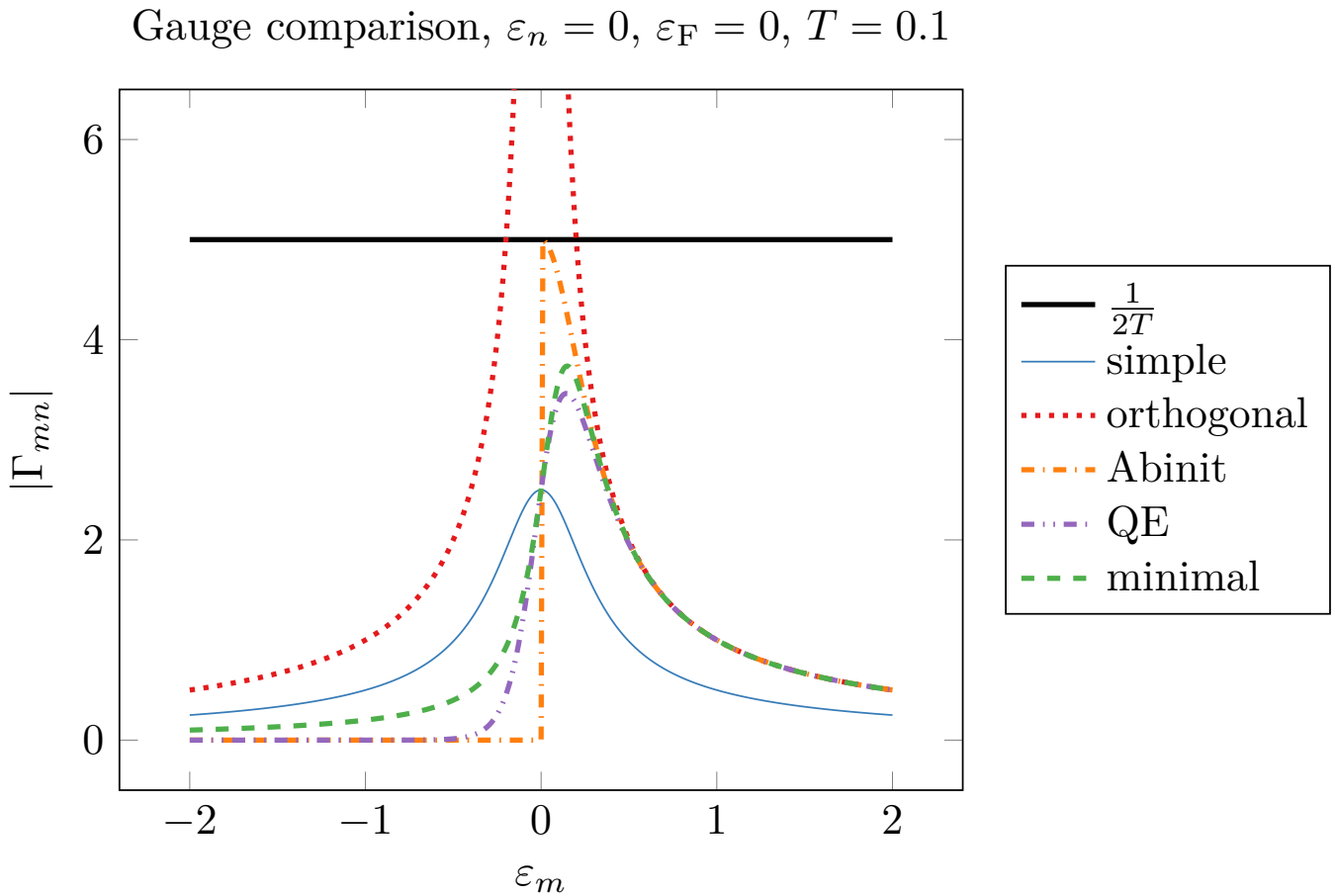
$$\text{s.t. } \Gamma_{mn} + \Gamma_{nm}^* = \Delta_{mn} = \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m} \delta V_{mn}$$

- **Minimal gauge:** Solution to above problem

$$\Gamma_{mn} = \frac{f_n^2}{f_n^2 + f_m^2} \Delta_{mn}$$

- Other gauge choices:
  - Quantum Espresso: $\Gamma_{mn} = f_{\text{FD}}\left(\frac{\varepsilon_n - \varepsilon_m}{T}\right) \Delta_{mn}$
  - Abinit: $\Gamma_{mn} = \mathbb{1}_{f_n > f_m} \Delta_{mn}$

