

Kohn-Sham Density-Functional Theory: Perspectives on various inverse problems

Michael F. Herbst

Mathematics for Materials Modelling (matmat.org), EPFL

10 April 2026

Slides: https://michael-herbst.com/talks/2026.04.10_Swiss_Numerics_Day.pdf



Acknowledgements

 MxMat group



Niklas Schmitz



Bruno Ploumhans

Kohn-Sham inversion



Vebjørn Bakkestuen
(OsloMet)



Andre Laestadius
(OsloMet)

- Noe Blassel
- Nathanael Bosch
- Benedikt Menges



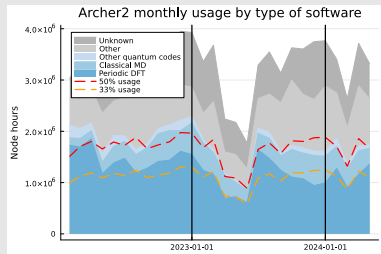
Impact of computational materials discovery



- 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 households¹
- ⇒ Complement experiment by **computational materials discovery**

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

Impact of computational materials discovery



- **>30% usage on key HPC clusters**
 - Many open mathematical problems
(Uncertainties, error estimation, robustness . . .)
- **This talk: Inverse problems; improving models**

⇒ Complement experiment by **computational materials discovery**

- 1 Kohn-Sham density-functional theory
- 2 Inverse Kohn-Sham using Moreau-Yosida regularisation
- 3 AD-DFPT: Efficient gradients for inverse problems
- 4 Wrapping up

Quantum-mechanical simulation of materials

- **Minimisation problem:** Ground state Ψ with energy

$$E(v) = \min_{\Psi \in W_N} \int_{\mathbb{R}^{3N}} \Psi^N(x_1, x_2, \dots, x_N) \hat{\mathcal{H}}(v) \Psi(x_1, x_2, \dots, x_N) dx_1 \cdots dx_N$$

- Physical system defines Hamiltonian $\hat{\mathcal{H}}$

$$\hat{\mathcal{H}}(v) = \sum_{j=1}^N -\frac{1}{2} \Delta_{x_j} + v(x_j) + \sum_{j < k} \frac{1}{\|x_j - x_k\|} = \hat{\mathcal{H}}^{\text{el}} + \sum_{j=1}^N v(x_j)$$

- **Antisymmetry** (Pauli exclusion principle):

$$\Psi_N \in W_N = \left\{ \Psi \in \bigwedge^N L^2(\mathbb{R}^3, \mathbb{C}) \cap H^1(\mathbb{R}^{3N}), \|\Psi\| = 1 \right\}$$

- **Remarks:**

- **Curse of dimensionality:** $N \simeq 100$ not unusual
- $v \in \mathcal{V}$ with standard choice $\mathcal{V} = L^{3/2} + L^\infty$ (\Rightarrow ground state exists)

- E.g. Coulomb potential $v(x) = - \sum_{m=1}^M \frac{Z_m}{\|x - R_m\|}$

M point charges at $R_1, \dots, R_M \in \mathbb{R}^3$, N electrons at $x_1, \dots, x_N \in \mathbb{R}^3$.

Density-functional theory or: Breaking the curse

- Electronic density: mapping $\mathbb{R}^3 \rightarrow \mathbb{R}$ with

$$\rho_{\Psi}(x) = N \int_{\mathbb{R}^{3(N-1)}} |\Psi(x, x_2, x_3, \dots, x_N)|^2 dx_2 \cdots dx_N$$

- Contains all physical information about the problem
- **Constrained-search formulation:** Minimise only over ρ

$$\begin{aligned} E(v) &= \inf_{\Psi \in W_N} \langle \Psi | \hat{\mathcal{H}}(v) \Psi \rangle = \inf_{\rho \in \mathcal{D}} \inf_{\substack{\Psi \in W_N \\ \rho_{\Psi} = \rho}} \langle \Psi | \hat{\mathcal{H}}(v) \Psi \rangle \\ &= \inf_{\rho \in \mathcal{D}} \inf_{\substack{\Psi \in W_N \\ \rho_{\Psi} = \rho}} \left(\langle \Psi | \hat{\mathcal{H}}^{\text{el}}(v) \Psi \rangle + \int_{\mathbb{R}^3} \rho_{\Psi}(x) v(x) dx \right) \\ &= \inf_{\rho \in \mathcal{D}} \left(F(\rho) + \int_{\mathbb{R}^3} \rho_{\Psi}(x) v(x) dx \right) \end{aligned}$$

with

$$F(\rho) = \inf_{\substack{\Psi \in W_N \\ \rho_{\Psi} = \rho}} \left\langle \Psi \left| \left(-\frac{1}{2} \sum_{j=1}^N \Delta_{x_j} + \sum_{j < k}^N \frac{1}{|x_j - x_k|} \right) \Psi \right. \right\rangle$$

- **Interpretation:** $E(v)$ is strictly concave in v ; $F(\rho)$ is **convex dual** (Legendre-Fenchel conjugate) to $E(v)$; \mathcal{D} dual space to \mathcal{V} .