Comparison of gauges



- Graph investigates the growth of $\delta\rho$ wrt. δV
- $\frac{1}{2T}$ gives a lower bound (from Δ_{mn})

$\Rightarrow$ Orthogonal should be avoided, all others reasonable

Contents

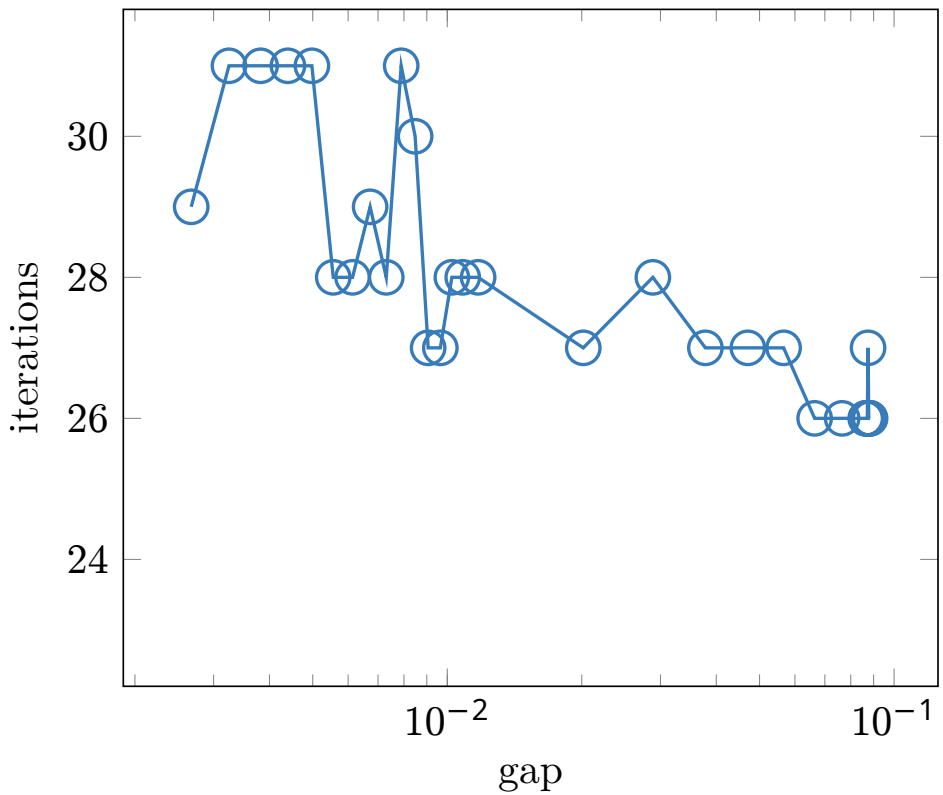
EPFL MXtMat

Extra SCF orbitals[1]

- Each *application* of χ_0 to a δV requires solving Sternheimer for all $n = 1, \ldots, N$

$$\Pi_Q (H - \varepsilon_n) \Pi_Q \delta\psi_n = -\Pi_Q \delta V \psi_n$$

- If gap $\varepsilon_{N+1} - \varepsilon_N$ closes (metals), conditioning gets worse



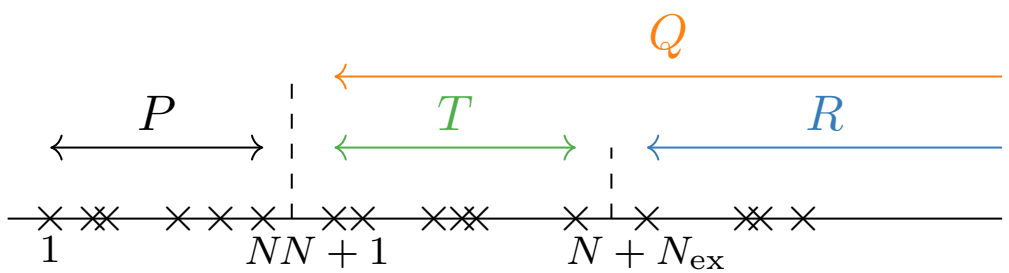
[1] E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023).

Extra SCF orbitals[1]

- Each *application* of χ_0 to a δV requires solving Sternheimer for all $n = 1, \ldots, N$

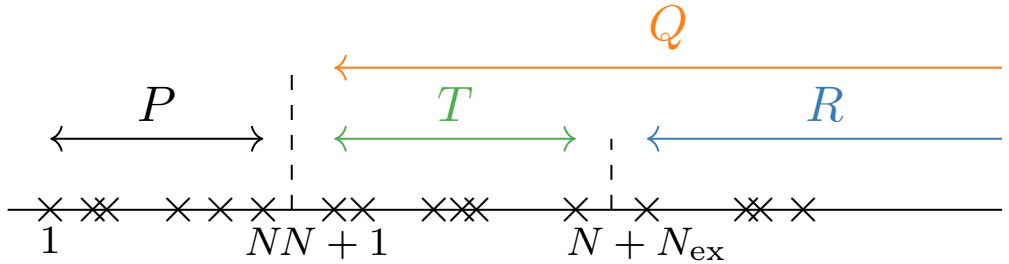
$$\Pi_Q (H - \varepsilon_n) \Pi_Q \delta\psi_n = -\Pi_Q \delta V \psi_n$$

- If gap $\varepsilon_{N+1} - \varepsilon_N$ closes (metals), conditioning gets worse
- But we have not used all we know:
  - Standard iterative diagonalisations (and thus SCFs) yield N_{ex} additional orbitals $\Phi = (\psi_{N+1}, \ldots, \psi_{N+N_{\text{ex}}})$
  - Notable property: $\Phi^T H \Phi = \text{diag}(\varepsilon_{N+1}, \ldots, \varepsilon_{N+N_{\text{ex}}})$
  - Not fully converged, i.e. $H\psi_n \neq \varepsilon_n \psi_n$ for $n = N+1, \ldots, N+N_{\text{ex}}$



[1] E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023).

Splitting the orbital space[1]



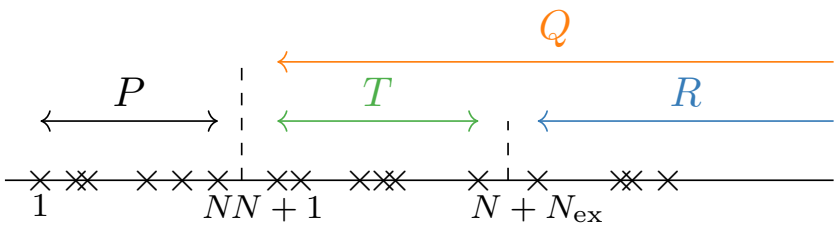
- Overview:
  - P: Fully converged, occupied orbitals
  - T: Non-occupied, not converged
  - R: Completely unknown states
- $I = \Pi_P + \Pi_Q = \Pi_P + \Pi_T + \Pi_R$
- Hamiltonian structure:

$$H = \begin{pmatrix} E & 0 & 0 \\ 0 & E_{\text{ex}} & \Pi_T H \Pi_R \\ 0 & \Pi_R H \Pi_T & \Pi_R H \Pi_R \end{pmatrix}$$

where $E = \operatorname{diag}(\varepsilon_1, \ldots, \varepsilon_N)$ and $E_{\text{ex}} = \operatorname{diag}(\varepsilon_{N+1}, \ldots, \varepsilon_{N+N_{\text{ex}}})$

[1] E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023).

Exploiting block structure[1]



$$\Pi_Q(H-\varepsilon_n)\Pi_Q\delta\psi_n = -\Pi_Q\delta V\psi_n$$

Sternheimer equation

$$H-\varepsilon_n = \begin{pmatrix} E-\varepsilon_n & 0 \\ 0 & \Pi_Q(H-\varepsilon_n)\Pi_Q \end{pmatrix} = \begin{pmatrix} E-\varepsilon_n & 0 & 0 \\ 0 & E_{\text{ex}}-\varepsilon_n & \Pi_T H\Pi_R \\ 0 & \Pi_R H\Pi_T & \Pi_R(H-\varepsilon_n)\Pi_R \end{pmatrix}$$

- Invert $\Pi_Q(H-\varepsilon_n)\Pi_Q$
- $n = N$: Possibly ill-conditioned as $\varepsilon_{N+1} - \varepsilon_N \to 0$

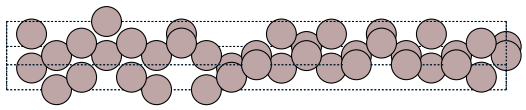
- $E_{\text{ex}} - \varepsilon_n$ diagonal: Inversion for free
- Only invert $\Pi_R(H-\varepsilon_n)\Pi_R$