Kohn-Sham density-functional theory (DFT)

- F is a **universal functional**: If known, could compute ρ for all v
- But F has **no closed form** and mathematically tricky¹:
 - Not necessarily convex, but can be made convex-lower-semicontinuous (by a double Legendre-Fenchel transformation)
 - Not differentiable (for $\rho \in \mathcal{D} = L^3 \cap L^1 = (L^{3/2} + L^\infty)^*$)
 - F is **NP-hard to compute** (Optimal transport with fixed marginals)
- **Way out: Kohn-Sham assumption**: ρ can *also* be obtain from ground state of a non-interacting system (with a different potential, i.e. $\hat{\mathcal{H}}^{\lambda=0} = \sum_{j=1}^N -\frac{1}{2}\Delta_{x_j} + v_{\text{KS}}(x_j)$).

⇒ Problem factorises; universal functional is minimisation over $H^1(\mathbb{R}^3)$ -functions ⇒ **Breaks curse of dimensionality**

$$F^{\lambda=0}(\rho) = \min \left\{ \frac{1}{2} \sum_{j=1}^N \int_{\mathbb{R}^3} |\nabla \psi_j|^2 \mid \psi_i \in H^1(\mathbb{R}^3, \mathbb{C}), \langle \psi_i | \psi_j \rangle = \delta_{ij}, \rho_\Psi = \rho \right\}$$

$$\text{where } \rho_\Psi = \sum_{i=1}^N |\psi_i|^2.$$

¹M. Penz, MFH, T. Helgaker, A. Laestadius. *Electr. Struct. Perspective on Moreau-Yosida Regularization in Density-Functional Theory* (2026).

Kohn-Sham minimisation problem

- Rewrite *interacting* universal functional as

$$F(\rho) = F^{\lambda=0}(\rho) + E_H(\rho) + E_{xc}(\rho)$$

where $E_{xc}(\rho)$ *defined* to make this exact and $E_H(\rho)$ is classical electrostat. term

$$E_H(\rho) = \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(x)\rho(y)}{|x-y|} dx dy,$$

⇒ Kohn-Sham DFT energy minimisation problem:

$$\begin{aligned} E(v) &= \inf_{\rho \in \mathcal{D}} \left(F(\rho) + \int_{\mathbb{R}^3} \rho(x)v(x) dx \right) \\ &= \min_{\substack{\psi_i \in H^1(\mathbb{R}^3, \mathbb{C}) \\ i=1, \dots, N}} \frac{1}{2} \sum_{i=1}^N |\nabla \psi_i|^2 + E_H(\rho_\Psi) + E_{xc}(\rho_\Psi) + \int_{\mathbb{R}^3} \rho_\Psi v \end{aligned}$$

under the constraint $\langle \psi_i | \psi_j \rangle_{L^2(\mathbb{R}^3)} = \delta_{ij}$ (and recall $\rho_\Psi = \sum_{i=1}^N |\psi_i|^2$).

- **Main challenge in KS-DFT:** Develop approximations to E_{xc}

Self-consistent field problem

- All practical approximations \tilde{E}_{xc} are made differentiable

⇒ Satisfy first-order optimality conditions

$$\left\{ \begin{array}{l} \forall i \in 1 \dots N : \left(-\frac{1}{2}\Delta + V(\rho_{\Psi}) \right) \psi_i = \varepsilon_i \psi_i, \\ V(\rho) = v + v_{\text{H}}(\rho) + v_{\text{xc}}(\rho), \quad \text{where } v_{\text{H}}(\rho) = E'_{\text{H}}(\rho); v_{\text{xc}} = \tilde{E}'_{\text{xc}}(\rho) \\ \rho_{\Psi} = \sum_{i=1}^N |\psi_i|^2, \end{array} \right.$$

- **Self-consistent field problem:** Fixed-point problem

$$D(v + v_{\text{H}}(\rho_{\text{SCF}}) + v_{\text{xc}}(\rho_{\text{SCF}})) = \rho_{\text{SCF}}$$

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

$$D(V) = \sum_{i=1}^N |\psi_i|^2 \quad \text{where} \quad \left(-\frac{1}{2}\Delta + V \right) \psi_i = \varepsilon_i \psi_i$$

- Note: **Non-linear** eigenvalue problem, but $\psi_i \in H^1(\mathbb{R}^3, \mathbb{C})$

⇒ No curse of dimensionality

- 1 Kohn-Sham density-functional theory
- 2 Inverse Kohn-Sham using Moreau-Yosida regularisation
- 3 AD-DFPT: Efficient gradients for inverse problems
- 4 Wrapping up

Motivation of Kohn-Sham inversion

- **Forward Kohn-Sham** problem:

Given approx. \tilde{E}_{xc} , thus $v_{\text{xc}} = \tilde{E}'_{\text{xc}}$ solve for approx. ground state

$$\tilde{\rho}_{\text{GS}} = \arg \inf_{\rho \in \mathcal{D}} \tilde{F}(\rho) + \langle v | \rho \rangle \quad \text{where} \quad \tilde{F} = F^{\lambda=0} + E_H + \tilde{E}_{\text{xc}}$$

- Dual pairing $\langle v | \rho \rangle = \int_{\mathbb{R}^3} \rho(x)v(x) dx$

- **Kohn-Sham inversion:** (Note F without approximation)

Given exact $\rho_{\text{GS}} \in \mathcal{D}$ find $v \in \mathcal{V}$ s.t. $\rho_{\text{GS}} = \arg \inf_{\rho \in \mathcal{D}} F(\rho) + \langle v | \rho \rangle$

- Naive motivation: If achieved, then zero gradient at ρ_{GS} yields:

$$0 = F'(\rho_{\text{GS}}) + v = (F^{\lambda=0})'(\rho_{\text{GS}}) + v_H(\rho_{\text{GS}}) + v_{\text{xc}}(\rho_{\text{GS}}) + v$$

\Rightarrow Deduce exact v_{xc} at $\rho_{\text{GS}} \Rightarrow$ Training data to **learn better** $\tilde{E}_{\text{xc}}!$

- There is a catch:

- Universal functional is **everywhere** non-differentiable (both F and $F^{\lambda=0}$, for standard choice $\mathcal{D} = L^1 \cap L^3$)

Motivation of Kohn-Sham inversion

- **Forward Kohn-Sham** problem:

Given approx. \tilde{E}_{xc} , thus $v_{xc} = \tilde{E}'_{xc}$ solve for approx. ground state

$$\tilde{\rho}_{GS} = \arg \inf_{\rho \in \mathcal{D}} \tilde{F}(\rho) + \langle v | \rho \rangle \quad \text{where} \quad \tilde{F} = F^{\lambda=0} + E_H + \tilde{E}_{xc}$$

- Dual pairing $\langle v | \rho \rangle = \int_{\mathbb{R}^3} \rho(x)v(x) dx$

- **Kohn-Sham inversion:** (Note F without approximation)

Given exact $\rho_{GS} \in \mathcal{D}$ find $v \in \mathcal{V}$ s.t. $\rho_{GS} = \arg \inf_{\rho \in \mathcal{D}} F(\rho) + \langle v | \rho \rangle$

- Naive motivation: If achieved, then zero gradient at ρ_{GS} yields:

~~$$0 = F'(\rho_{GS}) + v - (F^{\lambda=0})'(\rho_{GS}) + v_H(\rho_{GS}) + v_{xc}(\rho_{GS}) + v$$~~

⇒ Deduce exact v_{xc} at ρ_{GS} ⇒ Training data to **learn better** \tilde{E}_{xc} !

- **There is a catch:**

- Universal functional is **everywhere** non-differentiable (both F and $F^{\lambda=0}$, for standard choice $\mathcal{D} = L^1 \cap L^3$)

Subdifferentials: Generalising differentiability

- Universal functional F is **convex & lower semicontinuous**¹
- Admits formulation based on **subdifferential**, defined as

$$\partial f(\rho) = \{v \in \mathcal{D}^* \mid f(\sigma) \geq f(\rho) + \langle v, \sigma - \rho \rangle \forall \sigma \in \mathcal{D}\}$$

(Think of set of all possible subtangents in dual of \mathcal{D})

$\Rightarrow \rho_{\text{GS}} = \arg \inf_{\rho \in \mathcal{D}} F(\rho) + \langle v | \rho \rangle$ implies $0 \in v + \partial F(\rho_{\text{GS}})$ or²

$$0 \in v + \partial \left(F^{\lambda=0} + E_{\text{H}} + E_{\text{xc}} \right) (\rho_{\text{GS}})$$

$$\Leftrightarrow 0 \in v_{\text{xc}} + \partial \mathcal{F}(\rho_{\text{GS}})$$

with the convex lsc. (non-differentiable) **guiding functional**

$$\mathcal{F}(\rho) = F^{\lambda=0}(\rho) + E_{\text{H}}(\rho) + \langle v | \rho \rangle$$

- How to find the set $\partial \mathcal{F}(\rho_{\text{GS}})$ in practice ?

¹For appropriate choice of \mathcal{V} and \mathcal{D} , see M. Penz, MFH, T. Helgaker, A. Laestadius. *Electr. Struct. Perspective on Moreau-Yosida Regularization in Density-Functional Theory* (2026).

²Note: Assumes ρ_{GS} to be interacting and non-interacting v -representable, see paper above for details.

Moreau-Yosida regularisation

Definition: MY regularisation

\mathcal{D} unif. convex & $\mathcal{F} : \mathcal{D} \rightarrow \mathbb{R}$ convex lsc. The MY regularisation is

$$\mathcal{F}^\varepsilon(\rho) = \inf_{\mu \in \mathcal{D}} \left(\mathcal{F}(\mu) + \frac{1}{2\varepsilon} \|\mu - \rho\|_{\mathcal{D}}^2 \right) \quad \text{for } \varepsilon > 0$$

- \mathcal{F}^ε is convex, differentiable and has unique minimiser at **proximal point**

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

- Consider proximal point of ρ_{GS} with respect to guiding functional \mathcal{F} :

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

- Stationarity conditions imply

$$0 \in \partial \mathcal{F}(\rho_{\text{GS}}^\varepsilon) + \frac{1}{\varepsilon} J(\rho_{\text{GS}}^\varepsilon - \rho_{\text{GS}}) \quad \text{where } J(\rho) = \partial \frac{1}{2} \|\rho\|_{\mathcal{D}}^2 \text{ (Duality map)}$$

- Comparing with $0 \in v_{\text{xc}} + \partial \mathcal{F}(\rho_{\text{GS}})$ and noting $\lim_{\varepsilon \rightarrow 0} \rho^\varepsilon = \rho$ we obtain¹

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

\Rightarrow We obtain v_{xc} from ρ_{GS} as desired

¹MFH, V. Bakkestuen, A. Laestadius. Phys. Rev. B **111**, 205143 (2025).