$\Rightarrow$ Better conditioned as $\varepsilon_{N+N_{\text{ex}}} - \varepsilon_N > \varepsilon_{N+1} - \varepsilon_N$

- Non-zero off-diagonal parts: Schur complement
  - A bit tedious $\Rightarrow$ Ask me for details

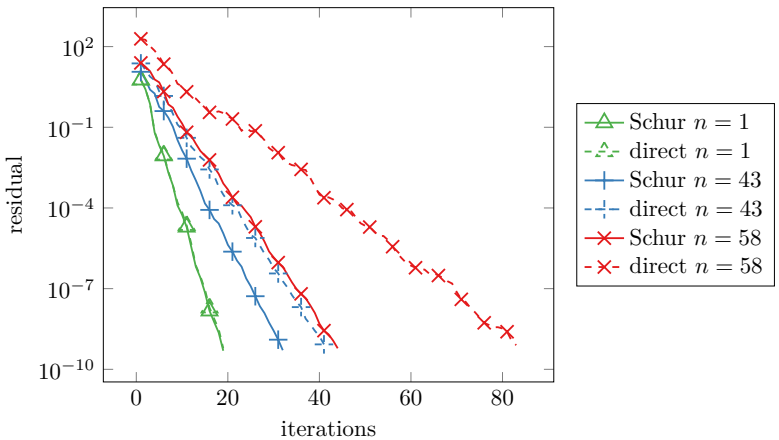
[1] E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023).

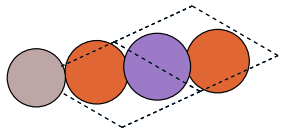
Schur-based response: Numerical examples[1]



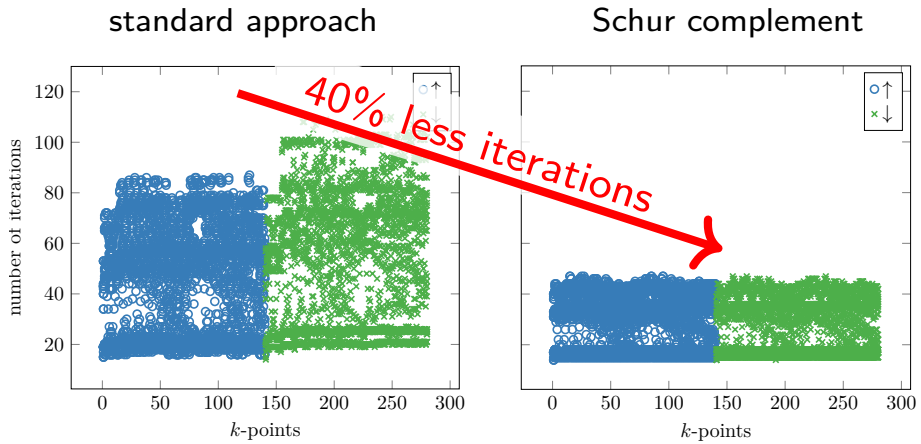
Al_{40} rattled supercell

k-point $[0.333, 0.0, 0.0]$





Fe_2MnAl Heusler alloy



- Largest reduction in iterations near Fermi level ($n = 58$)
  (where gap is smallest)
- Overall 17% less iterations

$\Rightarrow$ Improvement comes for free
(extra bands needed during SCF)

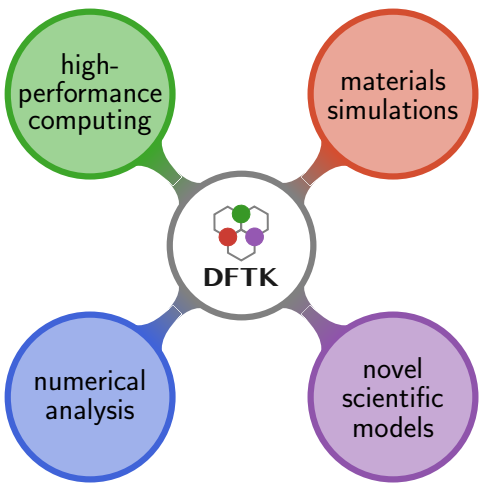
- Relevant materials class with unusual magnetic properties
- Translates to challenging numerical behaviour
- Schur-based approach tames CG
- ca. 40% less iterations

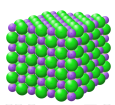
[1] E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023).