Taking inspiration for forward KS for KS inversion

- KS inversion:

$$\rho_{\text{GS}}^{\varepsilon} = \arg \min_{\rho \in \mathcal{D}} F^{\lambda=0}(\rho) + E_H(\rho) + \langle v | \rho \rangle + \frac{1}{2\varepsilon} \|\rho - \rho_{\text{GS}}\|_{\mathcal{D}}^2$$

- Forward Kohn-Sham DFT (more established; codes readily available):

$$\rho_* = \arg \min_{\rho} F^{\lambda=0}(\rho) + E_H(\rho) + \langle v | \rho \rangle + \tilde{E}_{\text{xc}}(\rho)$$

- Based on orbital parametrisation (compare SCF discussion):

$$\rho_{\Psi}(r) = \sum_{i=1}^N |\psi_i(r)|^2 \quad \text{with} \quad \langle \psi_i | \psi_j \rangle = \delta_{ij}$$
$$\implies F^{\lambda=0}(\rho_{\Psi}) = \frac{1}{2} \sum_{i=1}^N |\nabla \psi_i(r)|^2 \quad (\text{Closed-form expression})$$

\implies Use orbital parametrisation also for inverse KS

Choice of function spaces

- MY framework: Only mild assumptions about choice of \mathcal{D} & \mathcal{V} .
⇒ Rigorous framework, which covers many existing KS inversion schemes¹
- **Here:** Periodic problems and plane-wave discretisations
 - Regular lattice $\mathbb{L} = \mathbb{Z}a_1 + \mathbb{Z}a_2 + \mathbb{Z}a_3$
 - $L^2_{\text{per}}(\mathbb{R}^3) = \{f \in L^2_{\text{loc}}(\mathbb{R}^3) \mid f \text{ is } \mathbb{L}\text{-periodic}\}$ & Sobolev analogues
 - Plane-wave basis discretisation

$$u(r) = \sum_{G \in \mathbb{L}} \hat{u}_G e_G(r) \quad \text{with} \quad e_G(r) = \frac{1}{\sqrt{\Omega}} e^{i r \cdot G}$$

- Choose density space $\mathcal{D} = H^{-1}_{\text{per}}(\mathbb{R}^3)$
⇒ potential space $\mathcal{V} = \mathcal{D}^* = H^1_{\text{per}}(\mathbb{R}^3)$
- Computationally feasible², e.g. duality map $J : \mathcal{D} \rightarrow \mathcal{V}$ becomes:


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

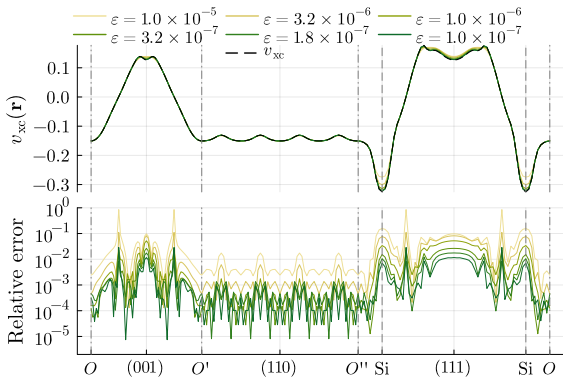
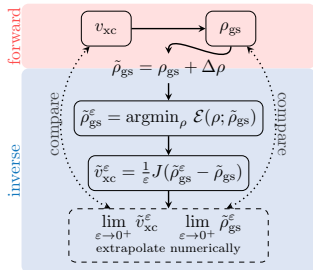
(Convolution with a Yukawa kernel)

¹M. Penz, MFH, T. Helgaker, A. Laestadius. *Electr. Struct. Perspective on Moreau-Yosida Regularization in Density-Functional Theory* (2026).

²MFH, V. Bakkestuen, A. Laestadius. *Phys. Rev. B* **111**, 205143 (2025).

Implementation & numerical tests¹

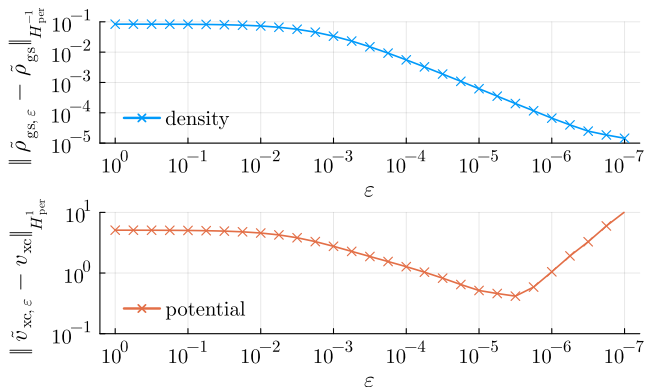
- Implementation in the Density-Functional ToolKit ( DFTK)
- Close match to theory & implementation able to treat realistic systems (Bulk silicon, simple insulators and semiconductors)
- First practical implementation to compute MY proximal point



Silicon crystal

¹MFH, V. Bakkestuen, A. Laestadius. Phys. Rev. B **111**, 205143 (2025).

Practical convergence of silicon



- As $\epsilon \rightarrow 0$ initially v_{xc}^ϵ converges, but then results get worse
 - Known in practice; no rigorous explanation so far
- Based on MY framework: Consider **effect of numerical noise**

First error bounds for KS inversion¹

- Reference densities usually inexact: $\tilde{\rho}_{\text{GS}} = \rho_{\text{GS}} + \delta\rho$
- Triangle-inequality decomposition of the total error

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

- Our analysis establishes non-expensiveness¹ of $\rho \mapsto \rho^\varepsilon$:

$$\|\rho_{\text{GS}}^\varepsilon - \tilde{\rho}_{\text{GS}}^\varepsilon\|_{\mathcal{D}} \leq C \|\rho_{\text{GS}} - \tilde{\rho}_{\text{GS}}\|_{\mathcal{D}} = C \|\delta\rho\|_{\mathcal{D}}$$

with $0 \leq C \leq 1$ depending on $\delta\rho$.

- Using the linearity of duality map J therefore

$$\begin{aligned} \|v_{\text{xc}}^\varepsilon - \tilde{v}_{\text{xc}}^\varepsilon\|_{\mathcal{V}} &= \frac{1}{\varepsilon} \|J(\rho_{\text{GS}}^\varepsilon - \rho_{\text{GS}}) - J(\tilde{\rho}_{\text{GS}}^\varepsilon - \tilde{\rho}_{\text{GS}})\|_{\mathcal{V}} \\ &= \frac{1}{\varepsilon} \|J(\rho_{\text{GS}}^\varepsilon - \tilde{\rho}_{\text{GS}}^\varepsilon) - J(\rho_{\text{GS}} - \tilde{\rho}_{\text{GS}})\|_{\mathcal{V}} \\ &\leq \frac{1}{\varepsilon} \|\rho_{\text{GS}}^\varepsilon - \tilde{\rho}_{\text{GS}}^\varepsilon\|_{\mathcal{D}} + \|\rho_{\text{GS}} - \tilde{\rho}_{\text{GS}}\|_{\mathcal{D}} \leq \frac{1+C}{\varepsilon} \|\delta\rho\|_{\mathcal{D}} \end{aligned}$$

⇒ Qualitative explanation for convergence behaviour

¹MFH, V. Bakkestuen, A. Laestadius. Phys. Rev. B **111**, 205143 (2025).

- 1 Kohn-Sham density-functional theory
- 2 Inverse Kohn-Sham using Moreau-Yosida regularisation
- 3 AD-DFPT: Efficient gradients for inverse problems
- 4 Wrapping up

Inverse design and inverse problems

- θ : Design or **model parameters** (e.g. strain, dopant concentration)
- SCF procedure yields fixed-point density ρ_{SCF}
$$0 = \rho(V(\theta, \rho_{SCF})) - \rho_{SCF}$$
- From these compute quantities of interest: $Q(\rho)$
 - ... to compare against experiment or higher levels of theory

- **Question:** How to tune θ to improve prediction of Q ?
- **Answer:** Follow gradient

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

- Core problem is implicit derivative $\rho'(\theta)$ (**Ingredient 1**)
- Want generality in $Q \Rightarrow$ Need flexible framework (**Ingredient 2**)

Ingredient 1: Density-functional perturbation theory

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

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

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

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

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

\Rightarrow Dyson equation (implicit differⁿ)

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

- Each product $\chi_0 \delta V$ requires solving N Sternheimer equations

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

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

\Rightarrow Nested iterative problem: GMRES + multiple CG

Algorithmic developments to solving DFPT

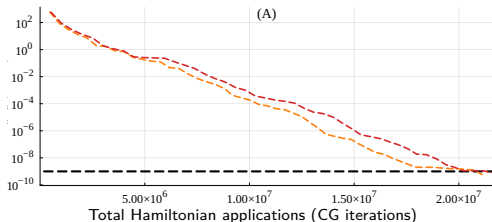
$$\frac{\partial \rho_{\text{SCF}}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta}$$

GMRES tolerance τ

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

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

1. Inner Sternheimer solver becomes badly conditioned for metals:
 - **Schur complement** approach: Re-uses orbitals of SCF stage¹
2. Tolerance for Sternheimer solver when applying χ_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 order to attain desired accuracy



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

¹E. Cancès, MFH, G. Kемlin, *et. al.* Lett. Math. Phys. **113**, 21 (2023).