Contents

EPFL MXtMat

Density-functional toolkit[1] — https://dftk.org





$H\Psi = E\Psi$

- julia code for plane-wave DFT, started in 2019
- Fully composable due to julia abstractions:
  - Arbitrary precision (32bit, >64bit, . . .)
  - Algorithmic differentiation (AD)
  - HPC tools: GPU acceleration, MPI parallelisation
- Low barriers for cross-disciplinary research:
  - Allows restriction to relevant model problems,
  - *and* scale-up to application regime (1000 electrons)
  - Sizeable feature set in 7500 lines of code
  - Including some unique features (Self-adapting algorithms)
- Accessible high-productivity research framework:
  - Key code contributions by undegrads / PhD students
  - AD support in 10 weeks (CS Bachelor)
  - GPU support in 10 weeks (Physics Bachelor)
  - Relevant contributions from outside collab. circle

Lattice constant sensitivities in DFTK

```
function dft_energy(a, θ)
    model = model_DFT(make_structure(a), PbeExchange(θ))
    basis = PlaneWaveBasis(model; Ecut=..., kgrid=... )
    self_consistent_field(basis).energies.total
end
optimise_lattice(θ) = optimise(a -> dft_energy(a, θ))

sensitivities =
    ForwardDiff.gradient(optimise_lattice, [κ, β])
```

(Å)	a_*	κ	$\frac{da_*}{d\kappa}$	β	$\frac{da_*}{d\beta}$
expmnt.	5.421				
PBEsol	5.449	0.804	0.713	0.0375	0.0058
PBE	5.461	0.804	0.550	0.0667	0.0194
APBE	5.465	0.804	0.482	0.0790	0.0269
PBEmol	5.467	0.804	0.456	0.0838	0.0301
XPBE	5.466	0.920	0.603	0.0706	0.0184
rev-PBE	5.467	1.245	0.744	0.0667	0.0099

Model sensitivities for the silicon lattice constant

- Optimal lattice constant sensitivities in one line of code

$$a_* = \arg\min_a \mathcal{E}(a, \theta) \qquad \text{sensitivities} = \frac{da_*}{d\theta}$$

- Practical challenges for derivation and implementation:
  - Nested iterative methods (eigensolver, SCF, lattice optimisation)
  - Unusual second-order derivatives (e.g. $\frac{\partial S}{\partial \theta} = \frac{\partial^2 \mathcal{E}}{\partial \theta \partial a}$)
  - Support for future DFT models? (with their different parameters θ)
- **DFTK** key achievements:
  - Integration with julia's frameworks for algorithmic differentiation (AD)
  - Floating-point agnostic design
  - Stable & generic response solver (this talk)
- Fully flexible in DFT model or targeted quantity:
  - Saves manual coding: Request gradient, AD delivers

  ⇒ Breaks "one PhD student per derivative" paradigm

Summary

- Challenges of response calculations for metals
  - Closing gap worsens conditioning of linear system
  - Ambiguity in representing density response (gauge freedom)
- Mathematical analysis of DFPT
  - Novel Schur-complement approach to response
  - Up to 40% faster, while no additional cost
  - Applicable to all "large basis set" methods
  - Readily available in DFTK
- Enables fast & robust derivative computations (in combination with AD)
  - Routine sensitivity analysis & UQ
  - Development of data-enhanced models
- DFTK: Multidisciplinary software development
  - julia-based framework for new DFT algorithms
  - High-productivity research framework
  - In one code: Reduced problems and scale-up to realistic applications

  $\Rightarrow$ Sketch new methods & test in HPC context

Open PhD & PostDoc positions in the MatMat group

Possible topics include:

- Uncertainty quantification for DFT:
  Error in data-driven DFT models, pseudopotentials, propagation to properties and MD potentials
- Self-adapting numerical methods for high-throughput DFT simulations
- See `https://matmat.org/jobs/`

- Interdisciplinary research linking maths and simulation:
  - Become part of maths and materials institutes @ EPFL
- Collaboration inside MARVEL:
  - Reproducible workflows & sustainable software
  - Computational materials discovery
  - Statistical learning methods

EPFL MatMat

Questions?