²MFH, B. Sun. SIAM J. Sci. Comp. *Efficient Krylov methods for lin. resp. in PW electronic structure calculations* (2025). arXiv:2505.02319

Algorithmic developments to solving DFPT

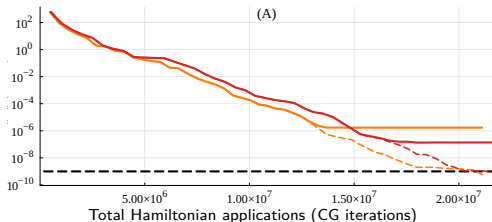
$$\frac{\partial \rho_{\text{SCF}}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta}$$

GMRES tolerance τ

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

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

1. Inner Sternheimer solver becomes badly conditioned for metals:
 - **Schur complement** approach: Re-uses orbitals of SCF stage¹
2. Tolerance for Sternheimer solver when applying χ_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 order to attain desired accuracy



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

¹E. Cancès, MFH, G. Kемlin, *et. al.* Lett. Math. Phys. **113**, 21 (2023).

²MFH, B. Sun. SIAM J. Sci. Comp. *Efficient Krylov methods for lin. resp. in PW electronic structure calculations* (2025). arXiv:2505.02319

Algorithmic developments to solving DFPT

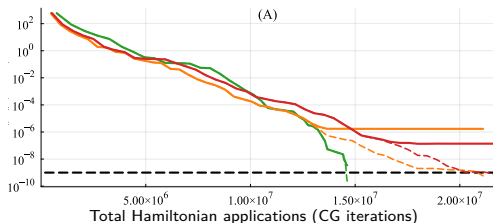
$$\frac{\partial \rho_{\text{SCF}}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta}$$

GMRES tolerance τ

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

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

1. Inner Sternheimer solver becomes badly conditioned for metals:
 - **Schur complement** approach: Re-uses orbitals of SCF stage¹
2. Tolerance for Sternheimer solver when applying χ_0 ?
 - Naive strategies: $\tau_{i,n}^{\text{CG}} = \tau/100$ and $\tau_{i,n}^{\text{CG}} = \tau/10$ for $\tau = 10^{-9}$
 - **Inexact Krylov**: $\tau_{i,n}^{\text{CG}}$ grows adaptively; accuracy retained²

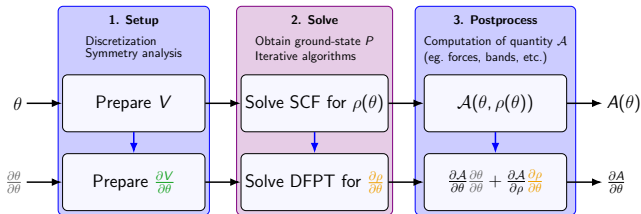


- Dashed: GMRES estimated residual norm
- Solid: Actual residual norm

¹E. Cancès, MFH, G. Kемlin, et. al. Lett. Math. Phys. **113**, 21 (2023).

²MFH, B. Sun. SIAM J. Sci. Comp. *Efficient Krylov methods for lin. resp. in PW electronic structure calculations* (2025). arXiv:2505.02319

Ingredient 2: Algorithmic differentiation (AD)



- Algorithmic differentiation (AD)

- **Generic framework** for derivatives: Request gradient, AD delivers
- **Chain rule**: Automatically compose gradients out of primitives
 - *Primitives*: Floating-point operations, diagonalisation, **SCF**
- Custom rule: For **SCF problem** gradient is DFPT

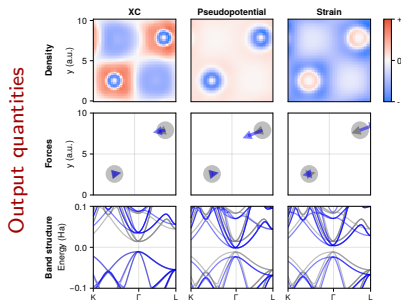
$$\frac{\partial \rho_{\text{SCF}}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta}$$

⇒ **AD-DFPT**: Robust DFPT solver plus AD framework in  **DFTK**¹

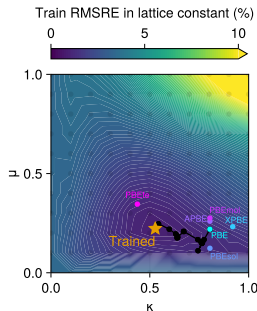
¹N. Schmitz, B. Ploumhans, MFH. npj Computat. Mater. 12, 6 (2025).

AD-DFPT showcases

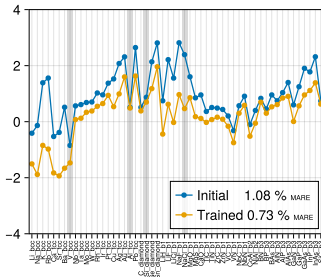
Parameters θ



Combinatorial sensitivity of key DFT output quantities¹

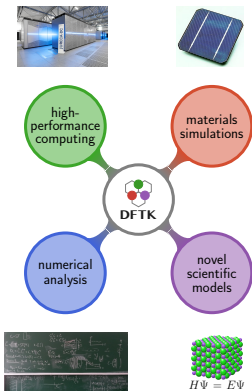


Test error in lattice constant (%)



- Inverse design of E_{xc} wrt. experimental lattice constant
- Note: Q involves another optimisation problem (lattice relaxation)

- 1 Kohn-Sham density-functional theory
- 2 Inverse Kohn-Sham using Moreau-Yosida regularisation
- 3 AD-DFPT: Efficient gradients for inverse problems
- 4 Wrapping up




- **Julia** code for **cross-disciplinary research**:
 - Allows restriction to **relevant model problems**,
 - **and scale-up** to application regime (1000 electrons)
 - Sizeable feature set in ca. **10k lines** of code:¹ meta-GGA, DFT+U, Hybrid-DFT (experimental)
- **Fully composable** due to **Julia** abstractions:
 - Algorithmic differentiation (AD)
 - HPC tools: MPI, **Nvidia & AMD GPUs**
- **High-productivity** framework & established **community**:
 - > 50 contributors in 6 years (Maths, physics, CS, ...)
 - Instrumental in a dozen of research works
- **Unique features**¹:
 - Self-adapting algorithms
 - Algorithmic differentiation
 - Numerical error estimates (e.g. basis set error in forces)

¹<https://docs.dftk.org/features>


Summary & Outlook


- Density-functional theory **understudied in mathematics**
 - Fascinating interdisciplinary topic
 - **Impact:** 30% supercomputer share; design of materials
- Inverse Kohn-Sham methods
 - **Kohn-Sham inversion:** Potentials from reference densities
 - **Moreau-Yosida regularisation:** Dealing with non-differentiability
 - **First error analysis** techniques for Kohn-Sham inversion
- Next steps
 - **Quantitative** *a posteriori* error analysis \Rightarrow Guided choice of ε
 - Understand **point-wise error** in v_{xc}
 - Improve **robustness** of inversion algorithms
 - Use potentials for **data-driven training** of E_{xc}


Questions?


 <https://matmat.org>


 mfherbst  @herbst @social.epfl.ch

 michael.herbst@epfl.ch

 https://michael-herbst.com/talks/2026.04.10_Swiss_Numerics_Day.pdf

 M. Penz, MFH, T. Helgaker, A. Laestadius. *Electr. Struct. Perspective on Moreau-Yosida Regularization in Density-Functional Theory* (2026).

 MFH, V. Bakkestuen, A. Laestadius. *Phys. Rev. B* **111**, 205143 (2025).

 DFTK <https://dftk.org>

- 5 Kohn-Sham DFT
- 6 Details on Schur complement approach
- 7 Details on inexact Krylov

Kohn-Sham theory

$$F_{\text{LL}}(\rho) = \min_{\substack{\Psi \in W_N \\ \rho_\Psi = \rho}} \left\langle \Psi \left| \left(-\frac{1}{2} \sum_{j=1}^N \Delta_{x_j} + \sum_{j < k}^N \frac{1}{|x_j - x_k|} \right) \Psi \right. \right\rangle$$

- Assume the ground state had the form of a **Slater determinant**

$$\tilde{\Psi}(x_1, \dots, x_N) = \det \begin{pmatrix} \psi_1(x_1) & \psi_1(x_2) & \cdots & \psi_1(x_N) \\ \psi_2(x_1) & \psi_2(x_2) & \cdots & \psi_2(x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_N(x_1) & \psi_N(x_2) & \cdots & \psi_N(x_N) \end{pmatrix}$$

where $\psi_i \in H^1(\mathbb{R}^3, \mathbb{C})$; $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ (det is exact for *non-interacting* electrons)

- In this case

$$\rho_{\tilde{\Psi}}(x) = \sum_{i=1}^N |\psi_i(x)|^2 \quad \left\langle \tilde{\Psi} \left| \hat{\mathcal{H}}^{\lambda=0} \tilde{\Psi} \right. \right\rangle = \frac{1}{2} \sum_{j=1}^N \int_{\mathbb{R}^3} |\nabla \psi_j|^2$$

- The equivalent of F for this system is

$$T_{\text{KS}}(\rho) = \min \left\{ \frac{1}{2} \sum_{j=1}^N \int_{\mathbb{R}^3} |\nabla \psi_j|^2 \mid \psi_i \in H^1(\mathbb{R}^3, \mathbb{C}), \langle \psi_i | \psi_j \rangle = \delta_{ij}, \rho_{\tilde{\Psi}} = \rho \right\}$$

- Minimisation over $H^1(\mathbb{R}^3)$ functions \rightarrow **Breaks curse of dimensionality**

- 5 Kohn-Sham DFT
- 6 Details on Schur complement approach
- 7 Details on inexact Krylov

Schur complement approach to response¹

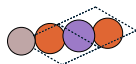
- Numerics of eigensolver:
We have N_{ex} “extra” bands
- Use these to partition \tilde{H} :

$$\tilde{H} = \begin{pmatrix} E_{\text{ex}} & \mathbf{C} \\ \mathbf{C}^\dagger & \mathbf{R} \end{pmatrix}$$

$E_{\text{ex}} = \text{diag}(\varepsilon_{N+1}, \dots, \varepsilon_{N+N_{\text{ex}}})$
& \mathbf{C} , \mathbf{R} projections of \tilde{H}

⇒ Use **Schur complement**:
Better-conditioned systems

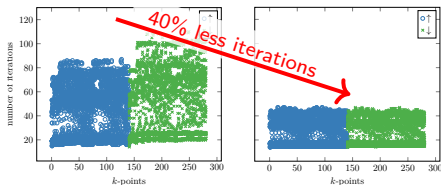
$$(\mathbf{R} - \varepsilon_i)x = b$$



Fe₂MnAl Heusler alloy

standard approach

Schur complement



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

¹E. Cancès, MFH, G. Kemplin, *et. al.* Lett. Math. Phys. **113**, 21 (2023).