MxtMat https://matmat.org

mfherbst

michael.herbst@epfl.ch

E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023). DOI 10.1007/s11005-023-01645-3

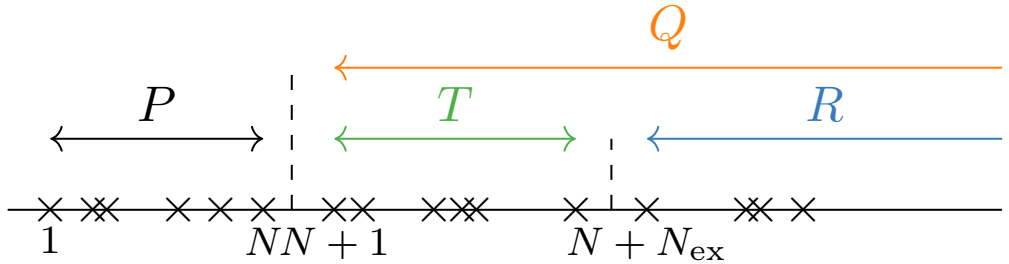
DFTK https://dftk.org

EPFL MxtMat

Contents

EPFL MXtMat

Schur-complement approach[1] (1)



- We want to solve

$$\Pi_Q (H - \varepsilon_n) \Pi_Q \delta\psi_n = \underbrace{-\Pi_Q \delta V \psi_n}_{:= b_n}$$

- Split orbital perturbation $\Pi_Q \delta\psi_n = \Phi\alpha_n + \Pi_R \delta\psi_n^R$ to obtain:

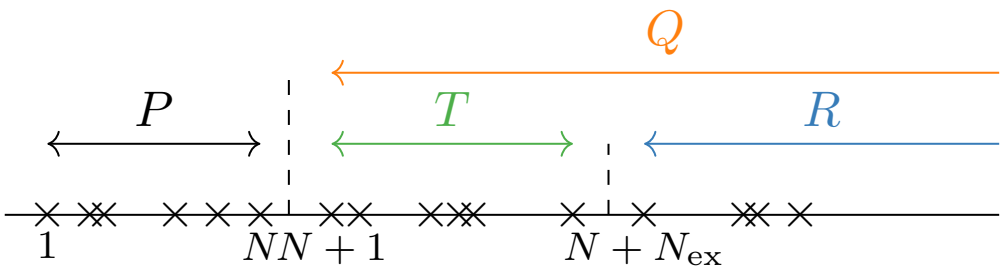
$$\Pi_Q (H - \varepsilon_n) \Phi\alpha_n + \Pi_Q (H - \varepsilon_n) \Pi_R \delta\psi_n^R = b_n$$

- Schur complement: Solve component in T (along Φ) explicitly:

$$\alpha_n = \underbrace{\left(\Phi^T H \Phi\right)^{-1}}_{= D^{-1}} \left(\Phi^T b_n - \underbrace{\Phi^T \left(H - \varepsilon_n\right) \Pi_R}_{= h_{RT}^T} \delta\psi_n^R \right)$$

[1] E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023).

Schur-complement approach[1] (2)



$$\Pi_Q(H-\varepsilon_n)\Phi\alpha_n + \Pi_Q(H-\varepsilon_n)\Pi_R\delta\psi_n^R = b_n$$

$$\alpha_n = D^{-1}\left(\Phi^T b_n - h_{RT}^T\delta\psi_n^R\right)$$

- Insert α_n back and project with Π_R from the left:

$$\Pi_R(H-\varepsilon_n)\Phi\left[D^{-1}\left(\Phi^T b_n - h_{RT}^T\delta\psi_n^R\right)\right] + \Pi_R(H-\varepsilon_n)\Pi_R\delta\psi_n^R = \Pi_R b_n$$

$$\Rightarrow \left[\Pi_R(H-\varepsilon_n)\Pi_R - h_{RT}D^{-1}h_{RT}^T\right]\Pi_R\delta\psi_n^R = \left[\Pi_R - h_{RT}D^{-1}\Phi^T\right]b_n$$

- This can be solved for $\delta\psi_n^R$ using CG
- Φ are almost eigenvectors of H

$\Rightarrow$ Π_R almost removes small eigenmodes of $H-\varepsilon_N$

$\Rightarrow$ Improved conditioning

[1] E. Cancès, MFH, A. Levitt *et. al.* Lett. Math. Phys., **113**, 21 (2023).

Shifted Sternheimer approaches

- Some codes avoid the split $\delta\psi_n = \delta\psi_n^P + \delta\psi_n^Q$ (e.g. Quantum Espresso)
- Instead they solve a *shifted* Sternheimer equation

$$(H + S - \varepsilon_n)\delta\psi_n = -(f_n - S_n)\delta V$$

(with S chosen to make this non-singular and S_n chosen to give the correct $\delta\rho$)