Sternheimer equations

$$\frac{\partial \rho_{\text{SCF}}}{\partial \theta} = [1 - \chi_0 K]^{-1} \chi_0 \frac{\partial V}{\partial \theta} \quad (2)$$

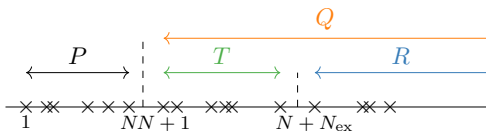
- Solving (2) (Dyson equation) is not cheap
- Each *application* of χ_0 to a δV requires **iteratively solving** (3) (Sternheimer equation) for all $n = 1, \dots, N$

$$\Pi_Q (H - \varepsilon_n) \Pi_Q \delta \psi_n = -\Pi_Q \delta V \psi_n \quad (3)$$

where

- $\delta \psi_n$: Orbital perturbation (to be determined)
 - $P = \text{span} \{ \psi_n \mid n = 1, \dots, N \}$: 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 .
- **Caveat:** (3) is badly conditioned if gap $\varepsilon_{N+1} - \varepsilon_N$ small
⇒ Response can be expensive for **metals**

Schur-complement approach¹ (1)



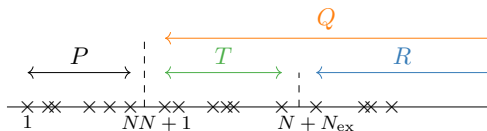
- SCF diagonalisations yield N_{ex} additional orbitals $\Phi = (\psi_{N+1}, \dots, \psi_{N+N_{\text{ex}}})$ spanning T .
 - Not fully converged, i.e. $H\psi_n \neq \varepsilon_n\psi_n$ for $n = N + 1, \dots, N + N_{\text{ex}}$
 - But: $\Phi^\dagger H \Phi = \text{diag}(\varepsilon_{N+1}, \dots, \varepsilon_{N+N_{\text{ex}}})$
- Split orbital perturbation $\Pi_Q \delta\psi_n = \Phi \alpha_n + \Pi_R \delta\psi_n^R$ to obtain:

$$\Pi_Q (H - \varepsilon_n) \Phi \alpha_n + \Pi_Q (H - \varepsilon_n) \Pi_R \delta\psi_n^R = \underbrace{-\Pi_Q \delta V \psi_n}_{:=b_n}$$

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

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

Schur-complement approach¹ (2)



$$\begin{aligned} & \Pi_Q(H - \varepsilon_n)\Phi\alpha_n \\ & + \Pi_Q(H - \varepsilon_n)\Pi_R\delta\psi_n^R = b_n \end{aligned}$$

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

- It only remains to iteratively solve the component $\delta\psi_n^R$:

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

- Π_R almost removes small eigenmodes of $H - \varepsilon_n$
- ⇒ Smallest eigenvalue of $\Pi_R(H - \varepsilon_N)\Pi_R$ is about $\varepsilon_{N+N_{\text{ex}}} - \varepsilon_N$
- For metals: Substantially larger than $\varepsilon_{N+1} - \varepsilon_N$
- ⇒ Improved conditioning

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

- 5 Kohn-Sham DFT
- 6 Details on Schur complement approach
- 7 Details on inexact Krylov

Inexact Krylov methods for DFPT (1)

- **Inexact GMRES:** Tolerable errors for $1 - \chi_0 K$

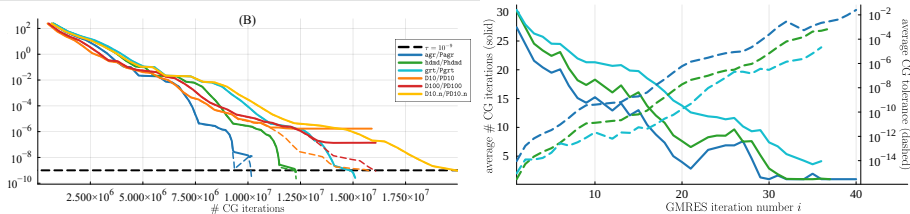
⇒ **Theorem:**¹ Guaranteed convergence of GMRES to τ when

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

- $\|\tilde{r}_{i-1}\|$: 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
 - 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 $\frac{\sqrt{|\Omega|}}{N} \searrow$)

¹MFH, B. Sun. SIAM J. Sci. Comp. *Efficient Krylov methods for lin. resp. in PW electronic structure calculations* (2025). arXiv:2505.02319

Krylov methods for DFPT (2)¹



- **Guaranteed (grt)** computes C exactly
- **Balanced (hdm)** 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 hdm, but can be a factor 10 off
- **Balanced:** From about 20M to 12M Hamiltonian applications (the expensive step)

¹MFH, B. Sun. SIAM J. Sci. Comp. *Efficient Krylov methods for lin. resp. in PW electronic structure calculations* (2025). arXiv:2505.02319